data.org

The Data Ecosystem Designer: Designing the Future of Digital Public Goods

by Danil Mikhailov, Ph.D., Uyi Stewart, Ph.D., Perry Hewitt, Jake Porway, Ph.D.

The Need for Data Ecosystem Design

Since the Big Data age of 2012, people have begun imagining a data-driven social sector. In this vision, folks imagine a world where the social sector is capable of creating new data-driven digital products for accomplishing its goals. This yearning for a tech-enabled social sector has only grown as the hype around AI has accelerated post-2023. Many efforts focused on creating more digital products in the nonprofit space have launched over the last few decades. However, to date, most have focused on helping one single organization develop a digital solution to one challenge they face. Digital solutions that truly scale have proven to be more difficult to create, largely due to the complex network of stakeholders, data sources, and funding streams needed to achieve them.

In their 2023 report "Accelerate Aspirations: Moving Together to Achieve Systems Change", data.org noted that *"improvement [to the use of data] lies in better coordination between diverse partners and actors, and long-term, well-resourced multi-sector (public, private, academic, SIO, philanthropy) partnerships based on honest and clear conversations about our diverse incentives.*" (Mikhailov, 2023) This quote points to the unique challenges of creating scalable digital products in the social sector. If we need an entire ecosystem of actors, datasets, and technologists aligned around a digital product for it to scale, then how do we codify that practice? What skills are needed to do this? And who do we depend on to make it happen?

We hypothesize that a new role is needed for the social sector called the "Data Ecosystem Designer" that is charged with helping digital products thrive and scale by organizing and aligning players within a data ecosystem. Though this role is not yet formalized, we identified and interviewed a number of people in the social sector who are currently informally holding this role. These folks are often largely responsible for the success of digital products that serve a collective or disparate set of actors and often have to invent new, unique ways of sustaining and maintaining their work. The people we spoke with were effectively responsible for recommending, designing, and aligning the ecosystem within which their digital solution would live. Until we better define the set of skills needed and the career paths for this role, organizations will not plan for them, individuals will not train in them, and funders will not recognize the critical importance of funding them. We believe that better defining, training for, and funding this data ecosystem designer role could create a step change in how many digital public goods are available and sustainable in the space.

Exemplars of Data Ecosystem Design

Before defining the role further, let us lay out some examples of digital public goods that a Data Ecosystem Designer might oversee:

1. Medic: Medic serves as a steward for the Community Health Toolkit (CHT), a mobile data collection and messaging platform that allows health workers to register pregnancies, track patient data, receive automated reminders and alerts, and communicate with supervisors and clinicians. The tool is intended to be used by community health workers across different geographies, different literacy levels, and different workflows. Moreover, the software was designed to be useful both to the health workers and the districts and governments they work within. In order to balance local context, global need, and a range of user features, Medic has put great effort into understanding the needs of multiple stakeholders and marrying that with sound product design. *Read more about their case study here.*

2. Ushahidi: Ushahidi evolved from a realization that more and more people were using personal devices and social media to document information about their environments, yet there was no simple platform to aggregate this information for those who could put it into action. Originally used for helping first responders crowdsource information from people on the ground during a humanitarian crisis, Ushahidi eventually developed use cases for election monitoring and human rights monitoring. Through a network of government and private partners, Ushahidi raised funding for their product, established data sharing partnerships with many organizations, and supported their clients in using their crowdsourcing tool to maximize impact. *Read more about their case study here.*

3. ISARIC: ISARIC is home to a set of data and tools used by a network of epidemiologists, researchers, and governments around the world to advance treatment and prevention of infectious diseases. As a result, advances have been made on tracking COVID19, monitoring non-febrile illness, and advancing inter-laboratory collaboration. The results would have been hard, if not impossible, for any one organization to achieve. Interviews with leaders at the ISARIC revealed that some of the key aspects in making this multistakeholder digital good operational were deep interviews with scientists on their usage patterns, managing complex international data sharing agreements, and finding funding support for the tool. *Read more about their case study here.*

4. data.org's Epiverse library: We would be remiss not to mention our own efforts with Epiverse, a library of digital public goods for pandemic prevention. Epiverse was designed from the start with a diversity of users and digital products in mind. Scientists and researchers will need varied sets of tools to conduct their data work around disease, all of which themselves will require diverse ecosystems of data sources and funding to function. By nature of the many partners they serve under many contexts, Epiverse tools

will require strategic design of their sustainability and use plans from the start. *Read more about their case study here.*

What makes these tools and their design unique? After all, digital products are nothing new, so why can't we apply the same principles to developing these digital products in the social sector? Looking at all of these examples, we can see that they differ from traditional commercial digital products in a few key ways:

Impact is as important as, if not more important than, use: In a commercial context, success is defined, crudely, by the number of sales of the digital product or the number of users that support an ad revenue model. For example, Spotify would rank its success by the number of subscribers to their service. There is a valid argument that corporations don't have to be concerned with social impact at all (Hull, 2006). Digital public goods, however, have an increased onus to deliver a specific societal impact to be considered successful. Moreover, that societal impact may not be "owned" by any one customer, but instead arises from the collective use, maintenance, and advancement of the tool by many actors in the community data ecosystem. For example, while it is impressive that GetCalFresh is used by every California county, it must also show that doing so reduces hardships on Californians that require food assistance. Therefore the creators and shepherds of digital public goods have to think past the "sale" of the product to the future use of the product to the ultimate impact it would create.

Profits don't scale with use: In many contexts, the users of the digital public goods will not be the ones paying for the service. Californians do not pay for GetCalFresh, nor do individual scientists pay to use Epiverse. Therefore, third party revenue streams need to be fostered, such as selling to government agencies, acquiring grant funding, or monetizing other elements of the product. Unlike private sector models, the monetization often cannot include the re-selling of user data. Other researchers have expanded on the many financial challenges to supporting digital public goods and their infrastructure (Saviano, 2022).

Customers are not uniform: In the for-profit context, many digital products and services are marketed toward a common set of customers with common infrastructure. For example, Spotify is designed for the broad class of "music listeners" and is supported on any mobile phone or computer. Industry-specific products, like customer relationship management (CRM) software, are tailored toward businesses with common use cases and digital infrastructures. Unfortunately, in the social sector, customers are much more diverse. We know that we need buy-in from leadership and an agility in using digital goods for organizational transformation (Jong 2023), but not all organizations have these qualities. Digital goods in the social sector may aim to serve a set of individuals with similar missions or goals, such as health care professionals, but the capability of and implementation with each partner can vary wildly. The

social sector's digital infrastructure is not as mature or uniform as that of business, nor are there as many strong regulations to dictate how certain processes work. As a result, the customers these goods aim to serve often have very different data, technology capacity, processes, and business skill levels and objectives.

Because of these peculiarities, digital public goods need to be thoughtfully designed to achieve social impact in the world. They need to incorporate a knowledge of the unique and diverse set of customers they serve, they need to work in varied data environments, and they need to be funded through methods that may not easily scale with adoption.
The definition above may still sound like a standard digital product. By comparison, here are some versions of digital public goods that we would not consider within the realm of a Data Ecosystem Designer as they don't involve creative engineering of an ecosystem of actors to accomplish their goals:

A market-based solution for the social sector: There are certainly digital products and services that are sold to nonprofits and governments in the social sector, such as fundraising management tools like Raiser's Edge or low-cost versions of commercial software like Salesforce for nonprofits. These are digital goods that many nonprofits and governments use, but they lack a common impact goal, make use of shared infrastructure that all customers have, and are either paid for by each customer or subsidized by the company's other profit centers.

An open source solution for the social sector: A common scaling solution for social sector digital products is to make them publicly available and open source. Such solutions are available to a wide set of customers and can contribute to widespread impact, but they fall short of the intentional customization and sustainability plans of the products we're focusing on. In their report on Open Source Software in the social sector, GitHub found that even very successful and useful projects suffered from lack of discoverability by the organizations that could use them, lack of design principles to make the tools usable, and a lack of long-term maintenance funding (Github, 2023). Open source is often necessary, but not sufficient, for making a useful digital public good without someone to support those needs. Equally, some of the products mentioned above, like Epiverse, also happen to be open source, but it is their intentional ecosystem engineering that distinguishes them, not their open source nature.

The Dimensions That Support Digital Products

If the role of the Data Ecosystem Designer is to foster a healthy ecosystem for their digital product, then we must identify what types of resources and actors they have to align within that ecosystem. From our early conversations with folks in this role, they identified a number of dimensions needed for their work to succeed[1]:

- **Impact Statement:** An impact statement, synonymous with the metric and population of interest in a Theory of Change, defines how much change, of what kind, and for whom is expected from this product. Examples could be “infant mortality within the following regions is reduced by 50% in five years”, “time for first responders to respond to humanitarian emergencies is decreased by 30%” or “number of research products produced by institutional scientists is doubled”. This is different from purely commercial products, which measure their success simply by customers acquired or profits made. The impact statement defines the people and resources within the ecosystem who need to be involved, as they are the ones aligned toward achieving this goal.
- **Direct Stakeholders / Users:** The next dimension within the data ecosystem is the direct stakeholders, or the users, of the digital product. These individuals are the people responsible for directly applying the tool to create the intended impact and are often nonprofit leaders, government agencies, or researchers. They are aligned toward the impact statement, but they often have different institutional practices, different workflows, or different contexts that they work in. For example, if one were creating a universal AI solution for traffic flow prediction for cities, they would be keenly aware that each city government can differ in how they measure, manage, and even define aspects of transportation. While a data ecosystems designer hopes that their users are as similar as possible, the overriding commitment to the impact statement requires some flexibility for various user needs.
- **Indirect Stakeholders:** The third dimension within the data ecosystem is the indirect stakeholders. Indirect stakeholders include the people and institutions that are not direct users of the product, but are either required for its success or impacted by its use. For example, imagine a digital tool that aims to reduce COVID transmission rates by implementing contact tracing and notifying people when they’ve been in proximity of someone infected with COVID-19. The direct stakeholders of such a tool would be the public who are using the contact tracing app and government officials in a Department of

[1] Note that there is a subtle distinction here between the ecosystem that supports a digital product and data.org’s work on community data ecosystems. In their ReCODE report (Gottschalk, 2023) data.org defined a community data ecosystem as “made up of the what, the who, and the how that enables data sharing and collaboration within a community. They include data infrastructure, tools, user capabilities, standards, and policies used to manage cross-organization data sharing." These community data ecosystems are defined by the community members working to solve problems by aligning data resources. In this report, we are interested in the specific types of actors in that ecosystem that are required to help a digital product thrive and scale within that community data ecosystem.

Health who need to act on the data. For the tool to reach its intended goal though, individuals need to download, use, and trust the application. Medical professionals might be needed to help advocate for the tool and persuade patients to enroll in it. Researchers might want to use the data to learn about more effective ways to reduce transmission. These two groups are indirect stakeholders because they are not the primary users of the application, but they are needed to ensure its impact. Commercial products, by contrast, are successful largely through only the actions of Direct Stakeholders.

- **Datasets:** Key to any data-driven digital product is the data itself! Digital products in the social sector often need to combine multiple non-traditional streams of data together in order to work. Examples include cases where the digital product depends on the user's internal data to work but the data is not standardized, cases where multiple organizations need to pool data for the digital product to work, or cases where multiple 3rd party data sources are needed for the digital product to work. A well-functioning digital product in this data ecosystem will have access to all the data streams it needs to run, and the privacy, security, and usage goals of all the data contributors will be honored. In commercial settings, the data is often internal to the organization building the product or purchased by them from a 3rd party for use in their tool.

- **Technical Capacity:** In order to design, build, and maintain the digital product, some amount of technical capacity will be needed within the data ecosystem. In commercial products, technical capacity is usually exclusively in-house at the organization building the tool. In community data ecosystems, technical capacity is often fragmented or temporary. For example, a group of university researchers may work together to build an improved COVID rate prediction model. If successful, such a tool will need consistent maintenance. However, software maintenance is not a common university function, so special funding, staffing, or other allowances would need to be made for the tool to be maintained there. Other digital products release their code open source in an attempt to cultivate enough of a volunteer technical community to continue to maintain the product.

- **Funders:** Digital public goods depend on funding streams to support their care and maintenance. Funding is needed to compensate all of the stakeholders in the ecosystem above and to allow for experimentation and improvement. In commercial products, funding eventually comes primarily from customers. In social sector products, customers are unlikely to be the funders, so funding needs to be creatively engineered. Many digital products cobble funding together from a number of sources, notably well-resourced clients that have the capacity to pay (often governments), grant funding, and corporate sponsorship.

These six dimensions broadly define the key components of the community data ecosystem that a data ecosystem designer has to align and organize.

A Note on Scale

One of the complexities to creating a digital public good comes from the varied way in which one “scales” a digital product. Below are some different ways in which a digital public good could scale to reach impact. The data ecosystem designer will have to navigate how to organize the data ecosystem based on which form of scaling the product needs to take.

- **Traditional scale:** Increasing the number of users who use the product to do X. For example, scaling our hospital bed inventory predictor from 10 hospitals to 100 hospitals.
- **Use case scale:** Increasing the number of users who use the product to do Y, without alienating the users who use it to do X. For example, expanding the hospital bed inventory predictor to be used as a homeless shelter bed inventory predictor.
- **Leveraged scale:** Identify a core set of users who, if they used the tool to do X, would have outsized impact. For example, convincing the largest national hospital system to use the hospital bed inventory predictor, thus reaching the largest population.
- **Geographic scale:** Increasing the number of users who use the product to do X in a new geography. For example, transferring the hospital bed predictor from Kenya to Chile.
- **Community scale:** Increasing the number of people contributing resources to the tool. For example, aligning a funder collective to provide long term monetary support to the hospital bed predictor, or creating an active community of volunteer engineers to keep the code up to date.

We believe the data ecosystem designer will have to name and account for the ideal version(s) of scale for their product.

Responsibilities of the Data Ecosystem Designer

Now that we've discussed the actors needed to sustain a digital public good and the ways a tool might scale, we can talk about the responsibilities the Data Ecosystem Designer will have to manage. The Data Ecosystem Designer is responsible for ensuring the product is adopted by a broad range of users, applied to create impact, and is sustained financially and technologically. When speaking with our interviewees, they cited two main prongs of responsibilities commonly held by someone serving in this role:

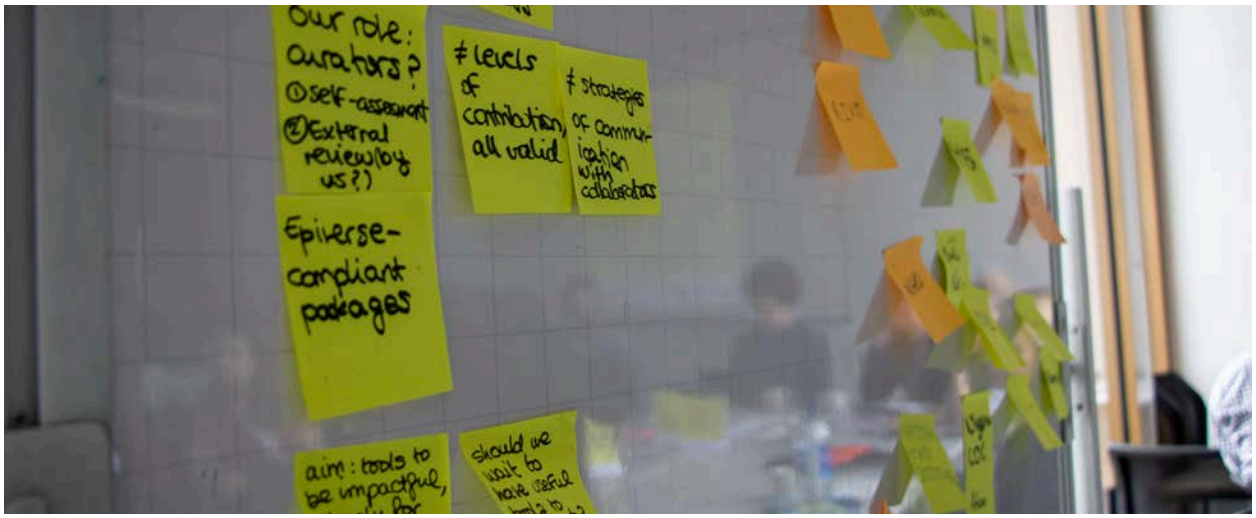
1. Product Ownership and Evangelism

- Successfully convey the vision for the product and how using it will meaningfully impact the world
- Recognize what classes of problems the product can address and identify customers with those problems
- Convey to partners how to use the tool and what ROI they will get from using it
- Invest in user-centered design, and learn from analytics and qualitative feedback once the product is in use so that the internal team can make improvements

2. Partnership Development

- Identify and foster the relationships with external parties within the community data ecosystem that are needed to sustain the tool.
- Technical partnerships: For most data-driven products we explored, data was generated by the customers or third parties, like businesses, in order for the tool to be maximally effective. As a result, Data Ecosystem Designers described having to identify partners that had data and technological capacity and then build data sharing and governance plans with them.
- Beneficiary partnerships: Many of the digital products we researched affected folks that were some degrees removed from the organization using it. As such, the Data Ecosystem Designer had to onboard and align those folks with the use of the tool. For example, the folks designing Ushahidi described how they first had to onboard disaster relief organizations like the Red Cross to use their tool, then had to evangelize the tool's use to first responders and governmental crisis teams so that they understood how Red Cross was using the tool and trusted the results. Some of these partners can become champions, invested in the product and ecosystem.

- Funding partnerships: Data Ecosystem Designers across the board described holding some responsibility for identifying and cultivating revenue streams for their projects. Many products started out using grant funding, but that funding stream would often expire once the product had reached maturity. Data Ecosystem Designers would have to design subscription models, individual donations, or alternative revenue streams to continue to fund the support and growth of the product.



What we can see from this list is that the Data Ecosystem Designer is responsible for a large amount of partnership management, evangelism, and creative revenue design[2]. When interviewing folks who held a role like this one, we found that they often had titles like “Digital Product Lead”, “Innovation Manager”, or “Lead Scientist”. All were quick to acknowledge that those titles weren’t perfect fits for what they were doing. They also felt that the role they were playing was distinct from some other roles within traditional nonprofit and product organizations, namely:

- **Chief Fundraiser / Chief Development Officer / Chief Digital Officer / Chief Revenue Officer:** Though the Data Ecosystem Designer has some responsibility for identifying funding and revenue, almost all of the interviewees we spoke with felt that a typical CDO/CRO had more broad-reaching fundraising skills but not as much experience selling technical products.

[2] Note that we refer to the Data Ecosystem Designer in this report as if they are a single person. However, they will of course need teams of folks to fulfill all of these skills. We see the role as analogous to a City Planner, which is a title that an individual holds but it is understood they are not required to perform all activities of city planning themselves. We will address team makeup in Teams with a Data Ecosystem Designer.

- **Technical Product Owner / Data/ML/AI Product Owner:** Though the Data Ecosystem Designer has a lot of responsibility for overseeing the digital product's lifecycle and entry to the market, many interviewees felt that traditional Technical Product Owners would not have the experience to marshal multiple stakeholders from different sectors to achieve success.

Skills of the Ecosystem Designer

Given the responsibilities in the previous section, the Data Ecosystem Designer needs to have a number of key skills to be successful.

Data Governance Experience: Separate from general technical know-how, participants cited data governance as a critical skill for success. Because the tools the Data Ecosystem Designer oversees will be digital products that rely on a number of different users and data providers, data governance is a critical tool for ensuring that all stakeholders are clear on how their data is used, what rights they have to the outputs of the tool, and how their data is kept secure. As the Data Ecosystem Designer is working within a community data ecosystem of actors, their

success will depend on their ability to honor that community's values in the use, combination, and reuse of its data. Interviewees also mentioned that those with strong data governance experience often had skills in ethical and responsible data use, which was critical for designing products that were safe as well as effective.

Technical Experience: All interviewees cited technical know-how as a requisite for this role. Clearly, in order to evangelize for and oversee a data science or machine learning solution, the Data Ecosystem Designer has to have enough capability with the technology to be able to recognize when and how such a technology could be used appropriately. How much technological skill the Data Ecosystem Designer needed, however, was not unanimous. All felt that the Data Ecosystem Designer did not have to get into the weeds of coding and data infrastructure design, but some felt strongly that they had to have had experience working at that level in order to fully understand the limitations of the technology and the constraints of the technologists building the tool. Others felt that the Data Ecosystem Designer would be successful if they could simply understand the technology stack overall and what problems each technology solved, paving the way for non-technologically trained folks to hold this role after some onboarding to the use cases of their specific product.

Product Development Experience: All interviewees also agreed that this person should have experience overseeing digital product development. Specifically, they felt people in this role needed to understand the lifecycle of technology product development and have experienced the challenges and pitfalls that go along with that. For example, one interviewee pointed out that user feedback and testing, an important step in product development generally, was exponentially more important for digital public goods, as there were so many different types of customers in the ecosystem using the product. A good Data Ecosystem Designer therefore would need to be well-versed in user research to carry out this task in a multistakeholder situation. Other aspects of digital product development, such as prototyping, data and product validation, and launch, are also critical for the Data Ecosystem Developer to have experience with.

Experience with Public Institutions: Digital public goods are explicitly built for not-for-profit uses. Therefore, Data Ecosystem Designers with skills navigating not-for-profit institutions and needs are better suited to align the product to their needs. They also understand the constraints that exist within governments and nonprofits that don't exist in for-profit institutions. Note that interviewees did not suggest that Data Ecosystem Designers *had* to have worked in the social sector, but they cited a deftness in working within political, low-resourced, or impact-driven contexts that often comes from this experience.

Partnership Management: Data Ecosystem Designers will have to negotiate the needs of multiple users and partners within the ecosystem. They will need to align the design of the digital product in service of a host of partners who have a similar high-level need, but also

specific needs and preferences. For example, health care workers may all want a system to digitally triage patients, but hospital staff may have slightly different workflows than community health workers. Ministry of health agents may have other workflows on top of that. They will also have to know how to communicate the vision of the product to new customers in a way that speaks to their unique needs. Lastly, they'll need to be able to work with external partners who can provide data, technical capacity, or funding to support the product. All of these partnerships require excellent communication skills, excellent interpersonal skills, and the savvy to be able to create win-win situations with folks who have different needs.

Empathy for all stakeholders: Perhaps the most oft-cited skill came down to a skill we're dubbing "empathy for all stakeholders". This skill was called "understanding the user", "experience with the industry", "playing the role of both vendor and customer", and "having first-hand experience with partner incentives" in our interviews. What was common across all of these terms was that the Data Ecosystem Designer needed to have some ability to empathize, most usually from personal experience, with the needs and experiences of all of the stakeholders involved. Said differently, the Data Ecosystem Designer must recognize with humility their place as coordinator within an ecosystem, not the unilateral top-down decisionmaker in that ecosystem.

Local context: As stated above, it is critical that a Data ecosystem designer be able to empathize with all stakeholders. To do so, they must have some connection to the context of the work being done. This could be experience living in the geographic region within which the tool will be deployed, having experience with one of the stakeholder groups, or having been involved with the organizations using this product. Even if they do not have direct experience, they must be acutely aware of the local contexts of all stakeholders so that they can accurately speak to their needs. Others have noted this skill in terms like "voicing the unheard", which allows the Data Ecosystem Designer to account for all users needs (Nicholson, 2022)

Ability to code switch: Data ecosystem designers need to have a range of different product and technical conversations that go beyond just a surface level understanding of the material. These folks must have the ability to go deep on an issue, but then summarize the key takeaways and clearly explain the value of the product and ecosystem in a way that resonates with the audience they're working with.

Appetite for risk: We heard a skill come up in many of our interviews that was difficult to put a finger on, but often translated to "bravery" or "chutzpah". Because this role is fairly new, the folks who hold it often have to blaze trails. They also have to negotiate partnerships and relationships with external parties that might be non-traditional, such as inventing new data governance models for a government agency or establishing new public-private partnerships. As such, many of the Data Ecosystem Designers we spoke with cited their willingness to take risks, invent new business models, and innovate where there was no playbook to follow. Perhaps over time, as

the role becomes more standardized, this quality won't be needed as much. For now, however, it seems critical to the people holding this role.

What is perhaps most striking about the list of skills above is that so few relate to technology. Many of the digital products these folks are designing are data science models, digital software, or algorithms that fall under the banner of machine learning and artificial intelligence. However, outside of technology and data governance skills, the remaining requirements rely very heavily on political savvy, local context, strategic thinking, and interpersonal relationships. It seems from our early research that the ability to shepherd large-scale multi stakeholder projects may be more important than technological skills to a Data Ecosystem Designer.

Teams with a Data Ecosystem Designer

The archetypes above help classify the types of skills different Data Ecosystem Designers may come with. However, no Data Ecosystem Designer is an island. In our workshop with folks who have held this role informally, they cited a number of key partnerships that needed to be formed in order to make their digital public good a success. Who they decided to partner with was largely a function of the skills they came with as well as the resources of the organization they

were working within. We identified three main factors that determined who filled out the Data Ecosystem Designer's team:

- **Their Skills:** Depending on the skills of the Data Ecosystem Designer themselves, they'll need to hire or partner with folks who complement them. For example, a Data Ecosystem Designer coming in with strong Data Governance skills may want to hire a technology product manager to round out their skill set.
- **Their Institution:** The Data Ecosystem Designer may find that certain skillsets are more or less available in their home institution. In our interviews, participants drew a distinction between institutions that have technology expertise in-house and those who tend to have less technology expertise in-house. For example, Data Ecosystem Designers that hailed from technology companies, like IBM's social impact team or Microsoft's AI for Good Lab, took technology and product expertise for granted within their institution. However, they recognized they would need to partner with trusted individuals in the local context to succeed. Inversely, Data Ecosystem Designers who had developed a digital product at their nonprofit institutions took for granted that they had built extensive trust with their direct stakeholders and had thought through the governance of their data, but needed technologists and partnership expertise to scale the product. Some other researchers have noted this institutional need and have suggested solutions to align skills, such as having nonprofits partner with social enterprises to create appropriate digital goods (Álvarez-González, 2023)
- **Their Experience in the Local Context:** One of the skills that we identified as crucial for a Data Ecosystem Designer is knowledge of the local context and trust from the ecosystem they're working in. Our interviewees all emphasized that no matter what the flavor of Data Ecosystem Designer and no matter what their organization type, they had to have knowledge of the local context. Interviewees felt that this was important for ensuring that the product met the actual needs of local users and for building trust with the partners they'd need to succeed. Sometimes a Data Ecosystem Designer is hired who has experience in the local context of the product, such as when Ushahidi founders hired local talent to lead the project. Other times the Data Ecosystem Designer is building a solution to a problem they understand themselves because they are embedded in the local context, such as when the Flying Labs Kathmandu built drone solutions for their own village's needs in aerial mapping.

Conclusion: A Path Forward for Data Ecosystem Designers

We feel strongly that formalizing this role is critical for more and better digital public goods to exist in service of the greater good. We also feel that well-trained Data Ecosystem Designers play a critical role in ensuring that all digital public goods are built responsibly and with the highest standards of data and AI ethics in mind. Therefore we have three recommendations for the social sector on engendering this role:

1. **Organizations should acknowledge this role:** Almost all folks who have played the role of Data Ecosystem Designer have not had that title. They have been promoted from a Programs Director or a CTO to carry out their functions. One interviewee in our workshop exclaimed “When I read this description, I finally thought ‘oh, THIS is what my role is called!’” By raising awareness of this role and the skills needed to carry it out successfully, organizations building digital public goods can begin planning for it and hiring for it.
2. **Funders should fund this role:** Many philanthropic funders have recognized the need for data scientists and AI researchers in nonprofit digital technology creation and have funded organizations to hire them. A natural next step would be to fund the role of the Data Ecosystem Designer. This funding could come in the form of direct funding for a person to serve in the role or through supporting fellowship programs, like Schmidt Futures’s Technologists for Global Transformation program, that place people with these skills in nonprofits.
3. **Programs should train this role:** Just as data science programs arose over the last decade to round out the skillsets of computer scientists and statisticians holding the title, universities and certification programs should launch programs to train Data Ecosystem Designers. These programs could live at the intersection of technology, policy, and product design programs.

With these three shifts, we can start building capacity in the social sector to bring digital products from the pilot stage through to the level of scale we need for digital public goods.

Acknowledgments

We extend our thanks to the generosity of our interviewees and workshop participants below. We could not have done this work without their insights and experience, and we owe them a debt of gratitude for their work building and scaling digital public goods.

Davis Adieno, Global Partnership for Sustainable Development Data
Aman Ahuja, Fenris Technologies
Caitlin Augustin, DataKind
Zameer Brey, Bill & Melinda Gates Foundation
Shanna Crumley, Mastercard Center for Inclusive Growth
Jon Furr, Strada Education Network
Matt Gee, Brighthive
Brigitte Hoyer Gosselink, Google.org
Elizabeth Grossman, Microsoft Cities
Mark Hansen, Columbia School of Journalism
Pritika Hingorani, Artha Global
Angela Oduor Lungati, Ushahidi
Bilal Mateen, Digital Square
Juan Mateos-Garcia, Google DeepMind
Patrick Meier, Ushahidi
Laura Merson, ISARIC
Charlene Migwe, East and Southern Africa Development Gateway
Oladimeji Mudele, Harvard University
Josh Nesbit, Medic Mobile
Jennifer Oldfield, Global Partnership for Sustainable Development Data
Anupama Shekhar, Microsoft Philanthropies

Case Study: Medic

Overview: Medic was created to provide support to Community Health Workers (CHWs), a group of local health caregivers that are tasked with doing 90% of what a physician would do with a fraction of the resources. The digital software Medic created eventually became the Community Health Toolkit (CHT), a digital public good that provides decision support to CHWs, as well as providing visibility into the state of community healthcare. The goal of the software was nothing less than to transform healthcare at scale, so data ecosystem design was core to the project from the start.

Josh Nesbit, founder and former Executive Director of Medic, shared his experience as a data ecosystem designer.

Data Ecosystem Design At Work: Medic was a rare organization in that it set out to build software for the social sector as part of its mission. “We weren’t a service delivery organization that had a software project, we were a software organization,” said Nesbit. As a result, the company prioritized many of the data ecosystem design skills. One of their core focuses was on

partnerships, right from the start. “We were obsessed about scaling pathways and partnerships from the start. We tried almost everything in 10 years - ‘What’s it like for a partner to have this capability? What’s it like to ask the government to fund this budget item?’”, recounted Nesbit. The founding team recognized the importance of an intentional strategy of experimentation, an appetite for risk, and partnership skills at the outset. When asked why partnership skills mattered to them, Nesbit replied “You are operating in a totally dysfunctional market. It is not a market where, if you build the thing that helps health workers the most or that they love the most, that it’s going to scale. No way.” He added, “You have to work the politics [within the data ecosystem] so your solution helps more people.” Their doggedness in pursuing partnerships and evangelizing for their cause resulted in quickly finding a sweet spot for their tools with Ministries of Health. They also recognized through this exercise of constant testing with partners which aspects of their tool could be scaled out through open source. The resulting product was released as the Community Health Toolkit, a tool that now serves over 100K people in Kenya alone and is the backbone for many national health systems today.

Where Data Ecosystem Design Could Have Helped More: Though Medic’s approach to data ecosystem design was quite intentional, there were still areas they wished they’d had more time to invest in. For example, they wished they’d been stronger on product development and data governance up front. “We were learning a lot of that on the job,” said Nesbit. “[If we’d had those skills], we’d have been able to be ahead of the curve. Without those skills, our imaginations were under-resourced in what we could do [with the product and with data governance.]” Nesbit also offered that, while they were working on all of the skills listed for a data ecosystem designer as a leadership team, the day-to-day realities of running the company consumed them as well. Organization and team building distracted from more time to work on the product. “You need to specialize in generalization,” he shared, laughing. Medic’s challenges are shared by many organizations who don’t have the time, especially in a startup environment, to invest as much time as they want into data ecosystem design, as daily requirements get in the way.

Recommendations to Future Funders: When asked what recommendations Medic’s team had for others in the field facing a data ecosystem design challenge, Nesbit emphasized the need for collectives and ecosystems of supporting players. “Expecting digital public goods to survive and succeed in our current political economy, on their own…is insane. It’s just not gonna work, basically. The only way that it works is if these products have an ecosystem and a movement around them. We had this at Medic…we had a home within a network of organizations that are going to get through good and hard times together.” Josh’s thoughts point to a world where funders fund supportive environments of related players and support organizations, instead of just funding individual capacities at these institutions. This suggestion is a fresh way for funders to potentially enable a collective of actors and speaks to the design lessons of this report - funders especially need to engage in ecosystem support for digital public goods to succeed.

Case Study: Ushahidi

Overview: Ushahidi is an open source platform aimed at centering citizen voices in critical conversations around social justice and humanitarian issues. Their main goal is to empower disenfranchised communities by providing access to a data platform that allows them to make their voices heard and aggregate the needs and viewpoints of the community. Overall, they want to ensure that people everywhere can easily gather data and generate insights that help tackle issues that matter most to them. By providing the software that facilitates this data-driven advocacy, Ushahidi hopes to drive positive change in line with the needs expressed by local communities themselves.

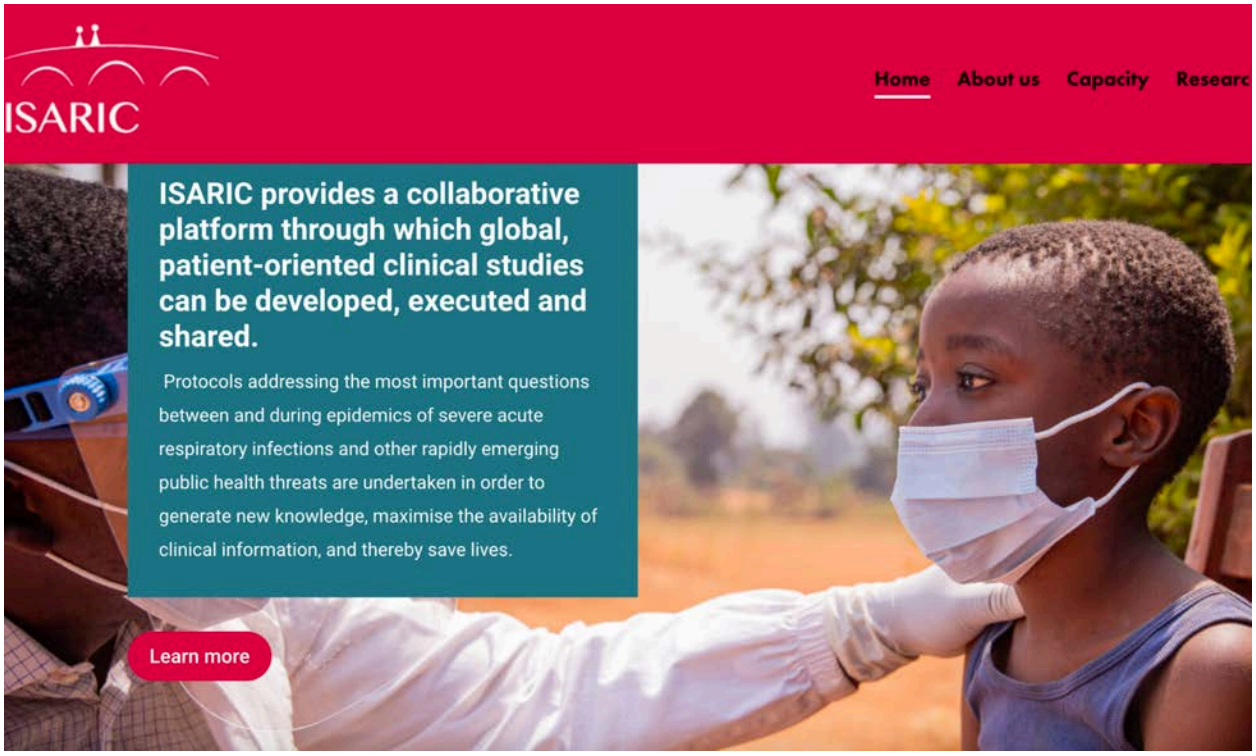
Angela Oduor Lungati, Executive Director of Ushahidi, talked with us about her experiences building this digital public good from the ground up.

Data Ecosystem Design At Work: In the course of developing and operating the Ushahidi platform, several key data ecosystem design skills proved invaluable. First off, Ushahidi started with core technology skills that set them off on the right foot, but they quickly realized the need for product development skills. “We first built a platform, it was able to scale, and people were able to use it. But as time went by, it became very clear that we have to productize it - we have to take into account how people are interacting with the tool, and how to make their work much easier,” shared Oduor. They shifted to investing in product development early on as a result. In addition, the team used deep community engagement and local context understanding in their work. The need for that skillset grew somewhat organically early on though, as Oduor recounts after seeing Ushahidi deployed internationally by communities in need. “Ushahidi was used during a problem in the Kenyan election, then people in Russia picked it up for reporting during the Russian fires. We knew there was only so much our tech can do on its own - it has to be appropriate for the context in which it's being used. And the only way you can understand or get to that level of appropriateness is by putting yourself in the shoes of the people who are going to be using it, whether it's the ones sharing the data, the ones who are digesting the insights, or anyone who deeply understands what the local issues are." Prioritizing local context awareness was responsible for the tool’s early success and has carried through to today.

Where Data Ecosystem Design Could Have Helped More: Though the team turned their software into a product early on, Oduor felt more formal product development capabilities throughout their process could have allowed them to improve and iterate on the platform much faster. At some point they needed to rewrite the code for the system after many successful deployments. “The tool would get deployed, and this group would say, ‘oh, we need this feature’. Okay, we’ll patch it in for them, then we’ll patch another feature for this group. The software became very clunky. The way we went about rewriting the product took years, and I think we lost people somewhere along the way,” Oduor shared.

Recommendations to Future Funders: Based on Ushahidi's experience, Oduor recommends that funders prioritize supporting nonprofits to build data ecosystem design skills. As emerging technologies like AI continue rapidly evolving, it is critical that civil society has a "seat at the table" to influence governance and policy. “Look at who's guiding the decisions around policy - It's big tech. Yet we know for a fact that a lot of the data that's being used to train AI models is coming from people like us and the communities that we work with. If we're talking about building responsible tech ecosystems, then we also need to have skin in the game. And the only way we can do that is by actually having these skill sets available to us in the nonprofit sector so that we can actually make a difference.” Oduor’s call for more representation by international nonprofits around tech policy decisions resonates with us. We feel this imbalance can only be addressed if funders fund more digital skills capacity to these nonprofits such that they can keep up with the global market and offer alternatives designed by, and for, the communities affected.

Case Study: ISARIC



Overview: ISARIC, based at the University of Oxford, focuses on research preparedness and response to epidemics and pandemics. They are a global federation of clinical research networks, providing a proficient, coordinated, and agile research response to outbreak-prone infectious diseases. They were founded in 2011 as a hub for coordination between scientists, but the COVID-19 pandemic in 2020 presented new challenges for standardizing and analyzing clinical data. We spoke with Laura Merson, Executive Director of ISARIC, about her experience in data ecosystem design.

Data Ecosystem Design at Work: When COVID-19 hit the world, Merson recognized the need for an ecosystem that could integrate and harness data from diverse global sources. “Because we already had tools we’d built for influenza,…we could quickly pivot those tools to enable collection of standardized data that would build an evidence-base rapidly. We developed data systems so that clinicians around the world had somewhere to put the data, using standardized data models and controlled terminologies.” This data store required a high-level of data governance to ensure data was stored correctly, accessed appropriately, and secured. Aside from the technological requirements of data governance and product design, the project of course also required Merson to design for the political needs and structural realities of her ecosystem. “Competing interests posed a challenge to what we were trying to do. There was a

clear global health need, but it required people to take a lot of risks and spend a lot of their time outside of their careers and interests [to take part in ISARIC]. So a key part of our work was understanding those priorities and then enabling an ecosystem that allowed each person, institution, or public health agency to slot into the part of the project that met their interests." ISARIC's empathy for all stakeholders allowed them to create an environment that worked for everyone.

When asked what other data ecosystem skills Merson used, she credited her networking abilities and collegiality, noting that in a chaotic environment with a lot at stake, it was critical for her to be able to work with others. She felt she succeeded because she "could identify things that other people can do and enable them to do that." "I can't code, like I couldn't write an R script to save my life…but I found excellent people to do some of the key things that take a lot of time. I have the nicest data manager you have ever met in your life, and the funniest." Hiring people was not enough on its own though, she had to make an inclusive environment for all stakeholders. "I did a lot of post-midnight calls with Australia or Latin America", she said. "We made friends, and we built trust really easily. So, the skill [a data ecosystem designer needs] would be 'creating personal connections'. And part of that is being very open to bringing people on board and making them a part of the initiative, incorporating their thoughts and ideas and interests into where it's going to go. I think openness and inclusiveness are more important attributes than any tools."

Where Data Ecosystem Design Could Have Helped More: ISARIC had learned a fair share of data ecosystem design skills from their decade of work serving scientists with epidemiological tools. Still, when COVID-19 hit, new challenges arose and Merson stated that they could have benefited from even more partnership management and ecosystem view. Specifically, she felt that they could have achieved greater scale if there had been a mechanism for them to convene with other similar efforts to understand how everyone could play the most effective role possible. "If we'd had better global alignment, we would've pre-negotiated with other institutions to understand what was the best way to do things in a way that could scale the impact across all of our different initiatives," she said. Merson also felt that they could have benefited from more knowledge of the ecosystem itself up front, such as understanding who was onboarding with them and which important players weren't yet involved in their work. "We were quite alone up until 2020," she said, until new entrants started flooding their platform. "Fortunately, we did have WHO quite on board early in the pandemic, but due to legitimate reasons, it broke off and they started their own initiative, which essentially competed with us. Had we been able to foresee the reasons for that, and repercussions of it, then we'd have been more prepared to avoid it."

Recommendations for Future Funders: Merson's experience led her to one main recommendation for funders – recognize the need for this role that she and others have been playing informally. "We need to find a way to communicate the fact that having a data ecosystem designer that has that oversight of all the different systems is going to save

over-investing in the many different parts that are, frankly, going to waste a lot of time heading off in directions that are not in line with an overall strategy.” She felt that funders would actually save time and money while increasing their impact if they funded this role for major digital public goods projects. Merson closed with an analogy on this need: “If you don't have the architects on board first to design the building and you just send a whole bunch of builders in there and say ‘go for it’, that’s going to fail. We need someone to give direction and leadership to all of the critical individual pieces that are trying to put something together themselves.”

Case Study: Epiverse

Overview: The Epiverse project represents a groundbreaking effort in the realm of digital public goods, focusing on enhancing the global scientific community's ability to collaborate through an open, standardized suite of tools. This initiative seeks to address the fragmented landscape of scientific tooling, which is characterized by duplicative, unsupported, and poorly documented software. The goal of the project is to foster a cohesive, supportive, and open-source ecosystem of epidemiological tools that can create better health outcomes during pandemics, global health risks, and day-to-day life.

We spoke with Danil Mikhailov (Executive Director of data.org), Zulma M. Cucunubá (Assistant Professor of Infectious Disease Epidemiology, Pontificia Universidad Javeriana), and Catalina González-Uribe (Director of Internationalization at the Vice Presidency of Research and Creation and Associate Professor at the School of Medicine) about this project. data.org hosts the Epiverse project and Zulma and Catalina lead its efforts in Colombia.

Data Ecosystem Design At Work: The primary challenge for Epiverse was the prevailing culture of tool creation within the scientific community. Many scientists developed tools solely for personal use, without consideration for broader applicability or user support. “Scientists don't care about product design, such as how well supported the tools are, how well designed they are. So often they just build a tool for themselves to use, which means they won't be useful for

other people," said Mikhailov. This approach led to a proliferation of isolated tools, lacking in documentation, community support, and standardization. The Epiverse team knew that, in order to change behavior in the scientific community, they would have to lean into their Data Ecosystem design skills of humility and partnership building so that they would be trusted by all participants. Mikhailov recounted this plan: "From the beginning, we thought that to be successful we needed to not be seen by other teams as just a competitor, just another team building their own tools. So we adopted an ethos of not building something if it already exists. Our software developers help existing tools from other communities be improved, better supported, and better documented. We are trying to signal to the rest of the community that we are not just another team competing with them." That thoughtfulness about positioning their work as a service to the rest of the ecosystem was critical to their success.

When it came to running Epiverse in Colombia, Cucunuba noted the importance of using their existing relationships and experiences to improve product design and support stakeholders. "We designed the project with the idea of co-developing tools based on our existing relationships with decision makers and stakeholders," she said. "[We wanted to] build on everything we'd learned from the pandemic, because we knew [from that experience] that some tools and features were missing. We wanted to provide a more reliable set of tools and a more transparent way of collaborating to build and to democratize the tools for everyone." Without those relationships and that experience, the team would not have been able to position itself as a knowledgeable and trusted support to the ecosystem.

Another area where data ecosystem design came in handy was in creating funding for this work. The ecosystem suffered from a lack of institutional and financial support, further exacerbating the fragmentation and inefficiency of these tools. "We knew we needed to design a funding program that allowed us to fund not just our own tools, but also the development of others. To do that, we created a consortium of funders," said Mikhailov. He described how that partnership and business mindset resulted in their consortium of funders, which provided a diversified financial support system for tool development within the ecosystem. Mikhailov commented on how foundational this coalition building work was. "In fact, the work started even before the program. I had to negotiate with funders and a lot of that involved persuading them [to fund our work] and I had to champion this idea of creating a community behind a single set of open-source tools. So that's one example of ecosystem design skills: the persuasion, the championing, the negotiation with indirect stakeholders through diplomacy skills."

Gonzalez-Uribe emphasized the data strategy skills needed, combined with empathy for all stakeholders, to create useful tools in the scientific community in Colombia. "I'm always concerned about the digital divide. I did a lot of health research with indigenous communities in the Amazon peasant communities, and the reports at the end all involved tables and models. None of the questions that they were asking were in there. I said, 'I want to learn the logic, the way of thinking from the other side.' I learned how to go into the head of those who are thinking

about how to model and how to think about the data. What do we measure? Why do we measure, is this a valid way of measuring it?" She felt that her ability to translate between the data strategists and the people her tools were serving were critical for any utility to come out of their tools.

Epiverse is a prime example of ecosystem design, focusing not just on the technical aspects of tool development but also on changing the behavioral patterns of scientists, funders, and institutions through humility and diplomacy.

Where Data Ecosystem Design Could Have Helped More: The Epiverse team planned for Data Ecosystem design from the start. However, they still felt there could have been some improvements in their approach. At a meta-level, Mikhailov felt they could have articulated the value of Data Ecosystem design to partners that didn't perceive its utility at first. "The partners didn't see the value of data.org as an organization or my role as this Data Ecosystem Designer. They thought it was bureaucracy. They thought I was there to get their funding for them in the way the admin team in their university would help them with funding. If I could go back, I would have made clear from the beginning that we're not just securing funding, we're negotiating a network of funders," he shared.

The team also shared that they wished the governments and universities they'd served had taken an ecosystem design approach from the start, instead of the silo-ed or piecemeal approach they ended up taking. "[During the pandemic, the government] tended to consult experts in silos. First, they talked to the epidemiologists and they produced a plan for lockdowns to stop transmission. Then, a few months later, the economy started crashing, so they forgot the epidemiologists and talked to economists. And the economists say, 'of course the economy is crashing, so we need to reopen'. But then the second wave spikes straight away because they reopened too quickly. So the experts are not talking to each other because they don't have a common language. If we had Data Ecosystem designers defined already, then you would've had a group of interdisciplinary people ready before the pandemic."

Recommendations for Future Funders: The team felt that funders have an exciting and important role to play in supporting digital public goods through a data ecosystem approach. Mikhailov felt that funders would naturally want to fund Data Ecosystem Designers because philanthropists care about impact and it is exactly this role that will make sure their impact is scaled and maintained. "Without this role, you can run programs and maybe get a few outputs, but that's it. The program ends. Organizations should embed the need for this role in their MEL strategy because it's how you'll get sustainability of impact." The team in Colombia also pointed out that current funding strategies for multidisciplinary projects could actually be counterproductive to data ecosystem design work. Cucunuba recounted the way this happened in their region. "At the start of the pandemic, The Ministry of Science (in Colombia) made a call

for anyone to apply for funding [for pandemic tools]. The organizations who were qualified were so busy responding to the pandemic that we didn't have time to apply for the funding. So you may end up in a situation where [your applicants are] people who are really technologically talented and want to do this type of thing, but have no experience or local context so they start learning from scratch. We should fund some new and disruptive teams, of course, but at the same time keep the teams that are already there and utilize their relationships and local context.” They recommended that funders use their funding to help free up the time of existing actors to connect, align, and work together as an ecosystem instead of chasing project-bound funding outputs. “Bring people together. You [funders] already have a network that is so powerful.”

Appendix

The table below shows the differences in digital public goods and commercial products with an example:

	Commercial Products	Digital Public Goods	Example: Spotify	Example: Hospital Bed Inventory Model
Impact Statement	Maximize profit (no social impact) No coordination of users needed	Maximize social impact Users often have to coordinate to achieve the goal	Sell subscriptions	Reduce negative COVID-19 outcomes by getting more critical patients to hospital beds
Direct Stakeholders	Paying customers. They often have identical digital infrastructures (e.g. iPhone)	The users of the digital public good. They often have radically different digital infrastructures and operating procedures	Customers who listen to music	Hospitals that need to coordinate beds. They may have different data systems / procedures.
Indirect Stakeholders	None	The organizations and constituents affected by the product or involved in its success. Government agencies and the public are almost always indirect stakeholders	None	Ministries of health that coordinate resources from the government The public that benefit from better healthcare
Datasets	All data is accessible to the company. They either create their own data or purchase the data needed	Almost always require multiple datasets and data providers. Data governance is critical to access, secure, and uphold the integrity of stakeholders' datasets	Music and user databases	This tool might require multiple databases of hospital inventories, in different formats, requiring different legal considerations.
Technical capacity	The company building the product provides the technical capacity.	Technical capacity may vary over the lifetime of a product. It may start at a nonprofit who pilots the good, then move to a government for maintenance. Oftentimes open source communities contribute to and maintain the tool.	Software engineers and data scientists on staff at Spotify	Researchers at a university working with hospitals to build this tool.
Funders	At the outset, VCs. Over time, paying customers.	At the outset, foundations and grants. Because users don't often pay for the tool, alternative revenue streams need to be devised.	Spotify subscribers	Potentially hospitals, but may also come from government, foundations, or others with commitment to the impact statement

References

Álvarez-González, L.I., Barroso-Méndez, M.J., Galera-Casquet, C. et al. The impact of technologies on society through NPO-social enterprise value co-creation. Humanit Soc Sci Commun 10, 711 (2023). https://doi.org/10.1057/s41599-023-02145-4

GitHub. (2023). Open Source Software in the Social Sector: Examining Barriers, Success and Opportunities. GitHub. https://socialimpact.github.com/assets/img/research/GitHub_tCF_OSSInSocialSector_FINAL_updated.pdf

Gottschalk, L., Zielinskie, G. (2023). Rising Equitable Community Data Ecosystems (RECoDE) data.org. https://data.org/reports/recode-report/

Hull, C. (2006). Innovation in non-profit and for-profit organizations: Visionary, strategic, and financial considerations. Journal of Change Management, 6(1), 53-65. https://www.researchgate.net/publication/247505468_Innovation_in_non-profit_and_for-profit_organizations_Visionary_strategic_and_financial_considerations

Jong, CLK., Ganzaroli, A. (2023). Managing Digital Transformation for Social Good in Non-Profit Organizations: The Case of The Felix Project Zeroing Hunger in London, Voluntas (2023) https://doi.org/10.1007/s11266-023-00597-5

Mikhailov, D., Zielinskie, G., Hewitt, P., & Stewart, U. (2023). Accelerate Aspirations: Moving Together to Achieve Systems Change. data.org. https://data.org/reports/accelerate/

Nicholson, B., Nielsen, P., Sahay, S., & Sæbø, J. I. (2022). Digital public goods platforms for development: The challenge of scaling. The Information Society, 38(5), 364–376. https://doi.org/10.1080/01972243.2022.2105999

Saviano, J., Rodriguez S., Morera, F. (2022). Financing Digital Public Infrastructure: Approaches to Sustain Digital Transformation. New America. https://www.newamerica.org/digital-impact-governance-initiative/reports/financing-digital-public-infrastructure/current-dpi-funding-models