

SCIENCE DATA SHARING: APPLYING A DISRUPTIVE TECHNOLOGY
PLATFORM BUSINESS MODEL

BY

ALAN EDWARDS

A dissertation submitted to City University of Seattle
in partial fulfillment of the requirements for the degree of

DOCTOR OF BUSINESS ADMINISTRATION

2022

Author Note

There are no known conflicts of interest. The conclusions and recommendations of this dissertation will ultimately be implemented within a FAIR compliant open data science platform.

© Copyright by Alan Edwards 2022

All Rights Reserved

DEDICATION

To my wife, Elizabeth. Always supportive.

TABLE OF CONTENTS

DEDICATION	iv
TABLE OF CONTENTS.....	v
LIST OF FIGURES	ix
ABSTRACT.....	x
CHAPTER 1: INTRODUCTION TO THE STUDY.....	1
Study Background.....	1
Current State of the Field in Which the Problem Exists	2
Historical Background.....	3
Problem Statement	4
The Extent of the Problem.....	5
Audience.....	6
Market Disruption Opportunity.....	6
Purpose of the Study	7
Methodology and Research Design Overview	8
Key Propositions.....	9
Research Questions.....	10
Study Limitations.....	11
Study Delimitations	12
Definitions of Key Terms	12
Summary.....	14
CHAPTER 2: LITERATURE REVIEW	17
Rationale for a Science-Data Sharing Business Model	17
Characteristics of a Technology Platform Business Model.....	20
Release of Excess Value.....	20
Providing Greater Reach	22
Platforms as Intermediaries and the Rise of the Sharing Economy	23

Platforms Positioned as a Utility or as Predatory	24
Network Effects, Supply Side Versus Demand Side	25
The Curious Anomaly of Price.....	26
Impact of Absence of Price Signals on Market Participants	28
Managing the Technology Platform	29
Role and Responsibilities of Technology Platforms	30
Unique Roles Within Platform Businesses	31
Content Providers as Social Entrepreneurs	32
Growth Strategies and Market Barriers for Technology Platforms	34
Ability to Scale Rapidly	34
The Evolving Nature of Web 3.0	35
Network Effects.....	36
Multisided Versus Two-Sided Platforms	38
Characteristics of Markets That Lend Themselves to a Technology Platform Strategy	41
The Business Transformation of IBM.....	41
The Business Transformation of Apple.....	42
Motivations for Sharing Data	43
Government and Academic Solutions or Lack Thereof.....	44
Role of Platform Technology in CS.....	46
The Isolation of Science Data and the Long-Tail Effect.....	48
Potential Market Disruption Strategies for a CS Technology Platform	50
Market Disruption	50
Science-Data Sharing in Underdeveloped Countries	52
Sustainable Business Practices	54
Platform Market Entry.....	55
Dethroning a Platform Incumbent.....	58
Market Failure: The Chicken or the Egg Conundrum	59
Monetary Subsidies	60
Product Feature Subsidies	61
User Sequencing.....	61

The Core Transaction.....	62
Ecosystem Binding.....	62
Value Ecosystem	63
Importance of Simplicity of Core Transaction.....	64
Complexity in Search Matching.....	65
Summary	66
Deficiency in the Market for the Sharing of Science Data.....	66
Is This New Business Model “Pandemic-Proof”?	67
CHAPTER 3: METHODOLOGY	69
Research Method	69
Research Design.....	70
Instruments.....	72
Participants.....	74
Data Analysis Methods.....	76
Limitations	77
Delimitations.....	78
Trustworthiness.....	79
Summary.....	80
CHAPTER 4: FINDINGS	81
Presentation of Findings	82
Thematic Analysis	85
Summary.....	86
CHAPTER 5: CONCLUSIONS AND DISCUSSION.....	88
Discussion of Findings and Conclusions	89
Reflection on the Questionnaire Findings.....	89
Barriers That Impede Science Data from Being Shared with the Global Science Community.....	94
Critical Attributes of a Market-Driven Technology Platform Business Model.....	95
Application of Findings and Conclusions to the Problem Statement	99

Application to Business	101
Recommendations for Action	101
Establish Frictionless Market Participation.....	102
Focus on the Core Transaction.....	102
Acknowledge Emerging Web 3.0 Concepts.....	102
Focus on Unmet Needs.....	103
Solve the “Chicken or the Egg” Conundrum	103
Platform Organizational Management	104
Viral Marketing and Platform Valuation.....	105
Facilitate Market Entrepreneurship	105
Recommendations for Further Research.....	106
Further expansion of platform concepts to facilitate data sharing	106
Survey Participant Ideas	106
Data Set Copyright	106
Decentralized Finance and Global Payments	107
Concluding Statement.....	107
REFERENCES	109
APPENDIX A Survey Questions.....	119
APPENDIX B Venn diagram of Themes by Category.....	141

LIST OF FIGURES

<i>Figure 1.</i> Number of Social Media Users by Platform.....	37
<i>Figure 2.</i> Venn diagram of Themes by Category	86

ABSTRACT

This research investigated the potential to enhance global science data sharing through market-driven strategies utilizing technology platform-based business models compared to the existing fiat strategy characterized by government-funded programs and highly fragmented information systems. With less than 1% of science data shared globally, and some estimates of data collection time–cost valued at USD billions per year, there is significant social value to be released in addressing this issue.

The purpose of this research is to identify whether a paradigm shift in business model is required to enable the release of excess value from science data gathering by facilitating data re-use; and incentivize increased market participation by currently marginalized groups that gather science data. A literature review was undertaken to understand (a) the characteristics of successful platform business models and (b) whether those lessons can be applied to science-data sharing to formulate a market-driven business model appropriate to satisfying the un-met needs of the global and frequently marginalized science community.

Participation in international science industry discussions, review of science data gathering grant requirements from USA and European institutions, and extensive literature review formed the basis of a questionnaire that was distributed to CEO of citizen science (CS) organizations globally who were used as a proxy for the broader science community to undertake a survey of unmet needs in both developed and underdeveloped countries. Responses were analyzed to determine the pattern of unmet needs from both legacy business models for sharing data, and from a hypothetical platform business model that had the attributes of successful market-driven businesses.

Evidence pointed towards a clear market failure to satisfy the unmet needs of a broad CS community particularly those residing in underdeveloped countries, and a significant opportunity to release excess value of global benefit. The research has recommended that a paradigm shift in business model towards that of an open platform-technology be established that rewards science data contributors based on free-market principles. i.e., centralized standards and data custodianship, but with distributed ownership, and rewards (both monetary and through recognition) based on the value that the market places on published contributions to the platform.

CHAPTER 1: INTRODUCTION TO THE STUDY

This research project pertained to whether global science data sharing could incrementally improve by establishing a purposeful technology platform to facilitate the sharing of science data. This platform would reflect successful strategies of commercial market leaders who have created excess value to the marketplace, and whose valuations have soared as a consequence of establishing a business model paradigm shift through applying a platform technology strategy to their chosen market on a global basis. The world's 10 largest public companies by valuation, as of mid-2020, were overwhelmingly platform-technology related (in descending order): Apple, Microsoft, Alphabet (i.e., Google), Amazon, Berkshire Hathaway (via its primary investments), Facebook (renamed Meta Platforms in 2021), Alibaba, JPMorgan Chase, Tencent, and Visa. This list radically differed from the situation just a decade prior, when hard-asset infrastructure companies, such as ExxonMobil, dominated the valuation hierarchy. In early 2021, the dominance of platform technology companies continued with the rise of Tesla, a software platform masquerading as a car manufacturer; and Nvidia, which is primarily a gaming platform and which also manufactures graphics chips to support other platform technologies.

Study Background

The aim of this study was to analyze (a) why the platform technology model has become the dominant business model, (b) which of those characteristics sustain their market position, and (c) what makes those characteristics so valuable, and (d) whether these principles could be successfully applied to satisfy the science data sharing needs of the global research community. A literature review was conducted to understand the

characteristics of industries that lend themselves to a technology platform business model and then an assessment was made as to whether the science-data sharing sector had these characteristics. Finally, the most important question on this topic was addressed: What facets of the technology platform business model can be applied to enhance science data sharing? This research considered the subsector of citizen science (CS) as a proxy for the more general science sector to define a business model that uses platform technology to facilitate science-data sharing. Through that platform, the aim of this study was to expand the science-data-sharing market by empowering entrepreneurial data scientists to contribute to solving social justice, environmental, and community health issues.

Accordingly, this study concerned (a) identifying barriers to establishing a technology platform for efficient science-data sharing, and (b) reviewing the potential reward or recognition mechanisms that can incrementally motivate the global community of CS data content owners to share their data in the public domain. The community benefits of incrementally increasing the shared data, which would constitute less than 1% of the total science data collected (Reichman et al., 2011; Wallis et al., 2013), were discussed and quantified by Bonney et al. (2016).

Current State of the Field in Which the Problem Exists

CS has diverse participants, including hobbyists, students, professional scientists, and geographically dispersed teams; it is a global and diverse field that involves the environment, ecology and epidemiology, and comprises numerous specialized subfields (e.g., water quality, air quality, ornithology, and entomology) and further subfields. When funded, research and design (R&D) projects are almost exclusively funded directly by government grants or indirectly via academic institutions. According to the National

Science Foundation (as cited by Jahnke, 2015), 29% of U.S. federal R&D money went to universities, 29% to industry, 29% to researchers working for federal agencies, and about 10% to federally funded labs operated by private contractors for the 2013 fiscal year.

Generally, science funding has been on a slow decline within the USA, however in 2018 USA budget funding had a windfall USD \$1.3-trillion allocated, although this level is not anticipated to persist (Morello & Guglielmi, 2018). Many CS grants have strict geographic boundaries aligned with the grantee institution's political boundaries.

Developed countries have a greater ability to fund CS projects (Nuñez & Pauchard, 2011). The consequence of this highly disjointed global industry is that its data repositories for the CS data collected are also disjointed, as identified by a general inability to obtain data in an efficient manner. Frequently, there is no public access to the data stored, and minimal ability to even identify what data repositories exist. Such disjointedness is remarkable, given CS data collection is labor intensive, characterized by volunteer labor, and a huge labor cost component of billions of USD per year in in-kind economic worth (Theobald et al., 2015).

Historical Background

Some of the world's longest-running CS records are from Japan; for instance, the timing of cherry blossoms have been recorded in Kyoto for 1,200 years and have even been used in climate reconstructions (Kobori et al., 2016). In the United States, the National Audubon Society surveyed wintering bird populations in 1900. CS was the norm prior to the advent of professional scientists. In the last two decades, the number of CS projects has dramatically increased, fueled in part by the use of smartphones with the ability to enter data, take photographs, and set accurate GPS positions. Theobald et al.

(2015) analyzed 388 English-speaking citizen projects, 89% of which were in the United States and the others in the United Kingdom, and estimated 1.3 million participants contributed up to \$2.5 billion in-kind annually; surprisingly, they found only 12% of the projects surveyed provided their data to peer-reviewed scientific journals. For professional scientists, the use of a network of volunteer citizen scientists presents the opportunity to gather data that would otherwise be cost and time prohibitive.

Problem Statement

A general problem for this research was less than 1% of collected science data are shared with the public (Reichman et al., 2011; Wallis et al., 2013), which is a great loss to humanity and a waste of academic and community resources. Funding for CS data collection is overwhelmingly regional or localized and primarily dependent on government largesse. This limitation has resulted in diverse data repositories that are difficult to locate and access (Bonney et al., 2016), and this is a significant part of the problem. There is no dominant technology platform to facilitate environmental science-data sharing, despite the recognition that such facilities would advance environmental awareness (Wallis et al., 2013). Wallis et al. also stated CS data collection is a product of the developed world due to the lack of funding in the underdeveloped world, and this provides an insight into a potential solution to facilitate greater science data sharing. The marketplace has, so far, failed to establish a technology platform to address this problem of lack of data sharing (Costa, 2016) and the funding of data collection despite the benefits of community involvement in science (Bonney et al., 2016). Not all markets (in the sense that the science community is a market) lend themselves to a technology platform solution (Spulber, 2019); therefore, it was important to examine whether there is

potential to satisfy unmet needs for science-data sharing through establishing a platform that is market driven and relatively independent of government funding.

The specific problem is how to overcome these barriers to global science data sharing, and whether a platform strategy has the potential to satisfy the unmet needs of the CS community utilizing market forces to become commercially viable. An analysis to find a solution required conducting surveys with CS organizations that represented both developed and underdeveloped countries, and selectively following up with national CS coordinating bodies. The global CS community, epidemiologists, and the general science community will directly benefit from studying this problem, and the successful implementation of a solution will provide society with an intangible benefit from the release of significant excess value.

The Extent of the Problem

As Theobald et al. (2015) assessed, accurate information on the extent of the use of CS data is difficult to obtain, but not impossible; however, their analysis was delimited to projects within the United States and United Kingdom. Further, researching barriers to data sharing has been predominantly restricted to the developed world. As such, it could be confidently assumed these barriers are even greater in underdeveloped countries. The Kenya Data Protection Act of 2019 evidenced these barriers following implementation that has remained controversial due to its negative implications on research despite good intentions. Various researchers have emphasized that the benefits of using citizen scientists are highly significant (Bonney et al., 2016); however, published results of their data collection is negligible for a variety of reasons, including a lack of publication on data availability, difficulties accessing databases, and inconsistencies in data definitions

despite the significant time and monetary costs of data collection (De Filippo et al., 2020).

Audience

Initially, the audience included those who participated in collecting and sharing environmental science data, both within developed and underdeveloped countries, and formal science institutions, as well as the relatively informal citizen scientist network. There were other specialist audiences, including those seeking social justice in urban areas, those concerned with the survivability of species in remote locations, and epidemiologists studying community health issues in under-developed countries.

Market Disruption Opportunity

The question was whether a market disruption opportunity exists for CS projects to reduce their overwhelming reliance on government and other grant-based support that has resulted in fragmented fiat information systems. Would a market-based demand approach using a single platform as a utility, combined with new business models that leverage reward and recognition market forces, create this market disruption? Researchers need to look no further than the example of Facebook (i.e., Meta) to recognize that technology platforms in the new economy have the potential to create significant opportunities in an industry that previously did not exist. Facebook was originally designed to facilitate socializing between Harvard college students at a single campus (Parker et al., 2016) and now dominates a global market of social interaction commanding a market capitalization of approximately \$918 billion as of December 2021 (Nasdaq, n.d.).

Purpose of the Study

The purpose of this grounded theory study was to understand how global science-data sharing can improve by potentially using the successful platform technology business models of the world's most valuable companies; moreover, the purpose was to determine whether those models can be used for science-data sharing and made commercially viable without relying on government financial assistance, the traditional support mechanism for science-data sharing. The goal was to foster an entrepreneurial opportunity and market-driven incentive for the CS global community to share their data content with their peers and the public. This opportunity is particularly important for CS in underdeveloped countries that face greater barriers to market participation, but whose local data content is no less valuable—perhaps arguably more so due to paucity. The research objectives were as follows:

- Discuss the barriers that have impeded CS projects from sharing their labor-intensive data collection efforts.
- Determine whether the market niche of CS data lends itself to a global technology platform solution, considering the pressure toward cloud-based storage platforms from the formation of intermediaries that displace traditional science data storage custodians (Plantin et al., 2018).
- Assess whether a dedicated platform technology has the potential to be commercially viable (i.e., independent of government financial support).
- Explore strategies to solve the biggest hurdle of platforms—the “chicken or the egg” problem of creating a critical mass of data that will attract others to participate.

- Identify a mechanism to ensure the proposed platform is market driven (i.e., provide an ongoing financial reward and recognition mechanism for culturally diverse data contributors; Gerhardt, 2007) independent of their legacy funding (if any).

Methodology and Research Design Overview

This research project sought to facilitate transformative change using qualitative methodology with the grounded theory design to identify and reflect on the issues and arrive at a recommended solution strategy. The hypothesis was a market-based disruptive technology platform for facilitating sharing of science data can be self-sustaining in the marketplace. A premise of this dissertation was a market-based solution is more cost-effective than a government's 100%-funded solution. Interviews and questionnaires were conducted with representatives of significant CS organizations in various countries to arrive at their common unmet needs in pursuit of a sustainable commercial business model independent of government funding.

The U.S. National Institutes of Health (NIH) launched an initiative in early 2019 to acquire a cloud-based solution to store and share health data from all countries in Africa and has a working budget of approximately \$62 million over the initial 6 years, renewable. This budget includes providing a cloud-based solution budgeted at \$11.25 million over 5 years to comprehensively establish the data facility, and \$45 million to contribute to the data take-up commencing September 2021 from a number of suitably qualified Africa-based organizations (Glass & Tromberg, 2019). This initiative provided a benchmark on the cost of establishing a comprehensive, pan-continental, open science-data platform.

Key Propositions

The conceptual framework revolved around the following key propositions:

1. Only one or two organizations, at most, can thrive in a global market niche that is supported by a platform-technology supported business model (Moazed & Johnson, 2016).
2. Platform businesses require a different definition of what a business is than the traditional linear business definition; as a consequence, organization, management, and growth strategies for platforms significantly differ from that of the traditional firm (Parker et al., 2016; Rochet & Tirole, 2014; Spulber, 2019).
3. The network effect of viral marketing is significantly more cost effective than traditional marketing strategy to grow a platform technology business (Costa, 2016).
4. A fundamental characteristic of global markets is a market-based reward or recognition system (Sun & Zhu, 2013). This absence is notable in the research community as evident in the lack of sharing of science data.
5. The market has not provided equitable funding mechanisms to entrepreneurial environmental scientists to pursue projects optimized for satisfying unmet needs of the global science data community (Spulber, 2019; Sun & Zhu, 2013).
6. Traditional science funding has been localized or based on political boundaries, which has resulted in storing data collection efforts in local dedicated-use information systems that have a narrow focus and are not

usually designed for global open access—at a detriment to global science data sharing (Glass & Tromberg, 2019; Plantin et al., 2018).

Exploratory research of the literature was conducted to investigate these propositions. Further, the grounded theory method was used to collect and analyze data through interviews and questionnaires, ultimately addressing the hypotheses. Grounded theory was used to investigate these propositions in the specific case of environmental science data, focusing on CS individuals and organizations as a proxy for the broader scientific community. Key questions were asked of CS organizations via use of a survey and selected follow-up interviews leaning on appreciative inquiry techniques. A global approach was taken in the study, surveying CS organizations in both developed and underdeveloped countries to look for common themes. This approach helped formulate a theory that was validated through the establishment of a platform-technology business model designed to facilitate the publication and open-access sharing of global environmental science-data sets.

Research Questions

The research questions were:

1. What are the barriers that impede science data from being shared with the global science community?
2. What are the critical attributes of a market-driven platform technology business model whose mission is to facilitate global science-data sharing and satisfy unmet science community needs?

Study Limitations

Science-data sharing, specifically CS, is a very broad and diverse subject area with a global audience. A limitation of the study was gathering representative opinions that reflected this diversity in science topics and among different countries to represent developed and underdeveloped economies, given many underdeveloped countries discourage the sharing and publication of environmental data (e.g., Kenya Data Protection Act of 2019). Significant study limitations of concern were:

1. Participants (i.e., environmental data gathering groups) tend to be funded through government grants and donations that are contingent on satisfying the government's policy (e.g., investigation of clean water, habitat sustainability). This funding structure limited the pool of independent data gathering groups to interview.
2. The scope of the study was not limited to the developed world. Outreach to CS groups from underdeveloped countries was included, but was more difficult. One-on-one data gathering was restricted to interview questionnaires rather than face-to-face interviews, along with a restricted sample size, although Zoom-based interviews to participants in Africa were found to provide adequate data.
3. The analysis was restricted to countries highly proficient in the English language.
4. It was assumed, excluding funding constraints, barriers to sharing environmental science data among professional groups in the developed and

underdeveloped world are the same, although there are “politically contrived barriers” in some countries where the sharing of data is inhibited by law.

5. An issue to study was the potential to leverage market forces to achieve a better outcome in sharing science data than the traditional command strategy. A concern was the interview pool might not be representative of a global market, and participants might not have the insights to recognize the entrepreneurial opportunity before them; however, this was not found to be the case, and many entrepreneurial ideas were presented.

Study Delimitations

A boundary condition was the focus on scientists and citizen scientists in the field of environmental science. CS was used as a proxy for the broader field of science. A modest delimitation of the study was it was predominantly researched in the English-speaking world, although that is not an overwhelming delimitation for the science community; in some science fields, 90% of the publications are in English and the global science community has considerable English-language representation (Fung, 2008). Interviews were implemented with a principal management representative of a number of CS organizations. The questionnaire was limited to 25 questions.

Definitions of Key Terms

Action research is a type of research that seeks transformative change through the simultaneous process of taking action and doing research, which are linked together by critical reflection. The act of participating in action research may have a positive intent of fundamental change.

Appreciative inquiry differs from the traditional strengths, weakness, opportunities, and threats (SWOT) analysis and strengths, opportunities, aspirations, and results (SOAR) analysis; it attempts to ask questions in a positive manner with the aim of envisioning the future through cooperative acts and relationships (Cooperrider & McQuaid, 2012).

Citizen science includes science projects that often involve large numbers of trained volunteers, due to their sheer size or sampling requirements, to undertake data collection. These projects are usually led by one or more professional scientists.

Developed countries have a mature and sophisticated economy, with advanced technological infrastructure and diverse industrial and service sectors. They are also characterized by the protection of private property, whose value can be realized separate from the physical asset (de Soto, 2001). A characteristic of severely underdeveloped countries is the lack of access to capital due primarily to the absence of a formal property system that would enable citizens to monetize their assets.

Environmental science is a type of research primarily concerned with nonspecific data (i.e., measurements of parameters over time), which comprise the most common environmental data sets. The parameters can be widely diverse, although the units of measurement tend to be modest in variety, whereas specific data applications are limited in the types of data stored (e.g., images of a virus).

Grounded theory was the qualitative research methodology applied in this research focused on the sociological aspects of market participation, leading to a hypothesis on market behavior.

Hard assets are generally physical assets, such as pipelines, refineries, vehicle fleets, and land. *Soft assets* are relatively intangible assets, such as brand names, intellectual property, business processes, and cloud storage.

Platform technology is synonymous with *technology platform*.

Technology platform assists two parties to the transaction, tends not to own the assets that are conveyed, and leverages cloud-based technology to affect the transactions with parties belonging to strong social networks.

Viral marketing is a feature of networked social groups wherein the individuals assist in generating positive marketing communication messages and practicing viral marketing by resending messages they have received, thereby delivering content to a large society segment (Fouad, 2017). The message transfer growth rate tends to obey Metcalfe's power law that a network's value is proportional to the square of the number of users (Metcalfe & Boggs, 1976).

Web 2.0 refers to websites that emphasize user-generated content, ease of use, participatory culture, and interoperability for end users (DiNucci, 1999).

Web 3.0 is the autonomous integration of data and services that increases preexisting capabilities of Web services and the creation of new functionalities (Rudman & Bruwer, 2016). There is a growing expectation by the content provider for revenue sharing by the organization to whom they crowdsource content, and to increase societal responsibilities of the platform (Sun & Zhu, 2013).

Summary

Grounded theory was employed in this dissertation to understand how the target audience operated, what their unmet needs were, and how their unmet needs could be met more effectively, (Charmaz, 2014). There were modest elements where the causal and solution aspects were explored in a collaborative manner with selected thought-leader participants, and addressed the concerns of the marginalized (Marshall & Rossman, 2016). The use of this research technique has some facets of active research methodology. According to the literature review by Abdel-Fattah (2015), this similarity and overlap between grounded theory and active research lends itself to borrowing from each methodology's approach.

The first area of research identified the barriers to sharing science data with the broader community, and the second sought to explore practical solutions for minimizing these barriers, independent of government largesse (i.e., to identify a market-driven solution based on a platform technology). Of particular interest was the exploration of mechanisms to facilitate data content contributions from undeveloped countries where project funding is particularly absent, while also facilitating the increased desire for scientists, citizens, and students of all ages to actively contribute and participate in debates on global warming, ecological diversity, environmental health issues, and environmental justice (Garrigos-Simon et al., 2012). The global effort to collect and synthesize environmental science data are very significant, with some measures costing USD billions of dollars per year; however, studies have shown usually less than 1% of this science data are publicly shared (Reichman et al., 2011; Wallis et al., 2013), a serious breakdown of the market.

The largest global businesses by market capitalization were reviewed to understand why their use of technology platforms has propelled their rapid growth rates, and to identify barriers to the successful implementation of these platform strategies. Lessons learned by these businesses were reviewed to determine whether a similar strategy could be emulated to create a science-data sharing platform that meets the needs of citizen scientists in particular (as a proxy for the broader science community) and how market forces could be harnessed for that platform to ensure long-term commercial viability.

CHAPTER 2: LITERATURE REVIEW

Beginning with the rationale for an open science-data sharing platform, the unique characteristics of platform business models were examined to differentiate them from the traditional linear business model. The different emphases in management functions required for a platform business were discussed, followed by unique opportunities for powerful growth that are not usually found in traditional linear businesses. Not all markets lend themselves to a platform-technology business model, which was consequently reviewed in the context of the potential for science-data sharing, followed by a review of market disruption strategies and opportunities as the usual rationale for establishing a platform business model. A common failure of platform businesses was also reviewed—namely, the failure to obtain critical mass to make the platform attractive to both parties to a transaction. This review was followed by early recognition of the importance of simplifying the core transaction considered the most impactful to each party involved in the transaction. The summary then assesses market opportunity during the COVID-19 global pandemic.

Rationale for a Science-Data Sharing Business Model

There are significant efforts set forth globally by citizen science (CS) every year, but relatively little to show for these efforts when measured by direct public access to the data collected. By extension, there remains little ability by secondary parties to use data collected for higher value-added decision making, such as social justice and proactive environmental community health intervention. CS data collection is labor intensive and characterized by volunteer labor (Theobald et al., 2015). As Theobald et al. calculated for U.S. CS projects alone (after applying the estimated participation hours and the U.S.

minimum wage), CS data collection has a large labor cost component worth billions of USD per year in in-kind economic worth. Nevertheless, researchers must be cautious when applying traditional costing metrics to a function with a high degree of volunteerism.

The lack of science-data sharing has been a long-term problem. A study by Wallis et al. (2013) on why data gatherers in scientific communities tend to not share data with the broader community identified various policies that have tried to rectify this issue. Wallis et al. identified that when requests are made, professional scientists are quite willing to share their data sets (with conditions such as recognizing authorship). Unfortunately, as these data repositories are government-led, there are no market-driven incentives to publicize esoteric data sets. There is also no market mechanism to incentivize nor reward data set owners to incur the time–cost expense of making their data available to the public, which results in an opportunity-cost loss to the scientific community, a lack of public understanding of science issues, and a lack of awareness of social justice issues. There is also a high cost barrier to sharing esoteric data sets. Wallis et al. presented a discussion of science data sharing and reuse, finding it is difficult to justify the cost of matching disparate data content providers with data content users. These market matching requirements and the lowering of transaction costs became part of the rationale for establishing a market-driven science-sharing platform technology.

In addition to efficiencies from a market-based science-data sharing platform, there is also the question of the nature of a market-based reward mechanism wherein data content contributors would be rewarded commensurate with the value placed by the market on their contribution. This market-driven reward mechanism may be financial,

based on recognition, satisfaction of personal development, or social justice goals—or any combination of these potential benefits. This reward system would present a radically different approach than the government fiat approach to deciding which projects to fund and publish (Jahnke, 2015). The benefit of such a successful strategy would involve rewards given to market participants who have historically been underrepresented or disenfranchised. These participants could include CS data collectors from underdeveloped countries where the government’s funding support for their projects is sparse (Nuñez & Pauchard, 2011), but their data sets are no less valuable—arguably even more so due to their scarcity. The solution set includes determining whether a commercially viable platform business model can be established that satisfies the general needs of the CS community as identified by numerous authors e.g., Moazed & Johnson, 2016; Parker et al., 2016; Perren & Kozinets, 2018; Reillier & Reillier, 2017; while also providing a reward or recognition function (Garrigos-Simon et al., 2012) independent of the government’s funding priorities. Van Alstyne et al. (2016) argued when a platform enters a pipeline firm’s market, the platform almost always wins; similarly, a platform business model that is market driven is expected to be more efficient and cost effective in providing comprehensive data sets and satisfying unmet needs of the global science community.

Establishing a platform technology that addresses this global issue is not a straightforward matter, and has its own set of unique problems to solve. These problems were identified in this study; however, the benefits to the global community are potentially significant. As argued by Lee and Stvilia (2017), “Increased access to research data increases the impact and efficiency of scientific activities and funding” (p. 1). Of

course, whether this goal can be achieved on a sustainable commercial basis is a different matter. In this literature review, impediments to a successful market-driven, open-science business model through a review of major barriers and their potential solutions were sought to be identified.

Characteristics of a Technology Platform Business Model

Before answering the question of why platform technologies have become a dominant business model, researchers need to understand what characterizes and differentiates them from traditional business models. A decade ago, the largest company by market capitalization was ExxonMobil, a firm that built an extensive infrastructure of hard assets composed of pipelines, refineries, storage facilities, a transportation fleet, and a distribution network. By the end of 2016, hard-asset firms were displaced, and the world's five largest companies by market capitalization were all platform-powered (i.e., their assets were dominated by soft assets such as IP): Apple, Google, Microsoft, Amazon, and Facebook (renamed Meta). Clearly, not all platforms are equal and can have attributes of varying importance; however, there are some key commonalities of successful platform businesses relevant for science-data sharing.

Release of Excess Value

A key feature of platforms is they produce excess value; consequently, their valuation is high compared to the cost of their physical assets. Porter (1985) used the concept of the value chain to analyze traditional organizations from a linear perspective, considering the transformation of inputs (i.e., raw materials) into outputs (i.e., products and services), along with sales to customers and the competitive advantage that can be

established when creating an aggregate value proposition for the customer. Platform business models have different value propositions as they focus on attracting, matching, and connecting communities; they do not own any inventory, but their value is contingent on facilitating the transaction between exogenous parties. The underlying economics of such platforms were first described by the Nobel Prize winner Tirole (Rochet & Tirole, 2014), who argued platform technology is almost always a two-sided market, and used the rather quaint phrase to “get both sides of the market on board” (p. 1) in terms of the buyers and the sellers. Reillier and Reillier (2017) provided a simple definition of platform businesses “as those connecting members of communities and enabling them to transact” (p. 5), stating platform business models have become mainstream. The researchers cited diverse business entities such as eBay, Airbnb, Uber, Facebook, and Tinder (Reillier & Reillier, 2017). Costa (2016) further challenged the traditional concept of a firm, declaring platform technology is a different mechanism in the seller–consumer interaction. This challenge placed entrepreneurship at a higher level than traditional management actions, such as those described by Porter (1985).

Parker et al. (2016) presented the platform’s overreaching role, which is to “consummate matches among users and facilitate the exchange of goods, services, or social currency” (p. 5). The authors identified the platforms’ key attribute, namely the delinking of ownership and access to assets, which creates value within the community by providing greater access to underutilized assets. Parker et al. described this process as a value increase through enabling the use of an asset for the greatest economic value. An illustrative example of the process for a physical asset (i.e., water) is water rights in Australia, which lower the negative impact of disastrous droughts by enabling farmers’

water rights to be traded to those farmers who can generate crops of greater value.

However, the effects of this phenomena on digital goods that can be traded endlessly are magnified significantly; this magnification bodes well for a science-data platform, which is comprised solely of digital goods to release excess value that would otherwise be limited or localized if they were a physical good, such as those typically traded on eBay.

The delinking argument of Parker et al. (2016) paralleled earlier conclusions by de Soto (2000) who, after a multidecade investigation of global economies, attributed the success of capitalism in the western world to the theme of strong preservation of legal rights leading to the decoupling of assets and the release of their inherent value. As a note, it was expected this freeing-up process of inherent value would become even more self-evident for digital assets where copies can be made indefinitely. Such a prediction was further explored in this dissertation. This feature of platform technology for digital assets predicts it will produce considerable value well in excess of a comparable platform that trades in physical assets alone (as per the Australian water rights example), and the implementation of this prediction may be the key breakthrough for establishing a commercially viable science-data platform that is market driven rather than government-fiat driven.

Providing Greater Reach

Traditional farmers' markets have always been around as a type of platform that enables buyers and sellers to transact at a convenient place and time. What differentiates a technology platform from traditional modes of transaction is a significantly greater reach provided by the matching of two parties. This expanded reach has enabled underutilized products and services to find a market that would otherwise not have been

cost effective to find in a more localized and physically constrained marketplace. One of the earliest platform businesses, eBay, demonstrates the ability to extend market reach by matching consumers with sellers. Spulber (2019) discussed how online platforms help entrepreneurs by allowing them to create their own business on those platforms, giving the example of an independent seller on eBay as analogous to the farmers' market. The platform's ability to extend the reach of transacting parties enhances the underlying asset's value in the transaction.

Platforms as Intermediaries and the Rise of the Sharing Economy

The sharing economy is a relatively new phenomenon that is facilitated by platform technology. There are many examples, such as Airbnb, which allows people with spare rooms to compete with hotel chains very competitively and where short-term rental accommodations account for more than 20% of the U.S. lodging market (Gold, 2019). Uber is another example as a platform that enables the matching of car drivers with customers. These are also examples of releasing value from what would otherwise be underutilized physical assets. The relevance for digital assets is that this sharing phenomena has a potentially significantly higher multiplier effect for generating value.

The study of platforms demonstrates the importance of participation and coordination in the formation of markets, showing how platforms affect innovation and entrepreneurship and pose significant implications for the theory of the firm. Spulber (2019) recognized the critical role of platforms as intermediaries in endogenous price adjustment and market clearing and provided explanations for the bid-ask spread, including market power, search costs, matching costs, adverse selection, and moral hazard. Spulber also provided a definition of a platform, adding to a plethora of

definitions; what was added through this study was a more recent review of the role of platforms by this well-respected academic and their thoughts on a sharing economy. Specifically, Spulber argued markets and platforms represent the same economic phenomena, requiring some traditional definitions of business to be refined to include the platform technology phenomena. Unfortunately, Spulber ignored the instance of platforms commonly providing price-free transactions, as per the example of Facebook. Spulber also discussed how online platforms facilitate entrepreneurs by enabling them to create their own businesses on those platforms (e.g., the independent seller on eBay). This finding was highly relevant to this research, as it provided a close approximation of entrepreneurial science-data content providers who sell and share their data sets on a science data platform.

Platforms Positioned as a Utility or as Predatory

The discussion of platforms as intermediaries must include their market positioning. Questions include: Is the platform a utility, or considered predatory? What is the optimal positioning for a platform that wishes to facilitate science-data sharing with the objective of long-term, sustainable profitability? Page and Childers (2012) discussed the market positioning of Twitter and whether it reflected antitrust actions. Their argument rested on whether Twitter has undergone system enhancements to “fill holes in its platform functionality” (Page & Childers, 2012, para. 18), or whether those enhancements are designed to exclude complementors of the platform by eliminating profitable opportunities of third parties. A balance must be struck between enhancing platform functionality and providing opportunities for entrepreneurs to add to the value proposition. If the business decision is to have a zero price for content providers and a

zero price for content consumers, it could be argued the platform is positioned as a utility and the opportunity for entrepreneurship is unfettered—enabling content providers to introduce esoteric, innovative, and high-demand data sets for their own benefit while enhancing the platform’s value proposition to the community.

Page and Childers (2012) concluded the utility strategy has a lower profitability path in the case of platforms such as Twitter; however, this research established a utility strategy as the least contentious strategy with both professional and citizen scientists. Further, a frictionless participation (i.e., zero cost) would lead to high network growth and instill confidence by the science community, as this commercial market-positioning strategy is sustainable over the long term and does not have sinister capitalist, monopolistic motives.

Network Effects, Supply Side Versus Demand Side

The rapid growth in the value of a platform business is a significant feature of its success. Costa (2016) presented key differences between the networks of two transport companies: Union Pacific and Uber. The former adds rail lines, stations, and carriage stock, thereby increasing the supply side of its value proposition. In contrast, most platform technology businesses rely on the demand side of economies of scale—the more customers join the platform, the more valuable the platform becomes, a phenomenon reflective of Metcalfe’s power law and the lower the unit cost (Metcalfe & Boggs, 1976). This inherent ability of software-driven platforms to rapidly create a more cost-effective scale is the primary reason why Costa predicted they will surpass traditional linear business models relying on hard-asset growth. The lowering of the unit cost for a

transaction due to the rapid scaling of the number of transactions is an example of the power of the network effect.

Further, increased numbers of users can potentially also have a negative effect on valuation. Costa (2016) presented the case of OKCupid, an online dating platform, with the initial problem of “too many men converging on too few very [attractive] and interesting women” (p. 383), resulting in those targeted women being inundated with messages and their subsequent departure from the site. This imbalance was solved by implementing a filtering system to match both parties more closely, which resulted in fewer but better quality transactions. This common issue has been termed the “naked hairy man” problem—a real issue, and first identified in a similar site, Chatroulette (Parker et al., 2016). This effect indicated not all network growth is positive; rather, growth has to be managed so as to not destroy the underlying value proposition.

The Curious Anomaly of Price

No linear business can provide products and services for free for very long, despite Drucker’s (1955) assertion that a business’s purpose is to create and retain a customer (not necessarily to make profits). Some notable platforms can provide their services without charging and have been very successful through monetizing their traffic, with Facebook as a key example. Parker et al. (2016) ignited an interesting discussion on monetization, viewing any pricing imposed on either the buyer or seller as a friction or dampening factor on the natural growth of the platform’s ecosystem. They further argued there is a marketing case to be made in subsidizing either or both parties’ participation in the transaction. Examples of this factor system for traditional products in linear businesses are the printer market, where a high-priced toner subsidizes the relatively low

cost of sophisticated printers; and the market for grooming products, where razor handsets are sold close to cost, but the product's replaceable razor blades contain the price mark up to generate profits. A potential example of the science-data sharing market would be providing a data storage and editing facility at zero cost (i.e., subsidizing the data content provider) while providing the consumer with a free search function and free access to the published data sets. These are examples of providing value and subsidizing both parties to the transaction while monetizing eyeball traffic through advertising and sponsorships. This system also reflects the Facebook business model, where the company obtains 98% of all revenue from advertising placement and whose business was valued at approximately \$918 billion as of December 2021 (Nasdaq, n.d.). Facebook demonstrates friction-free price mechanisms can be established for both parties to the transaction while the firm also establishes profitability and market share.

Pricing models for platforms can be highly complex, particularly when deciding which party to the transaction is to be subsidized, the point in the transaction where a cost is imposed, and the manner in which this fee is to be undertaken. Tirole (2015) recognized technology platforms' goals of pricing policy involved maximizing platform value through the network effect, where the larger the participants' network, then the greater the value of the platform. Tirole described platform pricing models for two-sided markets, emphasizing both "monopoly and competitive platforms design their price structure so as to get both sides on board" (pp. 1674–1676). This pricing policy could be through cross-subsidization of one side of the transaction, such as in the razor or printer market examples, or through platform subsidy for both sides of the transaction and monetization obtained through other independent mechanisms (e.g., the monetization of

traffic volume through advertising placement such as in the Facebook model). A key conclusion from this analysis of pricing policy was that platform technology business models have the potential for significant price competitive advantage over traditional linear business models (Van Alstyne et al., 2016). These models have the potential to offer a free service on not only a temporary promotional basis, but also an ongoing, sustainable basis. An additional advantage of a platform when it transacts in digital goods is that it has unlimited supply due to the model of providing copies of a data item, rather than individual, physical items at a price. Given digital exchange tends to have minimal transaction costs, market participation can be optimized by eliminating participation costs using a business model where organizational costs and profit margins are met through third-party participation. This optimization also reflects the Facebook model of monetizing eyeball traffic rather than direct transactions.

Impact of Absence of Price Signals on Market Participants

The absence of price signals creates new dilemmas for platform market participants. What is the impact of zero cost of digital items for consumers on the goal of profit maximization for content providers? With zero revenue from direct sales, what should be the incentive for publishing data sets with no profit maximization incentive? Would it be fame, as is the case with TikTok; or social interactions, as with Facebook; or career opportunities, as with LinkedIn? If the impact is a monetary incentive, it could be in the revenue share from advertising or sponsorship generated by the platform. Fame could be measured by the social interactions generated by the network. Is it sufficient to induce content owners to absorb a modest data take-up cost of the sharing of their science data? Should data content providers be subsidized in their data take-up costs as part of the

strategy to reach a critical mass of data sets? These questions were addressed in the research methodology section of this dissertation.

Another unique price-related issue was the problem of two suppliers of digital assets with the same price, possibly zero. How does the purchaser or content consumer decide which of the two sellers is the most reputable, given there are no price differential signals? Should the platform provide metrics on the number of downloads or a customer feedback rating, as is the case with Uber drivers and Amazon book resellers? Should the platform provide in-depth information on the seller's project? Who is the original content owner, given the data set may have been copied and rebranded? Should blockchain technology be used for digital assets to ensure immutable proof of original ownership by applying a non-fungible tokenization (NFT) strategy? These are important issues to be resolved for a technology platform transacting in digital assets that have minimal or zero price signals to the content's consumers.

Managing the Technology Platform

It is increasingly apparent that professional management of the technology platform business model differs from that of traditional linear firms. Perren and Kozinets (2018) added to the diverse terminology around technology platforms, referring to them as "lateral exchange markets" (p. 21) and investigated deeper the concept of the two-sided platform and the role and responsibilities of technology platforms. An interesting concept they highlighted is *sociality*, which is the tendency to form cooperative relationships with others in groups. Sociality's relevance to technology platforms is that new mechanisms for social interaction can be established relatively easily, leading to stronger bonds among participants—in turn, these bonds lead to a more stable network

and an enhanced network value. Perren and Kozinets (2018) maintained providing “socially rich exchanges” (p. 23) have a positive effect on consumer trust, generating “community-based governance” (p. 21) on their subsequent intention to participate in the platform’s services. Consequently, the technology platform has a unique opportunity to leverage engagement through sociality. An example of a comparable sociality in a traditional linear business model would be a car enthusiast club for car manufacturers. The concept is not new, but what has changed for platforms is the ability to expand membership significantly and rapidly due to the availability of sophisticated communication channels.

Role and Responsibilities of Technology Platforms

The market failures of platforms have also become more evident via the unauthorized sharing of user data, lack of security from data breaches, “fake news” accusations, and lack of responsibility for stealing personal data (Navia et al., 2022). Navia et al. discussed many of these issues with a case study concerning Facebook being investigated in Germany and Italy under competition and consumer laws respectively, and they bemoaned the lack of informed consent and abuse of market power by Facebook. The matter of trust has always been an issue in business, and it becomes even more apparent when using a platform where the consumer is a step further removed from the seller. This issue was ably discussed by the Nobel Prize winner, Akerlof (1970), whose study “The Market for Lemons” included a discussion of car sales; and where he postulated that without an adequate trust mechanism, the price points of all cars would inevitably converge toward the price of the worst possible car—another example of market failure. There are both opportunities and responsibilities for platform businesses

that have rarely been considered by the traditional linear business, leading to the conclusion that heightened governance is required of platform businesses and establishing a well-thought-out trust framework is an essential component of a professional platform.

Unique Roles Within Platform Businesses

As platform businesses organizationally differ from a traditional linear business, it is necessary to discuss some key roles within a platform business's organizational structure.

Content Curation

Content curation is a unique feature of many platform technologies; it is in the interest of the platform community to have reliable content. Content curation is typical for social platforms such as Twitter, where egregious use of extreme viewpoints and fake statements can inflame social groups—the latter reason was given by many technology firms behind cancelling the social medium platform Parler, in January 2021 (Baines et al., 2021). Facebook has also made statements on its policy concerning blocking fake political advertisements that have the potential to sway electoral voting patterns. For this study, where the potential for a technology platform to share science data was reviewed, the comparable role was for data content curation to which data standards were adhered and the potential for fake data were investigated and accounts potentially frozen until the issue was resolved. Lee and Stvilia (2017) analyzed the role of data content curators at 13 large research universities in the United States and found 80% of a curator's time went toward communicating with content providers to assist them in ensuring the integrity of

their data sets; however, this direct communication was not always possible when large volumes of items were supported on a technology platform. eBay was one of the first technology platforms to implement user feedback to rank a seller's integrity, which enabled buyers to discern whether or not they were dealing with a reliable seller. Mickey (2010) found a seller's negative rating significantly impacted the price they ultimately received. The ride-sharing services Uber and Lyft, along with the room sharing platform, Airbnb, have a similar user-defined feedback system, which has become an essential platform feature.

Fiduciary Duty

Firms selling products have long had the practice of offering warranties on product sales. Technology platforms have emulated this practice through a fiduciary function that is empowered to roll back transactions if a consumer has proven they have not received what was advertised. For a cost-free transfer of science data, a roll back of the transaction is somewhat mute, but a negative review of the quality of the downloaded data sets could result in a downgrade of the data content owner's status, which can impact others using their data. This issue is an example of losing reputation within a community.

Content Providers as Social Entrepreneurs

Much has been written on entrepreneurship and the role of management within traditional linear businesses, which Drucker (1955) referred to as *intrapreneurship*. A more recent topic introduced particularly with respect to two-sided technology platforms is *social entrepreneurship*, which Phillips et al. (2015) defined as the creation of social value as opposed to personal or shareholder wealth. Here, content providers assist their

host platform with a social mission to accelerate their organizational level of entrepreneurship. There is probably no better opportunity for social entrepreneurship than science data sharing, synthesis with other variables, cross-pollination with diverse ideas and debates, resulting analysis, and recommendations for community action for social justice based on hard data.

In an interview with Porter on the subject of social entrepreneurship, Porter decried the “crisis of legitimacy of capitalism” (Driver & Porter, 2012, p. 422) and the importance of leadership that is appropriate for the societal times in which an organization resides. Fundamental to organizational leadership is the recognition of “organizational Zeitgeist” by the chief executive officer (CEO) in particular or “contextual intelligence” among the values of society (Driver & Porter, 2012, p. 423). Organizational Zeitgeist is also an area where a technology platform that facilitates science-data sharing—particularly epidemiology or environmental science-data sharing—has the ability to capture societal imagination and provide a mechanism to project concerns and views based on hard data and subsequent discourse. In this manner, the platform becomes a social entrepreneur by seeking out nonmainstream data and shifting the debate, which is what the platform Parler attempted before it was shutdown (Baines et al., 2021). Shifting the debate is a unique power of a platform in supporting and facilitating the social entrepreneurship of its content providers. In combination with organizational Zeitgeist, social entrepreneurship is another differentiator of platforms from traditional linear businesses where innovation is overwhelmingly driven by in-house management.

Growth Strategies and Market Barriers for Technology Platforms

One of the defining characteristics of successful technology platforms is their rapid growth. What is the primary driver of this rapid growth, and what favors one market over another? Are there any new and more compelling growth drivers?

Ability to Scale Rapidly

A key feature of platform technology businesses is their ability to scale exceedingly rapidly once platform participants identify an opportunity to obtain a positive return on their participation. This ability is partially catalyzed by the capability of software structural businesses to scale very cheaply compared to manufacturing or even service businesses; moreover, a pure platform business tends not to own the resources being sold (i.e., this model does not have inventory to finance, create, warehouse, and distribute products). There are numerous examples of this scalability: Uber, the world's largest taxi service, does not own vehicles; Facebook, the world's most popular social networking firm, owns no content; Amazon, the world's most valuable retailer, owns no inventory; and Airbnb, the world's largest accommodation provider, owns no hotel real estate. Parker et al. (2016) discussed this phenomenon and argued platform information systems are "inverted" (p. 11), and they provided the example of how a traditional linear business would have back-office enterprise resource planning systems, whereas platform technology businesses are more outward-looking and characterized by customer relationship management systems. Platform business strategy

has shifted its focus from controlling internal resources and creating competitive barriers to engaging vibrant communities and ecosystem governance.

A key feature of technology platforms that experience rapid growth is they enable users to easily provide feedback on their experiences and add value to the platform (Reillier & Reillier, 2017) through a direct positive feedback network effect, such as feedback on books on Amazon, feedback on user experiences with Airbnb room stays, and driver experiences on Uber (i.e., the coveted 5-star experience). Uber, to its credit, allows a driver to also rate the experience with their customer, which is a breakthrough because it is possible that Uber drivers will, one day, be able to add a surcharge or discount for a customer based on the ratings provided by other Uber drivers (Chan, 2019). Elon Musk has recently announced he will enter into the car insurance business for Tesla drivers by using data gathered from their Tesla vehicle, which will enable the company to offer personalized and lower cost insurance to good drivers who historically have had to subsidize the average driving habits of more reckless drivers (Chan, 2019). These various big-data-enabled features of technology platforms have contributed to a high network growth that traditional linear businesses have had difficulty replicating.

The Evolving Nature of Web 3.0

There is a broad discussion about the technical nature of Web 3.0 in existing literature. Garrigos-Simon et al. (2012) analyzed the importance of social networks in this new context, the effects of Web 3.0 on management and marketing of organizations, and the way organizations can exploit these changes. Specifically, researchers have concentrated on the importance of community managers and the relevance of crowdsourcing processes in coping with new changes. Other studies have identified user-

generated content as the key concept of both Web 2.0 and 3.0; the primary finding of Garrigos-Simon et al. was crowdsourcing content through social networks can be the most cost-effective and fastest mechanism of value creation. Still, what their research omitted was a discussion on incentivizing participants to use emerging Web 3.0 strategies of revenue sharing with content providers, a new field of research that has become the latest frontier in technology platforms. An important empirical study from Sun and Zhu (2013) identified in a site with 4.4 million blog posts, popular content increased by 13% after a 50% revenue sharing policy was introduced, along with a content quality improvement. Sun and Zhu (2013) anticipated Web 3.0 revenue sharing will serve as an integral component for motivating content providers, and this feature was explored for science-data content owners to mitigate the data take-up time–cost expense of sharing their data sets, particularly in underdeveloped countries where the financial reward multipliers and reliable science data are sparse.

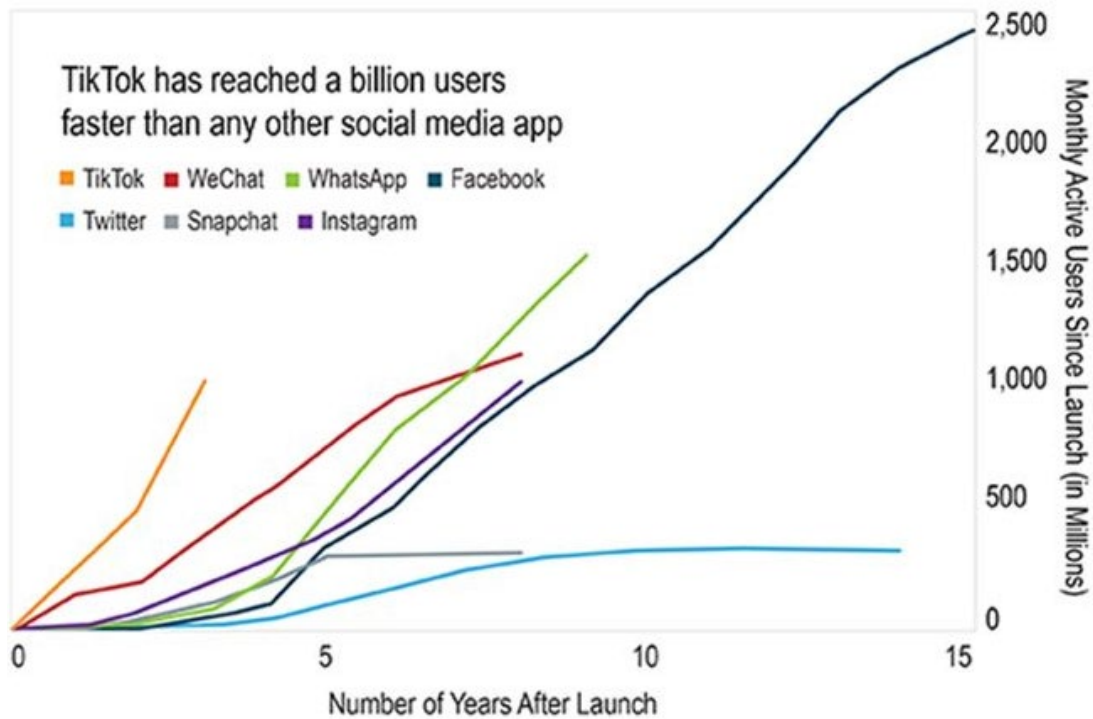
Network Effects

Network effects are a feature of technology platforms and present the cheapest and fastest route to growth. What makes a technology platform valuable is the number of users (Metcalfe & Boggs, 1976). Parker et al. (2016) argued the network effect can be used to generate significant growth to a technology platform through *viral growth*, a pull-based process encouraging users to urge other users to join the network. When this growth is successfully implemented, a network becomes the driver of its own success. Wan et al. (2017) shared this view when referring to a virtuous circle of the positive feedback loop that benefits both sides of the platform by increasing user base growth and engagement. TikTok is a recent fad that some regard as a silly platform (Vijay & Gekker,

2021), and was initially popular exclusively among teenagers, where the service resonated particularly strongly (Wee, 2017); yet, due to the COVID-19 global pandemic, TikTok saw a spike in adult users. The platform was initially comprised of 15-second videos (now being extended) that have been crowdsourced and is, remarkably, the fastest social media platform to reach 1 billion users (see Figure 1)—demonstrating the power of a viral network with very low friction to participate in two sides of the transaction (i.e., both providing and acquiring content; Kennedy, 2020). Due to the June 2020 military clash between India and China in the Himalayas, the Indian government declared the Beijing-based ByteDance, the owner of TikTok, a security risk, inhibiting its use by Indian citizens who comprised 31% of total users (Mishra et al., 2022). This action may inhibit the platform's stellar growth.

Figure 1

Number of Social Media Users by Platform



Note. From “TikTok/Hong Kong: Worlds Apart,” by Financial Times, July 2020

(<https://www.ft.com/content/bdf76bd9-897b-47df-87b1-14c7ae662331>). Copyright 2020

by Financial Times.

Multisided Versus Two-Sided Platforms

This study was primarily concerned with two-sided platforms; nevertheless, multisided platforms and their potential role in facilitating science-data sharing and creating barriers to potential market entrants must also be mentioned. One of the most well-known multisided platforms is Microsoft’s PC platform. In high-tech markets, firms such as Microsoft often depend on third-party developers to build applications on their platforms by opening access to their platform, and establishing standards for all

developers with application development tools to make that development more efficient. This increase in the number of applications on the Microsoft platform makes the platform more valuable to consumers, another example of transaction subsidy. This function is also an example of multisided transactions, wherein Microsoft negotiates with other parties in the transaction.

An example of complex multisided transactions is Intel and Microsoft's collaboration to create the Wintel platform. Intel, in turn, collaborated with other hardware suppliers for specialist chip components. This nexus of extended multisided collaborations generates considerable value for the consumer; however, as Page and Childers (2012) pointed out, it also establishes a case for breached antitrust laws due to the monopolized market position and barrier to entry created by this fortress of innovation and close collaboration. The U.S. federal government's case against Microsoft was based on "applications barrier to entry" (Page & Childers, 2012, p. 3) consisting of network effects. In this antitrust case against Microsoft, the government was referring to the network of application developers with which Microsoft has negotiated terms. This is quite different to the consumer network and which anti-trust law usually targets. An important point here is both Microsoft and Intel benefit from indirect network effects that pose barriers to competitors, which is an attribute of multisided platforms and is significantly less attainable for linear business models in their competitive environments.

A science-data sharing technology platform could become multisided by negotiating deals with instrument manufacturers to provide free or subsidized equipment to data content providers. Similarly, negotiating with large-scale data content owners to ingest their data into a technology platform—hence subsidizing their participation cost—

makes the platform more attractive to consumers due to a greater volume of diverse data sets. The latter has been the strategy of the U.S. National Institutes of Health (NIH), which requested proposals to establish a cloud-based technology platform commencing September 2021 to identify health issue correlations and trends in Africa (Glass & Tromberg, 2019); budgeted at \$11.25 million over 5 years, this initiative plans to establish a data repository called “Open Data Science Platform” to manage the data take-up process and provide access to the data via a portal. The NIH has also published a request for proposal offering six \$7.5 million grants over 5 years to data aggregators and content providers to source data for the selected cloud-based technology platform. This budget is solely for data from sub-Saharan Africa, indicating the scale of these types of projects, and is entirely consistent with this study’s market size estimates.

The NIH has not offered a solution for data take-up during the post contractual period, but the organization has asked for suggestions, returning to the problem of remunerating data content providers. A science-data sharing technology platform could guarantee a minimum revenue to data content providers to ensure diversity of data sets to the consumer as per the NIH’s six “Data Hubs”. On the other side of the transaction, the platform could provide additional benefits to data consumers (in the NIH example, health data scientists), such as priority alerts when new data sets are published or direct introductions to content providers. These examples provide incremental additions to the value proposition of a technology platform, creating additional barriers to competitors’ market entry, and solidifying its multisided status.

Characteristics of Markets That Lend Themselves to a Technology Platform Strategy

Not all market niches are appropriate nor have the potential to lend themselves to a platform strategy. With the notable exception of IBM, all world-dominator firms of a decade ago have been displaced in their relative importance by platform businesses. An understanding of key characteristics for a platform business to thrive in its chosen market is of fundamental importance to a successful platform strategy. This thorough comprehension of the disruptive potential of platforms has been demonstrated by Apple, a company that rose from relative obscurity to become its market dominator. As previously discussed, the market is also a key participant in the success of the platform business, and has to be convinced there is greater value to be realized in participation than in refraining from participation. Sometimes a catalyst is required, which may take the form of a well-funded entity. The market may be highly fragmented, and the platform may provide opportunities and incentives at the margin that may not have been realizable prior. These issues are reviewed in the following subsection.

The Business Transformation of IBM

The transformation of IBM's legacy linear business model demonstrates what can be done even by very large legacy businesses to transform themselves into a dynamic platform technology business model, maintaining growth and market share. Jetter et al. (2009) investigated the influence of information and communication technology (ICT) on business transformation and how IBM responded to market changes, effectively reinventing itself. The findings of their study showed that as the global economy has migrated toward that of a service economy, where 70% of a developed economy is in

services, there has been use of information and communication technology to facilitate that shift in customer service orientation.

Enhanced application of information and communication technology use plays precisely into the technology platform's strengths, and Jetter et al. (2009) claimed this is the primary reason technology platforms have grown significantly in importance, as it is the enabler of customer service on a global scale. Further, Jetter et al. identified a company's ability to radically change as one of the characteristic properties of a global platform business. This was a seminal study, as it demonstrated for a legacy dominator company (e.g., IBM) to remain a global market dominator, it must change rapidly from a product-centric to a customer- and service-centric focus, where added value to the customer can be created jointly with the customer—the hallmark of a technology platform (Jetter et al., 2009).

The Business Transformation of Apple

Apple was initially a traditional linear business that manufactured the iPhone. In 2007, five major competitor mobile manufacturers controlled 90% of the industry's global profits. By 2015, the iPhone generated 92% of all global profits (Van Alstyne et al., 2016). Apple's surge to market dominance arose from its transformation into a platform technology business, connecting app developers on one side and app users on the other side of the market with the App Store as the marketplace. This surge shows how platform strategy usually prevails over a linear business strategy. Van Alstyne et al. (2016) identified three key shifts in transforming a linear business into a platform business:

- From resource control to resource orchestration, where it is recognized that the network of producers and consumers become the primary asset.
- From internal optimization of the linear business to external interaction persuading participants to interact.
- From a focus on customer value to that of the overall ecosystem, where the size of the network of platform users is usually the best metric of platform value creation.

As discussed previously, this changed focus requires a management skill set different from that of a traditional linear business.

Motivations for Sharing Data

Belous (2014) introduced the topic of identifying motivations behind data sharing, and this topic was also central to this study on science data sharing. Belous found that a wide variety of motivators might influence the decision to share information online of which the most dominant motivators are intrinsic rewards (i.e., the feelings of joy, happiness, satisfaction and fulfillment that accompany the act of sharing), and self-efficacy (i.e., the sharer's belief in his ability to supply valuable information to others). Belous also suggested self-efficacy and expertise in the area of information sharing can help predict the influence of other motivators. Experts tend to be more driven by extrinsic rewards, but also experience a stronger obligation to give, whereas nonexperts might seek a sense of community and experience a stronger obligation to repay. Pounds and Bostock (2019) found in a survey of educators and students involved in aquaculture research, "altruistic drivers, recognition and encouragement, motivated both students and educators more so than financial rewards" (p. 1561), with over 80% of respondents sharing this

opinion. Pounds and Bostock also found students are more likely to upload and share their data rather than educators, although they hedged their conclusions by suggesting this finding may also be due to both time constraints of educators and copyright issues with their academic governing institutions.

In the specific market niche of CS, there is a diverse range of participants, including home hobbyists, high school students, esteemed scientists and institutions, and more. Cowell et al. (2017) postulated degrees of empathy toward the environment will also vary between cultures and between developed and underdeveloped economies. Cowell et al. found morality is an evolved aspect of human nature, but heavily influenced by culture, explaining over 20% of the variance worldwide. The global and shared impact of the COVID-19 global pandemic will probably result in furthering common awareness of environmental and health safety issues. These problems were explored in this study, because understanding market participants' motivation helps guide recommended strategies for establishing a viable commercial solution. Further, according to Belous (2014), there are parallels with participant diversity in other social network platforms and those supporting exchange markets; for instance, the richness of participant diversity tends to provide positive feedback to network growth, a view shared by Gerhardt (2007) when discussing the profile of organizations in general. Consequently, a global and diverse community of likeminded citizen scientists would be a positive driver of network growth.

Government and Academic Solutions or Lack Thereof

A global problem with science-data sharing involves identifying a government authority that has adequate funding and will take up responsibility across political

boundaries and establish coherent long-term rules. Plantin et al. (2018) discussed the requirement for grant-issuing agencies to facilitate data sharing through targeted grants. The NIH made it a requirement in 2003 for grantees over \$500,000 to have their science data published; the U.S. National Science Foundation established similar data sharing requirements in 2010 (Plantin et al., 2018). In early 2020, the NIH implemented its plan to issue requests for proposals seeking data hubs (i.e., data content providers) exclusively from African countries for their proposed technology portal, and budgeted a total of \$7.5 million per year to be disbursed to these six data-hub projects over the subsequent 5 years (Glass & Tromberg, 2019). The overall 5-year budget of approximately \$62 million, which included a request for the market to provide a Coordinating Center and an Open Data Science Platform, together budgeted at \$2.25 million per year. This NIH initiative for Africa science-data sharing indicates the cost simply to aggregate already collected science data from various projects in a single continent. The NIH has not suggested how to sustain data collection efforts post-project, although it requested applicants recommend a strategy (coincidentally, the topic of this research). These various initiatives provide opportunities to develop innovative archiving and storage systems, although the majority of CS data sets are on a more modest, single-project scale.

In the United Kingdom, Figshare was commercially established in attempts to respond to the requirements of large academic projects supported by grants; yet, Plantin et al. (2018) cast dispersions on Figshare's commercial motives. They acknowledged Figshare's simplicity in alleviating some technical disincentives for data sharing, but criticized it for not addressing directly "the systemic lack of positive incentives for data sharing" (Plantin et al., 2018, p. 9) when referring to the lack of financial remuneration to

the data content owners. Plantin et al. also made a point about Figshare providing the potential for elements of “fame but not of fortune”, which they criticized as a market failure on its part.

This criticism, however, is unfair, as Figshare is a for-profit organization and has no commercial obligation to share revenues with content providers; yet, their moral obligation is up for debate as per prior comments concerning the Web 3.0 debate of sharing remuneration with content providers. Plantin et al. (2018) argued the case against private or for-profit platforms for storing scientific data, maintaining the role should remain in the control of education and government entities. Plantin et al. appeared to be heretical to the concepts of free-market economics and market efficiencies and disparaged the concept of private enterprises fulfilling this function, instead referring to the eviscerating impact of FedEx and UPS on the U.S. Postal Service. They reasoned against private enterprise participation in science-data sharing on the grounds of their profit motive’s obscurity (specifically citing the case of Figshare), and their dubious ongoing commercial viability, concluding access to data sets can be tenuous. However, Plantin et al. failed to mention that less than 1% of science data are shared with the public (Reichman et al., 2011; Wallis et al., 2013), despite political and funding barriers, and they also ignored the potential of private enterprise to fill needs left unsatisfied by government or academia (i.e., a clear case of market failure as evidenced by the paucity of data set sharing).

Role of Platform Technology in CS

This section of the literature review discusses topics relevant to CS, which served as the proxy for the broader science community to bring focus to this study.

Social Justice Reforms

Returning to the theme required for a CS technology platform, Bonney et al. (2016) asked whether CS can enhance public understanding of science, and what role a platform technology has to play. The researchers came to the important conclusion that participation in CS gave people a voice in local environmental decision making, and an example they gave included social justice reforms in a community science project in Oakland, California. The project mobilized people from one of the city's poorest African American and Latino neighborhoods to collect air-quality and health-impact data, documenting the degree to which air pollution affected area residents. One project outcome was a recommendation preventing short-haul truckers from idling while waiting for Oakland port pickups (Bonney et al., 2016). This outcome provided an excellent example of the increasing use of CS community projects to facilitate social justice reform, a topic Bonney et al. felt would be a significant new growth area for CS (i.e., empowering local citizens).

Supporting Educational Curricula of Science

Another area Bonney et al. (2016) identified as a potential for high growth was supporting educational curricula in science through CS projects to which students could relate within their communities, and where they could use facilities (e.g., technology platforms) with graphical tools to explore those data further. The goal is to expand the public's understanding of science. Bonney et al. found strong evidence that CS's outcomes, particularly for data collection and data processing projects, were facilitated by platform technologies, although their analysis suggested outcomes could have been stronger.

The answer can be found within other conclusions of the nexus between science-data gathering and public health patterns. Bonney et al. (2016) identified a motivational factor for gathering environmental data in the community and solving health issues by identifying the relationship between environmental factors and real-world health outcomes. Communities can relate to this relationship more readily rather than collecting data for nebulous science projects, which are more common due to the priorities of funding sources (i.e., the process is pure science driven rather than community benefit driven). This need to generate a community benefit gives an insight into prioritizing CS projects to make them more mainstream and community relevant. Surprisingly, Bonney et al. did not address the need for a sense of community through CS, nor did they address third-world issues and participation where environmental factors have a greater impact on health—which, to their credit, was the impetus for the NIH's request for a proposal for Africa. Instead, Bonney et al. focused exclusively on CS in the developed countries of the United States and United Kingdom.

The Isolation of Science Data and the Long-Tail Effect

Science-data isolation refers to the phenomena of data sets deemed to have little value compared to the cost of publication, and as a consequence, are lost or isolated from the general science community. Research by Wallis et al. (2013) was particularly relevant for a CS technology platform, as it discussed science data and the long-tail effect. They provided a pivotal argument and justification of a CS technology platform, as its primary role is to facilitate frictionless science data sharing and to recognize data collectors or data owners. Wallis et al. underwent a 10-year study of the Center for Embedded Network Sensing (CENS), a National Science Foundation Science and Technology

Center, and found CENS researchers were willing to share their data; yet, few were asked to do so. In only a few domain areas did their funders or journals require them to deposit data. This finding was critical, particularly as Wallis et al. also found, “CENS researchers are willing to share data if they receive credit and retain first rights to publish their results” (Table 2).

Why are science data not shared more broadly? Wallis et al.’s (2013) study was concerned with a large research organization, CENS. Wallis et al. suggested science data are regarded as “goods [which] are bartered between trusted colleagues rather than treated as commodities” (p. 1). Relatedly, the current study sought to determine whether the same factors that inhibit data sharing as identified by Wallis et al. also apply to small research organizations, individuals, and CS groups. Moreover, strategies were sought to break down the silo mentality and provide incentives for sharing science-data sets more widely. Wallis et al., to their credit, recognized science-data platform availability in itself does not fully incentivize scientists to share their data sets, and other motivators are a prerequisite. Wallis et al. also implemented a survey based on interviews and ethnographic observation from 2002–2012 at multiple sites associated with CENS and found 75% of scientists were willing to share their data and were strongly willing to use data from other projects if those data were easy to retrieve, which was generally not the case. Further, Wallis et al. survey respondents stated, if the infrastructure and incentives to put forth that effort were readily available, they would be more inclined to share their data.

Wallis et al. (2013) also found the bulk of science-data gathering funding goes to smaller projects, to which they referred as “the long tail” (para. 16)—presumably because

these projects are disparate and their data are difficult to locate due to limited data sharing options. This funding allocation is both a problem and an opportunity, because if copies of long-tail data were located on an appropriate technology platform with sophisticated meta-data search functions, those data could be more readily found. Consequently, this availability would maximize the inherent value from recognition of the effort to collect, collate, and interpret through appropriate commercializing mechanisms. Both of these themes were analyzed in this dissertation.

Wallis et al. (2013) left many questions unanswered, such as who should fund the development of a technology platform, who would curate, how this platform would be funded on an ongoing basis, and what incentives would help share data sets by data content owners. These questions are not new, and social network stakeholders have actively discussed these areas, as evidenced by the current mainstream media discussion concerning shadow banning accounts by Twitter and Facebook (e.g., the account of President Trump; Minot et al., 2021).

Potential Market Disruption Strategies for a CS Technology Platform

Resistance to change is expected. This section discusses strategies for market disruption and sustainable business practices within the CS community.

Market Disruption

For a business model to succeed, it needs compelling motivation to overcome legacy industry inertia. The science community is not generally known for entrepreneurship, although it is characterized by innovation, as evidenced by huge strides in engineering, information technology, and more recently, biological sciences spurred by

economic and humanitarian demands for a solution to the COVID-19 global pandemic. Christensen (2016) discussed the paradox in which successful managers find themselves when confronted with a disruptive innovation in conflict with their legacy business model, making their experiences incompatible with a new approach and leading to resistance to a new concept. Christensen emphasized more successful market disrupters do not compete with traditional market leaders head on, but first satisfy unmet needs of peripheral market participants, who place a greater value on the disruptive technology and are consequently more receptive to it. Christensen concluded disruptive technology should be framed as a marketing challenge rather than a technological challenge.

For the specific case of citizen scientists and CS in the underdeveloped world, consideration must be undertaken to identify their unmet needs so a new value proposition can be defined that leads to rapid market acceptance of the disruptive business model created by the technology platform—and which does not conflict with their legacy information systems, if any. Identifying these unmet needs should be the central aim of any analysis to create a successful business model by asking: ‘Are the unmet needs of underdeveloped countries significantly different from those of the developed world?’. This question highlights an important aspect that must be addressed if a global science-sharing technology platform is to be established. Hart (2018) mentioned, “Any serious attempt to understand an aspect of the world is almost inevitably based on some dimensions of the intellectual tradition of Western knowledge” (p. 72), referring to Marx’s analysis on capitalism and the nature of people as “creative adaptive beings with unrealized potential, trapped by social forces that disempower/exploit” (p. 80). If a

capitalist-oriented business model for science-data sharing is established, would it gain traction in the undeveloped world given the barriers inherent in the society structures?

The seminal work of de Soto (2000) explored this issue at great length, and was referenced and discussed earlier within this research. Would providing a zero-cost, but sophisticated, cloud-based data storage application satisfy citizen scientists' unmet needs? Is the central unmet need of CS the ability to conveniently share and borrow data sets, as identified by Wallis et al. (2013)? Is there a motivation by CS to obtain recognition for their work? Is the ability of CS to generate a modest revenue stream to offset their data collection time—costs a motivation? These are some questions concerning unmet CS needs that those involved in the field must answer. As Christensen (2016) concluded, a powerful protection of disruptive technology is to undertake something that simply does not make sense for established industry leaders to compete against.

Science-Data Sharing in Underdeveloped Countries

Many barriers to science-data sharing in underdeveloped countries are not acknowledged in developed countries; for example, many parts of Africa currently have nonexistent or, when available, slow internet connectivity that is not conducive to transferring large files; these regions may also be subject to electrical grid disruptions, which negatively impact productivity. East Africa has been one of the world's most "digitally excluded" regions, with just about 2% of the population connected to the internet (Echezona & Ugwuanyi, 2010). Echezona and Ugwuanyi (2010) stated the situation was ameliorated somewhat with the advent of improved electrical infrastructure and deployment of satellite technology and construction of 9,900-km fiber optic system.

Aside from basic infrastructure issues, de Soto (2000) also identified another dimension that negatively impacts science-data sharing in underdeveloped countries and that is security over intellectual property rights. Notably, de Soto's (2000) work is important because it sought to understand, as reflected by the title of his text, *Why Capitalism Triumphs in the West and Fails Everywhere Else*. Based on over a decade of extensive research across the world, de Soto did not foresee the rapid rise of China's version of a capitalist economy; nonetheless, his work is important, as it described barriers to economic advancement in the underdeveloped world. He claimed capitalism is the only feasible way to rationally organize a modern economy due to its inherent complexity and the failure of central planning. His conclusion was *capital* is the "force that raises productivity of labor and creates the wealth of nations" (de Soto, 2000, p. 5); capital cannot be created without security of title over assets, and, consequently, it cannot alone be used to create surplus value. The relevance of this profound insight for sharing science data in the undeveloped world is there must be security of copyright and attribution, a factor often taken for granted in the developed world where formal property rights are ingrained.

Consequently, any platform technology solution for science data sharing must assure all parties their intellectual property (IP) rights are securely established and maintained irrespective of the country of their domicile. This assurance establishes the asset's potential to become fungible, create surplus value, and enhance the production of capital, thereby creating wealth and becoming an incentive for the underdeveloped world's participation in science data sharing. In practical terms, in an underdeveloped country, this extrapolation of de Soto's (2000) insights would facilitate the establishment

of a cottage industry of citizen scientists gathering data sets at relatively low cost compared to costs of data gathering in the developed countries. A networked science-data sharing platform potentially free of direct costs for participation would enable users to monetize their efforts with a suitable business model based on market demand for access to their IP, while also providing a public good to the global commons and their local community.

Sustainable Business Practices

A discussion on sustainable market dominance cannot be separated from sustainable business practices. Cici and D’Isanto (2017) reviewed literature on sustainability and interviewed business leaders on their definitions and strategies for sustainability. They concluded a modern definition is to be able to “guarantee long-term profitability” (Cici & D’Isanto, 2017, p. 54) and a key component in achieving this profitability is creating social value. What has become clear is technology platforms require innovative business models to remain competitive through market differentiation while creating this social value, which must become a high-visibility feature.

For a science-sharing technology platform, this definition by Cici and D’Isanto (2017) becomes self-evident for both sides of the transaction. For those wanting access to the shared data, their social value hinges on the ease of search, quality rating of the data sets, visualization of the data sets, and ease of download for personal use. For data content providers, their social value includes tools that allow them to upload, securely store, update, edit, and visualize their own analyses; use the ease of publication of selected data sets to the commons; and benefit from the ability to publicize their projects (Wallis et al., 2013). This definition morphs into the function of Facebook, a platform

that uses the content of others to generate revenue. However, considering societal change for Web 3.0 as discussed by Garrigos-Simon et al. (2012) and increased expectations by the content provider to share revenue, the Facebook model may not be sustainable in the long term unless it modifies its business model.

An important note is with Facebook's name change to Meta (short for Meta Platform) in late 2021, the platform does indeed appear to be modifying its business model to reflect Web 3.0 expectations. In a survey of business leaders on the topic of integration of sustainability in business, Cici and D'Isanto (2017) identified 68.3% of leaders responded integration of sustainability in business was "measuring and monitoring environmental, social and economic impact on business" (p. 55), and evolving societal impact on the expectation of enhanced social benefits and revenue sharing is anticipated to grow. Consequently, to ensure complete social value is attained, stakeholders in Web 3.0 business models must formulate a perceived equitable approach to sharing business revenue while guaranteeing long-term profitability.

There are several economic funding models for revenue generation. Pounds and Bostock (2019) provided a useful summary of common revenue models and concluded excessive reliance on donor-driven resource models (e.g., grants) inevitably leads to the collapse of the business model once the donor support ends and does not secure long-term sustainability. Pounds & Bostock instead appeared to recommend a more eclectic, revenue-driven model.

Platform Market Entry

An important and highly relevant insight to this dissertation research was a networked market is more hostile to innovation than a nonnetworked market. The

rationale behind this insight was attributed to the Nobel prize winner Nash (1950), who identified markets crave equilibrium and every player believes they are making the best possible decision; consequently, any innovation in a market, by definition, upsets the equilibrium and is met with resistance. The larger this networked market, the greater the resistance to innovation, based on Metcalfe's power law that a network's value is proportional to the square of the number of users (Metcalfe & Boggs, 1976). This insight into the value of networks was originally applied to telecommunications networks, and subsequently formulated by the futurist Gilder (2018), albeit still attributed to Metcalfe. This resistance to innovation by networks should give pause to an innovator who wants to break into an existing strong market network.

Consumers are inundated with marketing messages. Chakravorti (2004) examined the difficulty in bringing new product or services to market and provided a refreshing perspective on strategies to consider for platform-technology market entry. These strategies are relevant for a CS technology platform tasked with persuading networks of CS groups to use a new and innovative platform.

Reason Back from a Target Endgame

A strategist is required to envision a plausible new market equilibrium their innovation will create and then focus exclusively on strategies that support the accomplishment of this new equilibrium. Such an approach would require dropping all weaker strategies, which optimizes marketing efficiency. For example, if the strategist anticipates copycat competitors, it behooves them to promote their service's branding as an integral component of their marketing strategies.

Complement the Power Players

The strategist needs to identify the power players in the network and share the value proposition with them, being careful not to be perceived as competing, but instead as complementing their market position. An example within the CS global networks would be to identify major national CS organizations and approach them with a complementary strategy that creates value, such as increasing their membership level or generating incremental revenue. Other key power players are environment instrument makers, who could be approached to offer a differentiated service (e.g., providing a data storage facility for their customer's environmental data 'free with every purchase').

Offer Coordinated Switching of Incentives

Traditionally, marketers seek to persuade early adopters of the superiority of their new innovation, incrementally convincing the total market. According to Nash's (1950) network equilibrium theory, this is not the best strategy for networked markets, as they will tend to revert to their prior equilibrium state. Chakravorti (2004) postulated the optimal strategy is to concurrently convince market participants that the decision criteria (for optimal self-benefit) have changed across the whole network. This convincing is difficult to accomplish, as it requires the identification of key players or market makers within three core groups that closely interact and provide positive feedback to each other on their decision making: (a) the players that add to the innovator's benefits, (b) the players that act as channels to adopters, (c) and the adopters themselves.

Preserve Flexibility

The strategist needs to have flexibility in marketing plans, of which pricing policy is the most common. Chakravorti (2004) gave an example of Adobe, where its Adobe Reader initially had a license fee, but the company subsequently decided to make its Adobe Reader license free by charging only for its Acrobat Creator program. Adobe provides another example of subsidizing one side of the two-sided platform, as discussed prior.

Dethroning a Platform Incumbent

The competitive strategy of differentiation by providing a superior product or service is not new. Suarez and Kirtley (2012) argued it is difficult, but not impossible, to “dethrone a platform incumbent” (p. 36) by moving the value proposition from what is valued today, to what the market will value tomorrow. This concept is refreshing due to the high level of high level of entrepreneurship and associated risk for a market incumbent in not making a correct assessment of future market needs. An example for the science community would be an assessment by inadequately funded organizations that are “data rich” and wish to monetize their data rather than provide their data freely—this process reflects the current expectations of government-funded data exchanges and nonprofits. This is not to say the data owner would necessarily charge a fee to the recipient of their data, but they may instead seek an alternate platform that provides an indirect monetary reward while also providing a frictionless (i.e., without price) data exchange experience to both sides of the two-sided platform. Such a transaction strategy would have the potential to dethrone the incumbent platform that expects free exchange of data, regardless of the inherent costs of facilitating that data exchange by the data

provider. Instead of just a transaction cost, the data provider would receive a financial reward, which would be further reinforced by its potential cost-free replication due to the long-tail effect, and which has the potential to fully eliminate the costs of data exchange and hopefully even monetize their data assets further. This scenario has the potential to dethrone an incumbent platform that would otherwise be prohibited from applying such a strategy through a policy of not being able to monetize their data assets due to having been funded for the creation of their data asset via the tax payer.

Market Failure: The Chicken or the Egg Conundrum

The core problem of the “chicken or the egg” conundrum is that the value proposition is established when the other party to the transaction is already present. For instance, for a researcher seeking data sets on a science-data sharing platform, a prerequisite for creating value to the researcher is that there is a critical mass of diverse and attractive data sets already published by data set providers. The converse is also a requirement, in that for data set providers to find it worthwhile to undertake the time–cost of uploading and publishing their data sets to the platform; they will want to be persuaded that there already exists a critical mass of researchers who will search for this data. Hence, this issue reflects the “chicken or the egg” conundrum. This conundrum is a platform-specific problem, significantly more prevalent and fundamental than was originally realized by early platform developers as a barrier to establishing a successful market-based platform. Significantly it was not mentioned by Chakravorti (2004) who wrote extensively on platforms, even though this conundrum remains a formidable barrier that all platform businesses must overcome before launching a two-sided platform. This problem became increasingly apparent as being acute and fundamental with various

researchers discussing it at length a decade later, such as Costa (2016), Moazed and Johnson (2016), and Thiel and Masters (2014).

The resolution of the chicken or the egg conundrum is particularly important for CS networks given the consumer half of the CS platform transaction becomes more valuable the greater the number of users or size of the network; Metcalfe's power law emerges again in this context. How does one establish a market platform that rises above that critical threshold of value for either side? Parker et al. (2016) admitted they have not found a conceptual strategy for solving this practical problem, arguing competitors in the same market may find different solution routes; however, they were sufficiently insightful to list eight potential strategies that could work. Clearly, a solution must be found as a prerequisite for platform success. Moazed and Johnson (2016) defined seven strategies, but their more helpful provision was they identified three categories of strategy to subsidize value and solve the chicken or the egg issue: monetary subsidies, product feature subsidies, and user sequencing.

Monetary Subsidies

The most direct approach to monetary subsidy is giving money away to a producer or consumer, which typically manifests as providing a referral fee most platforms use to attract new users. Uber is a common example of this business subsidy, and PayPal used this strategy in the early 2000s to achieve growth rates of 7%–10% daily (Moazed & Johnson, 2016). Another strategy is to reduce the price possibly to below cost for one side of the transaction. In Uber's early days, drivers were provided with minimum earning guarantees to ensure there was a critical mass of drivers available in a city so customer wait time was minimal. For a science-data technology platform, this strategy

may manifest as providing a guarantee of a minimum revenue stream that covers the data take-up costs of content providers. The NIH employed this strategy for its Africa Open Data Science Platform, funding six data hubs to collect and submit data to the platform over an initial five-year period budgeted at \$45 million, which will probably be extended for another five years (Glass & Tromberg, 2019). This strategy is expensive, and almost guarantees only government authorities with extensive taxation revenue have the depth of funds to implement and take the financial risk.

Product Feature Subsidies

Twitter uses the verified user strategy to identify key influencers or celebrities with a large following who then attract other users to their platform. This verified user program provides additional software features, such as the ability to tweet exclusively to other such users, which have been remarkably powerful tools for world leaders to rapidly communicate public health information with citizens (Rufai & Bunce, 2020). This tactic could be useful for a CS community with a relatively low number of high-value content contributors, and could also be used as a mechanism to guarantee the quality of the data set shared.

User Sequencing

The strategy of user sequencing involves prioritizing a specific high-value group by possibly using the prior two categories of subsidy, a move expected to attract others to the platform. The objective is to reach the point of providing sufficient platform value at which positive network effects commence to drive future growth. Tinder is a strong example of this strategy. In its early days, Tinder's cofounder, Whitney Wolfe, met with

female sorority groups to enroll them on the platform. She then gave a presentation to a corresponding male fraternity group, who saw these attractive young women had already signed up. Tinder initially had 5,000 users; after Wolfe's marketing trip, they had more than 15,000, reaching their tipping point for sustainable network growth (Moazed & Johnson, 2016). Tinder accelerated to 1 million monthly users within the 1st year in 2012, and had a market valuation of over \$40 billion in 2021 (Nasdaq, n.d.). Wolfe's Tinder strategy could potentially be used within the U.S. school system to entice school student projects to enroll and introduce them to similar groups in other schools around the world. Millennials and Generation Z have shown a propensity to form virtual social groups and are highly influenced by opinion leaders on social media (Nesi et al., 2018).

The Core Transaction

The identification of a platform's core transaction is an essential feature that has been rarely discussed in the literature. The core transaction is the event that binds the loose ecosystem together and is more important than generally realized. Platforms make a mistake if they focus on multiple core transactions in their launch phase, which is not a sound initial strategy.

Ecosystem Binding

The arrangement of organizations around a platform is referred to as *ecosystem binding* due to its mutual benefit of association, which differentiates its governance from other governance forms (Jacobides et al., 2018). This organization of entities around a multisided platform has a central platform coordinator who sets standards, roles, and accountability, creating a loose amalgamation of complimentary entities with a common

interest and without formal or legal binding. Jacobides et al. argued that complimentary entities have an investment in participation that is not fully fungible, although there will be a cost in departure from this ecosystem, and this is a fundamental structural feature of ecosystem interactions that make them strategically distinct. This investment by complimentary entities helps the platform strategist understand the core transaction enabling the technology platform to provide value and the binding to the ecosystem. As discussed earlier, a platform's biggest resource is its network, which is why every effort must be made to build its network of complementors, who then create ecosystem binding.

Value Ecosystem

A platform can be considered an ecosystem of transactions of varying importance to each side of the platform. Moazed and Johnson (2016) used similar terminology to Jacobides et al. (2018), referring to the platform as having a value ecosystem composed of a set of primary activities that directly create value along with a set of secondary activities that support the value creation. For a science-data sharing technology platform, the primary value proposition may be the ability to publish selected data sets to the commons easily, which provides the greatest value to data content owners. The analogous secondary feature could be enabling the ease of upload of data sets, the editing, the online visualization of the content owner's data sets, or reporting on how popular data sets are with the global community and consequently providing a feedback mechanism.

The other side of a platform is the data consumer. The data consumer's primary value proposition may be the ability to search for published data sets together with project background information and its data quality certification. Conversely, their secondary activities may involve the ease of visualizing selected data sets online (a

common secondary value among content owners), and also the ease of download of published data sets that meet their needs. If this assessment of the primary value proposition for the CS data consumer is correct, extra effort must be made to provide an efficient search interface, as that single feature will be the primary criterion judged by CS consumers. This core transaction between data content owners and data consumers (i.e., being able to publish and search) has been recognized as the primary value creation for a science-data sharing platform; as such, it must be efficient and economical to ensure continued commitment to the ecosystem by both parties in the transaction. Moazed and Johnson (2016) emphasized that getting this core transaction correct is the most essential aspect of platform design, and defined it as the set of actions consumers and producers must complete to exchange value. They emphasized a platform does not own the asset transacted but, rather, is the facilitator of the transaction, a characteristic that differentiates it from a linear business.

Importance of Simplicity of Core Transaction

Is there value in multiple core transactions? Moazed and Johnson (2016) strongly recommended against that strategy, at least in the early days of establishing a platform business, claiming it confuses the market and diffuses the effort to dominate a market segment. All major technology platforms have “initial simplicity” in common (Moazed & Johnson, 2016). As recounted by Kirkpatrick (2010), Zuckerberg was particularly noted for his insistence on simplicity in the early release of Facebook, where the core transaction was initially just the ability to invite more friends. Early dating platforms experienced problems from not focusing on the core transaction; for instance, OkCupid

and Match.com initially required both parties to send a message of interest to the other user to match, effectively constituting a double transaction.

Compare those early transactions to the simplicity of Tinder, where both parties simply have to swipe right to connect. LinkedIn commenced in 2002 with a simple profile that allowed professionals to connect with others, and it was not until exceeding 2 million users (Moazed & Johnson, 2016) that LinkedIn started building other significant features into their core transaction. Moazed and Johnson (2016) explained that the core transaction must be as simple and straightforward as possible and that the benefits remain highly apparent to the transacting parties. Further platform sophistication should be explored as a subsequent phase rather than incorporated into the launch phase.

Complexity in Search Matching

Once the initial simplicity of the core transaction provides success in attracting both sides of the two-sided platform and scale is achieved, platforms with large content providers and large numbers of content consumers will face an evolving problem as described by Malgonde et al. (2020), which requires making searches and matches more effective. Malgonde et al. commenced to visualize the issue with the well-worn example of the ‘naked hairy man issue’ (Costa et al., 2016; Parker et al., 2016), giving the example of the OKCupid online dating platform, where the search parameters by men were originally one-sided; to solve that issue, a two-sided search mechanism was established where the search parameters of the women had to match those of the men searching (i.e., the reciprocal recommendation). As a consequence, a better “fit” was made in the matching process.

Malgonde et al. (2020) went on to describe the more sophisticated needs of an online educational platform matching large numbers of students with considerable selections of courses from diverse educational institutions. The problem they raised was both sides of the two-sided platform continue to evolve; as a consequence, the matching process is complex (in the mathematical sense), leading to suboptimal matching. Another example of this evolution of search needs is a student who advances through a course of study. As the student's grades become more advanced, and the courses offered by competing institutions also continue to evolve with new courses being introduced, the search criteria used to match the student's evolving needs against newly introduced courses also needs to evolve to become more sophisticated. The solution to increasing complexity is to establish what Malgonde et al. termed a *recommender system*, noting platforms are complex sociotechnical systems. These platforms require an algorithm to model the emergence or evolution of each side to the transaction to gain optimal matching, updating its recommendations in each search iteration responding to the changing environment. Empirical results by Malgonde et al. demonstrated the two-sided recommender system, as with the example of matching students with courses, improved students' and universities' fitness by 50% (buyers) and 87% (sellers), respectively, compared to the base case of no recommendation. Consequently, this level of search sophistication is justified for a mature platform with large volumes of participants with evolving needs.

Summary

Deficiency in the Market for the Sharing of Science Data

Through the literature review, a significant deficiency in the market for science-data sharing and a major opportunity for a market-disrupting business to fill the void left by government-dominated decision making were identified. The solution strategy was based in part on leveraging a technology platform, a new field as evidenced by the plethora of definitions of what comprises a technology platform. Further, many factors can cause this strategy to fail, requiring astute business analysis and thoughtful strategy. The fact remains that the business world's most successful businesses by market capitalization are overwhelmingly based on technology platforms.

The other major factor to achieving a solution for sharing science data was to harness a new business model applicable specifically for technology platforms to incentivize both sides of the two-sided market and minimize the friction of participation. Science-data content owners need to have incentives to overcome their time–cost of data contribution to the platform, and content consumers need incentives to participate in the market for searching and retrieving science data, the core transaction. These needs are further complicated by the prerequisite to reach a critical threshold of content, the chicken or the egg conundrum all platforms must overcome.

Is This New Business Model “Pandemic-Proof”?

The world faces the ongoing COVID-19 global pandemic and biggest economic crisis since the Great Depression. Is this a good time to launch a new global business? In this dissertation, the very significant intellectual efforts and human resources that contribute to the rapidly growing field of CS each year were identified. One could ask whether there will be a slowdown in government funding initiatives for CS projects, but that would be the wrong question, because this study has determined there is no shortage

of CS project initiatives with legacy data sets. The problem is in identifying a commercial business model and a marketing strategy that will provide the appropriate incentives to reward science-data content owners to participate, combined with a rich content user experience while ensuring mutual and equitable long-term profitability.

The COVID-19 global pandemic has significantly negatively impacted government revenues and placed into question the ability of groups to physically interact. Clemens and Veuger (2020) estimated COVID-19 led to a shortfall of roughly \$106 billion in U.S. sales and income tax revenues for the third quarter of 2020 through the second quarter of 2021, which is 11.5% of pre-COVID-19 sales and income tax projections. This shortfall is anticipated to negatively impact the funding of CS projects within the United States and elsewhere in the world (Clemens & Veuger, 2020). Counterintuitively, the framing of an appropriate business model to solve the issues identified in this literature review should not be significantly impacted by the COVID-19 global pandemic, given the very significant initial requirement to undertake data take-up of legacy data sets to reach the threshold of usefulness for content consumers. Data take-up is almost always an individual rather than a group activity, and consequently is not impacted by social isolation rules. Indeed, it may motivate CS to realign objectives away from field work, prioritize the most pressing social injustice issues, and generate a greater interest in CS through studying legacy data sets, while providing access to a CS technology platform that will greatly facilitate data sharing and scientist-to-scientist interactions.

CHAPTER 3: METHODOLOGY

The research methodology for this grounded theory study included elements of action research. The approach allowed the study of both dynamic, platform-based businesses via the literature review and conservative organizations via questionnaire, active participation in industry focused events, and personal interview. The latter method sought out participants' unmet needs while coaxing potential acceptance for a paradigm shift in their operations toward replicating dynamic aspects of successful platform-based businesses. There were two broad research questions:

1. What are the barriers that impede science data from being shared with the global science community?
2. What are the critical attributes of a market-driven technology platform business model whose mission is to facilitate science-data sharing and satisfy unmet science community's needs?

Research Method

The selected methodology was qualitative, interviewing different categories of user groups, participating in group discussions with special interest groups, then comparing and contrasting their unmet needs to seek a common solution to the problem of facilitating science-data sharing gathered by disparate groups. The key component of the solution set was to determine whether a technology platform has the potential to address common unmet needs and whether such a solution also has the potential to be commercially viable, given funding of science-data collection projects is often localized in nature in their data storage systems, which is not conducive to global data sharing. The

potential for a suitable reward system, both financially and through recognition, had not emerged in existing literature, and the methodology of this study was developed to explore this potential as another key component of the solution set.

This research was exploratory and context driven; the setting and participants' frames of reference or worldviews were important to capture. The nature of the questions was subjective and evolved as knowledge was gained. This question structure led to the selection of a qualitative methodology (Marshall & Rossman, 2016).

Research Design

In this study, government-funded information systems were considered to seek out a new paradigm for the business model for data sharing by citizen scientists in particular and scientists in general. Abdel-Fattah (2015) found there is evidence that “an interpretive paradigm is more suitable for evaluating e-government systems” (p. 310) and recognized more than one design method can be used, ultimately deciding to focus on the two that felt most appropriate for information systems (i.e., grounded theory and action research). Abdel-Fattah concluded each had a valid role, and action research had a superior advantage for developing or improving a case study; yet, he also concluded for e-government systems specifically, grounded theory was superior due to the stability of systems under review in that they tend not to change in a short time. Grounded theory is superior in formalizing, interpreting, and analyzing data and encompasses classification of data into categories to postulate networked relationships among these categories.

Abdel-Fattah (2015) also concluded, in his context of a single e-government system at an Egyptian university, that action research is not able to evaluate case studies that require evaluation of systems with more than one perspective; however, the topic of

this research study involves multiple e-government systems in various jurisdictions. Consequently, this focus made action research methodology less efficacious and tended to preclude it. In deciding between which methodology (i.e., grounded theory or action research) to use in this study, the following specifics were considered.

First, very little research has been conducted on the possible solution set for using a technology platform to satisfy the needs of citizen science (CS) participants applying a commercial market business model; consequently, a qualitative study was the most appropriate approach to better understand and resolve important issues (Creswell & Creswell, 2018) and test the theory generated. A qualitative study has more latitude to enable a more innovative approach and be more transformative.

In this study, a new paradigm was sought, and appreciative inquiry techniques were found to be valuable for selected interviews to enable the interviewees to visualize a possible future unfettered by existing constraints, such as funding or political barriers to the free flow of information. However, because of the cultural and jurisdictional diversity of organizations that were reviewed, a step back from collaborative analysis was required. As such, a general assessment of needs was made, based on a consolidated assessment of unmet needs, a feature more akin to grounded theory.

During the data-collection phase, conducting in-depth interviews in interviewees' personal spaces often generated significant insights that were not always evident from existing literature. These insights provided tacit knowledge, which was difficult to transcribe and had a higher likelihood of emerging during the interview process. Charmaz (2014) praised the grounded theory methodology when an interviewer sees past this

particular method, “combined with insight and industry” (p. 26), and where the method provides “flexible guidelines rather than rigid prescriptions” (p. 26).

The personnel who undertake these tasks in their day-to-day working environments had the firsthand, in-depth knowledge of the issues being studied. The skills required in applying grounded theory was to coax them into surrendering these insights and knowledge, many of whom imparted such information willingly if the correct questions were asked under viable conditions, such as authorship recognition or anonymity (Wallis et al., 2013).

Instruments

The instruments for the analysis included both questionnaires and interview questions, primarily with executives from CS organizations and from leading academics in the field of community health in underdeveloped countries. The questions evolved as the issues impacting the industry were identified (Charmaz, 2014). Commencing questions were grouped corresponding to the two primary issues being researched:

1. What are the barriers that impede science data from being shared with the global science community?
 - 1.1 Determine whether it is the political/administrative boundaries of the jurisdiction that is funding the project that creates artificial barriers to sharing the data outside of those boundaries.
 - 1.2 Explore the motivational factors that exist for initial data gathering.
 - 1.3 Identify what extraneous incentives exist to share the data gathered outside of the immediate community.

- 1.4 Identify what mechanisms exist for coordination of environmental data gathering projects with the broader community.
- 1.5 Identify what mechanisms, if any, there are for the broader community to gain access to project data; and identify whether these mechanisms tend to be effective and efficient.
- 1.6 Identify whether additional incremental funding would induce projects to continue and expand their data collection efforts.
- 1.7 Explore whether there are unique barriers to citizen science participation, particularly by developing countries.
2. What are the critical attributes of a market-driven platform technology business model whose mission is to facilitate global science-data sharing and satisfy unmet science community needs?
 - 2.1 Discover whether the CS community in particular would support the concept.
 - 2.2 Explore motivational factors that would encourage the CS community to participate in a platform technology.
 - 2.3 Review mechanisms used by other platform technologies in overcoming the chicken or the egg problem of surpassing the critical threshold of content that encourages others to participate in a two-sided platform.
 - 2.4 Discuss mechanisms to provide a combination of financial and recognition motivators that encourage ongoing platform contributions on a commercial basis.

The questionnaire and interview analyses were coded, and themes were drawn from these analyses. Other instruments included the analysis of requirements published by the U.S. National Institutes for Health's (NIH) Africa Health Discovery Open Data, as these documents represented a detailed official viewpoint and requirements of big government.

Participants

The focus was on scientists in the field of environmental and community health science. Within this domain are the following subdomains:

- 1) Citizen scientists
 - a. Nonprofits that are primarily government-funded or subsidized through donations
 - b. Individuals who are hobbyists
- 2) Academics
 - a. School groups that have environmental science as part of their curriculum
 - b. University students who study environmental science
- 3) Government entities whose mandate is to collect environmental data; one example is the U.S. NIH, which seeks to identify environmental health hotspots (Glass & Tromberg, 2019)
- 4) For-profit entities that collect environmental data during the course of business operations for government environmental compliance permission and/or to support business operations

All four subdomains are global to varying degrees. The first three groups were prioritized because they tend to overlap in their mission and outlook: citizen environmental scientists, academics in environmental science, and community health scientists, together with a large government entity with global aspirations, the NIH, who had a timely environmental health project spanning all of Africa.

Grounded theory (Charmaz, 2014; Creswell & Creswell, 2018) was used and identified the interviewee list by purposive selection (i.e., snowball sampling; Simon & Goes, 2018). The sample size was not unduly limited, but continued until saturation (Creswell & Creswell, 2018), which occurs when no further significant understanding can be acquired. Consequently, a number of group meetings were held via Zoom and individual interviews were held via telephone or Zoom, as well as 15 detailed interview/questionnaire responses from chief executive officers (CEOs) from diverse countries.

Receiving only 15 interview responses was disappointing, however the excellent quality and comprehensiveness of the responses provided balance. The reason for the poor response rate has been attributed to the number of in-depth questions which would have been time-consuming to respond to. Charmaz (2014) made the case for a smaller number of in-depth interviews but also concurred the number of interviews depends on an interviewer's objective to identify only themes or to gain a more fundamental understanding, stating, "Twelve interviews may generate themes but may not command respect" (p. 105). In this dissertation, the business model was defined for a specialized platform technology (i.e., software application) by reviewing different perspectives through many in-depth interviews and group meetings to ensure comprehension of issues.

Interview selection requests for possible participation included the following:

- the executives of national CS organizations in the United States, and other countries, together with significant state representative organizations;
- heads of faculties and postdoctoral research fellows at universities from both developed and underdeveloped countries;
- organizations that have global health monitoring mandates; and
- for profit and not-for-profit entities that gather environmental data for monitoring health impacts on society.

The identities of the 70-plus participants in the NIH Open Data Africa virtual conference were gathered. These participants represented a broad range of government, academic, not-for-profit, and global for-profit organizations that were primarily Africa-based, and they were targeted for group interview meetings. In addition, a list of all U.S.-based CS organizations and a list of non-U.S. organizations were canvassed for detailed interview and questionnaire participants.

Data Analysis Methods

The sample size for detailed interview and questionnaire solicitation was over 500, and 15 CEO questionnaire responses were received. Solicitation responses were then coded manually and grouped into themes, with further analysis of particular direct questions. This thematic analysis and its coding helped discover implicit meanings, explore additional directions, and “suggest emergent links between processes in the data to pursue and check” (Charmaz, 2014, p. 121). The generation of themes was an important phase wherein common issues were identified and considered further. Charmaz

(2014) provided a thorough understanding of this process, identifying that the coding process served as a pivotal link between collecting data and developing an emergent theory of explanation. Triangulation was an important facet of both the questionnaires and interview analysis, particularly as there was representation from both developed and underdeveloped countries—an essential aspect of this research.

Limitations

Science-data sharing, specifically CS, is a very broad and diverse subject area with a global audience. A limitation of the study was gathering representative opinions that reflect this diversity in science topics and among different countries to represent developed and underdeveloped economies, given many underdeveloped countries discourage the sharing and publication of environmental data (e.g., Kenya Data Protection Act of 2019). Significant study limitations of concern were:

1. All participants (i.e., environmental data gathering groups) were funded through government grants and donations contingent on satisfying government policy (e.g., investigation of clean water, habitat sustainability). This funding structure limited the pool of possible entrepreneurial data gathering groups to interview.
2. The study scope was not limited to the developed world, and outreach to CS groups from underdeveloped countries was also undertaken.
3. It was assumed, excluding funding constraints, barriers to sharing science data among professional groups in the developed and underdeveloped world are the same, although there are politically contrived barriers in some countries.

Delimitations

A boundary condition was the focus on scientists and citizen scientists in the field of environmental science. CS was used as a proxy for the greater field of science. A modest delimitation of the study was it was predominantly researched in the English-speaking world, although that is not an overwhelming delimitation for the science community; in some science fields, 90% of the publications are in English and the global science community has considerable English-language representation (Fung, 2008). Interviews were implemented with a principal management representative of a number of CS organizations. The questionnaire was limited to 25 questions in number and abided by the following conditions:

1. A boundary condition was this study focused on the scientists and citizen scientists in the field of environmental science and community health.
2. Another boundary condition was all citizen scientists from the United States had the same or very similar viewpoints on knowledge transfer. Consequently, the analysis of results was assumed to be generalizable to all U.S. participants.
3. Only subject groups who are at least moderately fluent in English were researched; therefore, the findings and results may not necessarily be generalized to other educational groups outside of the United States, nonscience subjects, nor countries that are not defined as western in outlook or non-English speaking.

Trustworthiness

The study participants were all executives at science-data collection organizations and/or senior academics within reputable universities and research institutions.

Trustworthiness of the study was achieved in the following ways:

1. Multiple sources of data were collected, quite separate from the interview process, as an aid to triangulate themes. This included active participation in a number of special interest groups via zoom that were webcast by the federal government and which featured presentations by government experts from the National Institute of Health (NIH) with audience question and answer sessions. Other sources were participation in various proposal writing with industry experts responding to government information requests to share science data. These sources of data were exceedingly valuable in assisting the framing of subsequent interview questions.
2. Rigor was established through a compilation of appropriate interview questions based on the questions posed in this dissertation and in parallel studies in the literature.
3. Accuracy was enhanced by ensuring interview notes were formerly transcribed as soon as possible after the interview.
4. Reliability of the interview notes was established through an ongoing process of comparing and contrasting various viewpoints until a consensus was reached through an iterative approach, and the process terminated when there was saturation or no new significant knowledge.

5. The methodology of the research was presented in the findings to enable study replication.

Summary

This study sought to identify a market-driven approach that emulated commercially successful technology platforms to define a disruptive technology platform strategy that motivates science-data content owners and science-data content providers to proactively share their data commensurate with suitable market-driven reward mechanisms. This market-disrupting business model to the existing government-funded fiat data collection programs has the potential to use market-demand forces and create a rich pool of diverse science data sets to be published to the commons. This objective is expected to result in more accessible and more comprehensive science data sharing, compared to traditional government-sponsored programs and dictates.

This project sought a transformative change. A qualitative methodology, grounded theory, was used to identify and reflect on the issues and then arrived at a recommended solution set. This solution set will take the form of a technology platform with appropriate business rules that facilitate participation and rapid market growth, a hallmark of successful technology platforms. Interviews were conducted and questionnaires administered with a range of CS organizations from both developed and underdeveloped countries to explore issues (e.g., barriers to data sharing and motivation to participate in science-data sharing) to identify common themes.

CHAPTER 4: FINDINGS

In this chapter, the raw data obtained from the survey issued to chief executive officers (CEOs) of citizen science (CS) organizations are presented. The CS sector was used as a proxy for the broader science community because the sector tends to comprise a relatively homogenous group that is easier to identify, particularly when seeking representation from both developed and underdeveloped countries. Over 500 CEOs of CS organizations were contacted, and 15 responded to the 22 in-depth questions. Although this was a relatively low response rate, the caliber of the responses from the CEOs was excellent and crystalized the study's themes. The questionnaire spanned the scope of the literature review, with some questions asking the inverse of a prior question to verify the veracity of responses. Charmaz (2014) made the case for smaller numbers of very in-depth interviews and questionnaires and advised the number of interviews depends on an interviewer's objective to identify themes or to gain a more fundamental understanding, stating, "Twelve interviews may generate themes but may not command respect" (p. 105).

The two primary categories of research questions were:

1. What are the barriers that impede science data from being shared with the global science community?
2. What are the critical attributes of a market-driven platform technology business model whose mission is to facilitate global science-data sharing and satisfy unmet science community needs?

Presentation of Findings

Detailed responses from the CS CEOs are presented in Appendix A. The questionnaire results provided some particularly strong issues that are numbered according to the questionnaire:

1. There frequently exists political and administrative barriers that inhibit the sharing of data and span jurisdictional constraints and internal reasons, such as embarrassment over lack of data organization or concerns the data would be misinterpreted by external parties.
2. There was a strong feeling of camaraderie and altruism in stepping in to embark on work the U.S. government has been remiss to undertake.
3. There were strong motivational factors related to civic responsibility to share collected data with the community (i.e., an opportunity to inform), along with personal satisfaction in having the opportunity to give back to the community.
4. Mechanisms for communication of special interest findings emerge primarily through conferences, and not through mass media to inform the public.
5. As for sharing data sets, there was strong consensus that existing mechanisms are either totally absent or inadequate.
6. The overwhelming consensus was if additional incremental funding becomes available, it would enhance data gathering, and this consensus was interpreted to mean funding is currently inadequate for all to reach their data-gathering goals.
7. Funding is particularly insufficient for organizations in underdeveloped countries.

8. There were two distinct themes that correlated well with other responses: (a) the desire to belong to a greater community with common interests, and (b) recognition of their time–cost efforts with at least a nominal reimbursement of personal expenditures.
9. Very few respondents identified an absence of unmet needs, indicating widespread frustration.
10. There were differences between the unmet needs of developed versus underdeveloped countries; the latter’s unmet needs concerned not only deficiency in resources, but also the inability to communicate issues with both the professional community and the public.
11. Half of respondents identified a desire to obtain recognition for their CS work, whereas others acknowledged their motivation was to satisfy their personal research interests.
12. There was a strong recognition that receiving a modest financial reimbursement for time–costs associated with data gathering and data sharing would be well-received.
13. Respondents had a strong and consistent desire for the acknowledged primary features of an open science-data sharing platform, in that it must be findability, accessibility, interoperability, reusability and open (FAIR/O) compliant (Chamanara et al., 2021)—indicating an essential requirement. A third of respondents expressed a desire for even the more esoteric data sets to be shared.

14. Overwhelming support was expressed for a commercial platform to be made available that facilitated the monetization of selective published data sets.
15. Financial motivation for sharing data was by no means the primary motivation (although further compensation would be well-received). Respondents possessed a high degree of civic-mindedness, desiring to give back to the community and inform the community of social injustice.
16. On the issue of solving the high barrier to establishing a viable critical mass of data, there were no suggestions; however, the majority of content providers were willing to take the time–cost risk of providing content in the hopes they would eventually be compensated for their efforts, thereby demonstrating a high degree of trust.
17. Respondents were very forthcoming in making suggestions to encourage their industry colleagues to provide data content to a platform. Common themes ranged from industry recognition to providing various data management tools, indicating areas that remain lacking for them.
18. Assistance on what data consumers desired of an open-science platform was requested, and respondents were asked to expand on the optimal nature of the core transaction for data consumers. The responses identified findability as the most important feature, followed by ability to rapidly review data sets online and subsequently download in a consumable format.
19. The reciprocal question was asked of what data content providers would most like to see in an open science-data platform, and there was broad consensus on expected features, such as tools to edit and visualize for data sets uploaded;

adherence to industry standards; support for team-based efforts; a cost-free facility; and strong support for other functions, such as ease of data set publication.

20. CS organizations are overwhelmingly supported financially by donations, not government grants.
21. Recognizing that respondents were from a mix of developed and underdeveloped countries, half of respondents had annual budgets of less than \$25,000, and 20% had budgets in the range of \$100,000–\$500,000.
22. Sixty percent of CS organizations from this survey had up to 25 staff, noting the industry is strongly characterized by volunteers. Twenty-six percent of CS organizations from this study had upward of 26 staff, and 13% had over 100 staff. Almost all CS organizations classified themselves as not-for-profit.

Thematic Analysis

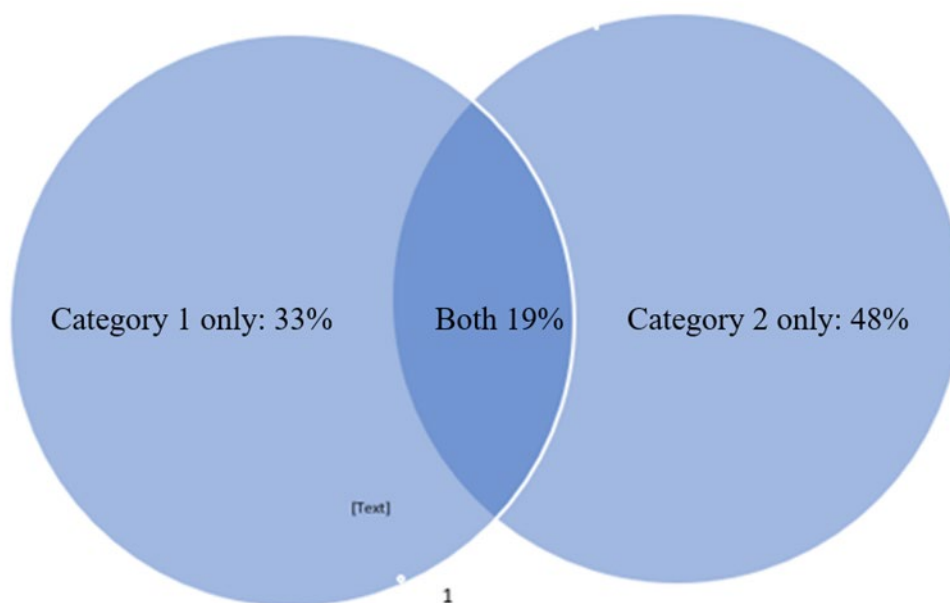
The thematic analysis was undertaken to identify, analyze and interpret patterns of meaning within the questionnaire responses contained in Appendix A, then they were categorized to identify category overlap to gain further insight and interpretation of unmet needs not satisfied by the existing business models, (refer Appendix B). Themes were not weighted for significance.

The study had two primary categories of research questions, namely a problem category of ‘barriers that impede science data from being shared within the global science community’, and the potential solution category of ‘the critical attributes of a market-driven platform technology business model whose mission is to facilitate global science-data sharing and satisfy unmet science community needs’. These two categories can be

considered the existing predominate strategy for science data sharing, and the second category a potential business model paradigm shift to sharing science data. The overlap between the Venn diagrams represents common requirements whichever business model is used, and contains such matters as adherence to uniform standards.

Figure 2

Venn diagram of Themes by Category



Summary

In this study, the CS sector was used as a proxy for the broader science community. Over 500 CEOs of CS organizations were contacted, and 15 responded to the 22 in-depth questions; the responses were presented in Appendix A in both percentage terms, and graphically. The questionnaire spanned the scope of the literature review. Thematic analysis identified that there is a significant range of unmet needs that are not or cannot be addressed by the prevailing global business models, but which have the potential to be addressed by a paradigm shift to a different business model that is

platform based where there are central management controls but decentralized ownership by data providers guided by market forces as to which data sets are in most demand, and the data sets owners remunerated commensurate with the value placed on their data sets based on global market demand.

CHAPTER 5: CONCLUSIONS AND DISCUSSION

In the literature review, the findings and publications of diverse authors over many years concerning evolving platform technology business models and the significant differences between those business models and traditional linear business models were investigated. Although platforms have resulted in a superior business model, the early adopters suffered numerous missteps in establishing themselves, e.g., failing to address the 'chicken or the egg' barrier to platform growth; however, the rewards for defining a successful platform business model and growth strategy have been immense—particularly with the emergence of global market dominators whose market position, once established, have been difficult to dislodge and which have released significant excess value. This study investigated whether those lessons could apply to the business of science-data sharing globally, a business some have valued at many USD billions per year in time–cost effort and which has been overlooked by the market for new strategies to efficiently share and reuse science data.

A questionnaire was issued to the chief executive officers (CEOs) of over 500 citizen science (CS) organizations around the world, where this market segment was selected as a proxy for the broader science community. The questionnaire results were concerned with the sharing of science data and were found to be entirely consistent with the literature review analysis, strongly demonstrating a range of unmet needs experienced by the CS community. CS organizations specifically are overwhelmingly supported by donations and direct government grants, and respondents regarded this support structure as a rather tenuous situation. The conclusion was that a well-designed business model for an open-science-data sharing platform could potentially satisfy many of these unmet

needs, provide market-driven incentives to encourage global science data sharing, and become commercially viable.

Discussion of Findings and Conclusions

The research questions had two broad topics of research:

1. barriers that impede science data from being shared with the global science community; and
2. critical attributes of a market-driven platform technology business model, whose mission is to facilitate global science data sharing and satisfy unmet science community needs.

Reflection on the Questionnaire Findings

The number of respondents to Appendix A Question 1 who advised there are political and jurisdictional barriers to sharing data with the global science community was surprising, at 67%, although entirely consistent with the literature research. Also surprising was that the survey participants felt there were only modest technical barriers to sharing data at 33%; however, there had been few requests to share. Of the respondents, 33% advised there were internal reasons for not wanting to share data, and comments reflected these reasons fell into two categories of (a) not trusting external parties to interpret the data correctly, and (b) their data archives were embarrassingly disorganized.

It was encouraging to see the strong response to Appendix A Question 2, at 67%, that identified the feeling of being team players with a common interest. This sentiment can translate to forming strong bonds laterally with other comparable groups around the

world. Responses to gathering data to meet academic requirements, at 33%, was also stronger than anticipated.

Consistent with the prior question, Appendix A Question 3 identified the sense of satisfaction in being able to give back to the local community and inform the community by sharing issues at 73% and 87%, respectively. This result augured well for a science-data sharing platform, and implied there is also a requirement to provide a narrative of projects to interpret findings. Receiving additional funds and project recognition also both rated moderately high, at 40% as motivational factors.

Participation in industry common interest conferences to promote CS projects and coordinate with other comparable groups was a key issue of Appendix A Question 4, at 73%, as a mechanism for communicating with the broader community, and indicated a platform that could assist in this communication would enhance this communication need.

A striking response to Appendix A Question 5, at 80%, was the need to be able to facilitate the sharing of data sets; respondents affirmed there are no formal mechanisms to be able to facilitate data sharing, a finding quite different to the earlier question on communicating project issues. This finding correlated with findings from the literature review—that there was a strong (and global) unmet need to readily share data (Belous, 2018; Bonney et al., 2016; Reichman et al., 2011; Theobald et al., 2015; Wallis et al., 2013). This finding was further reinforced by 53% of respondents claiming there was no straightforward data distribution channel.

A unanimous response from all participants was the notion that they could do more with additional funding. This Appendix A Question 6 finding was also consistent

with the literature in that project funding has historically been scarce, and there is a heavy reliance on volunteerism (Bonney et al., 2016; Theobald et al., 2015). A mechanism to provide additional incremental funds, independent of legacy project funding, would clearly be well received and a strong motivator to further facilitate global data sharing.

Appendix A Question 7 identified there were some underlying issues of mistrust toward government in underdeveloped countries, and this mistrust was compounded by traditional low prioritization toward funding environmental projects in those countries. These findings indicated if funding was made available outside of the traditional in-country sources, and the data generated could be privately stored and secure under the control of the data gathering groups, there would be strong support for a service such as the one presented, as it would empower these independent entities into a virtuous circle.

Appendix A Question 8 sought verification from prior questions; as anticipated, respondents reinforced their earlier responses. In that question, the negative responses were important; only 13% advised they were not aware of any unmet needs. Clearly, the CS sector's needs are not adequately met, and there is an opportunity to remedy this limitation and address these needs.

Appendix A Question 9 supported findings of the literature review in that CS projects in underdeveloped countries remain considerably underfunded; further, the public as well as professional groups may not even be aware of what research projects are being conducted. This finding indicated both projects and the availability of data sets need recognition.

The response to Appendix A Question 10 concerning motivation by citizen scientists to obtain recognition for their work was equally split; however, there was a

clear message that CS participants had a strong sense of giving back to the community, which could be assisted by the availability of accessible channels for this work.

Even though there was an altruistic component of undertaking CS data collection efforts, 73% responded to Appendix A Question 11 that a modest financial recognition of their efforts would be well received in offsetting their time—costs, with two thirds of respondents identifying such recognition would be a welcome motivator for a variety of reasons.

Respondents to Appendix A Question 12 overwhelmingly and consistently confirmed the critical attributes of a platform business model whose mission is to facilitate science-data sharing and satisfy unmet science community needs, and even modestly supported publication of more esoteric data sets.

Again, there was overwhelming confirmation to Appendix A Question 13 that participants would be willing to share their data sets as long as this task was not onerous, and respondents would be receptive to receiving modest revenue on that basis.

Respondents demonstrated an overwhelming and clear sense of altruism in sharing their data. Appendix A Question 14 identified that being able to receive a remuneration was the lowest motivator, at 40%, with the sense of satisfaction of contributing to the community being the highest, at 87%. This latter motivator was followed closely by providing a teaching opportunity and gaining recognition for their contributions to science and the community.

There were no strong viewpoints nor suggestions from Appendix A Question 15 as to how to overcome the ‘chicken or the egg’ conundrum to grow a platform for sharing data. Fewer than half of respondents (47%) would even consider being paid to upload

their data, and a majority (54%) were content to undertake that time–cost task given the expectation they would be fairly remunerated for their efforts in the future.

Appendix A Question 16 requested suggestions to provide incentives to encourage data content contributions. Participants were gracious in being most forthcoming with innovative ideas. This willingness further reinforced the demand for a data-sharing platform to be made available to the global CS community.

Appendix A Question 17 requested feedback on what data consumers would desire in a science-data sharing platform. Responses were all highly consistent with the literature review, and also verified the assessment of the core transaction (87%) in being able to find published and peer-reviewed data sets, being able to review the data online graphically (80%), and subsequently being able to download it easily (74%) in a consumable format.

Appendix A Question 18 sought to confirm the needs of data contributors, again strongly confirming the literature review; the most important feature (80%) included the ability to create, edit, and visualize data sets uploaded, followed closely by adherence to standards (73%). Other significant needs of data contributors were identified as the ability to securely store data sets in the cloud at no cost (67%), which presumably reflected the lack of resources available to CS organizations to fund their own data storage facilities. Respondents nearly equally answered at 60% regarding the ability to easily upload or download data sets and the ability to selectively publish data sets.

When seeking to determine the source of funding for data collection, Appendix A Question 19 surprisingly found the primary source of funding came from donations (73%) rather than government grants (7%). This finding places CS organizations in a

very perilous and volatile funding state. This survey was restricted to CS organizations, whereas the broader science community has a strong element of direct government grant funding as identified by the literature review. What this survey demonstrated was, again, additional, incremental revenue would be well received and would assist ongoing data gathering expansion.

An interesting survey response to Appendix A Question 20 was the profile of typical annual budgets, with half of all respondents having a budget of less than \$25,000. However, 20% had annual budgets in the \$100,000–\$500,000 range, which demonstrates strong community financial support even if not quite sufficient for CS organization needs.

The staffing resource profile from Appendix A Question 21 was also interesting, demonstrating 60% of CS organizations have 1–25 staff members, and 13% have in excess of 100 staff. As anticipated, 93% of CS entities identified as not-for-profit. More interesting was the private entity conducting CS activities.

Barriers That Impede Science Data from Being Shared with the Global Science Community

CS organizations surveyed broadly agreed with the literature review consensus of the critical technical attributes required of an open access technology platform. These attributes are commonly termed findability, accessibility, interoperability, reusability and open (FAIR/O). The last attribute is optionally open, because many platforms have a subscription model to enable access to their data. The existence of technical barriers was not identified as a major reason for the absence of sharing science data; many study participants indicated they would be happy to share their data if asked and subsequently

acknowledged; however, a remarkably high proportion of CS organizations advised that there was an absence of effective formal data sharing mechanisms available within the industry, both locally and globally.

The conundrum remains there is a lack of an efficient opportunity and incentive to promote the availability of data sets that owners are willing to share (i.e., findability), combined with lack of perceived benefits to the data set owner in expending this time–cost to upload their data sets to the commons (i.e., accessibility). This problem should not be surprising given project funding is not from an open and efficient market that provides market demand signals for data sharing, but is instead a limited requirement from predominately western government grantors for science projects which mandate data set publication for larger grantees only. The result is market inefficiency in the matching process between data owners and data consumers. This barrier to data sharing is due to an absence of an efficient market demand mechanism to induce participation in data sharing (i.e., primarily a marketing issue, not a technical issue that needs to be overcome).

For a platform business to succeed, the literature review identified that it is essential for that platform to identify and streamline its core transaction. The survey responses were highly consistent with the literature review in identifying the value of the core transaction for science data consumers, which is ease of findability of data (i.e., a communications issue). This consistent response acknowledged a central finding of an unmet need, which is the lack of a facility that enables ease of search and retrieval for data sets.

Critical Attributes of a Market-Driven Technology Platform Business Model

There is an exceptionally strong sense of altruism and fellowship in gathering science data, as evidenced by extremely high reliance on volunteer labor by CS organizations. There is also an exceptionally strong desire within CS organizations to inform the community of societal issues, filling an information gap left unsatisfied and underreported by government or mainstream media. Respondents expressed frustration that there are inadequate channels available for communicating issues. Those in the CS industry have a strong preference for networking with their peers and relying heavily on conferences to network rather than remote networking (e.g., a listserv), although the latter remains an important channel of communication. CS organizations are only moderately motivated by recognition, but for those who require recognition, they decried the paucity of channels of communication to enable them to achieve this requirement, particularly for international communication. The conclusion was a communications facility that could provide a forum for common interests internationally would be well received, satisfy the desire for a sense of global community, and would facilitate the sharing of project findings along with sharing and reusing science data.

A unanimous finding from respondents was that CS projects are underfunded. This underfunding stands in contrast to the USD billions per year spent in volunteer time–cost, and this situation is quite remarkable. The absence of adequate funding is even more acute for CS organizations located in developing countries. This inequity can be attributed to cultural issues, as well as less discretionary spending within the local economies. There was a wide disparity in revenue levels between survey respondents, and this disparity was expected, given that a number were from underdeveloped countries. This finding also identified that for CS organizations with lower budget levels,

even a modest increase in revenue can potentially make a significant difference in funding their activity levels, particularly given the leverage provided by large numbers of volunteers.

Overwhelmingly, CS organizations felt they were not satisfying unmet needs of their data collection participants. These unmet needs equally ranged from sense of community and reimbursement of direct costs, to recognition of volunteer work effort. Unmet needs in underdeveloped countries were particularly pronounced when it came to the ability to provide community awareness of issues or even identifying which scientific projects were being undertaken; respondents also noted the shortage of resources to undertake that data gathering. There was overwhelming consensus that sharing science data with the global community is a constructive task to creating a public good, with the proviso that the task does not have a significant time–cost barrier. Respondents welcomed the opportunity to monetize their data set assets, even if this effort resulted in marginal income. The sense of satisfaction in being able to contribute to their community was the highest motivator for CS participants, followed closely by providing a teaching opportunity. Financial remuneration was their lowest motivator; however, in the context of being able to facilitate expanding projects, a modest financial return generated through data sharing was viewed as a well-regarded mechanism to support these activities.

Funding of CS projects of those surveyed is overwhelmingly gleaned by donations. The literature review identified that this business model provides a very perilous circumstance for organizational continuity, and is the probable reason for the enthusiastic acceptance by survey respondents of the concept of additional incremental revenue from nondonor sources by monetizing their data assets. This shift in funding

would diversify their revenue base and potentially provide some level of funding continuity. Over half of respondents noted their willingness to undertake the time–cost risk of uploading data sets in anticipation—but not a guarantee—that their efforts would be financially rewarded. This willingness to take risk demonstrated cautious optimism. Consequently, a strategy of modest initial data contribution from CS organizations with rapid positive acceptance by the research community would lead to the CS organization generating a modest remuneration from a share of third-party funds. This remuneration, combined with recognition, would be expected to provide a virtuous and positive feedback mechanism. This mechanism would encourage further data contribution and incrementally grow the data provider side of the platform.

CS organizations are clearly not strongly motivated by remuneration for their extensive efforts in collecting data; however, they do feel strongly that even a modest financial return would be welcome as an acknowledgment of their efforts. Similarly, survey respondents were also highly consistent, both among themselves and with the literature review, in identifying the significant unmet needs of data contributors. They seek sophisticated online tools for data editing and visualization, adherence to industry data standards, data security, and facilitation of team-based projects, all at no cost to themselves. The conclusion was any global facility must be frictionless from a price perspective, providing basic technical tools while also providing the opportunity for a modest revenue stream that at least has the potential to cover the time-costs of their data ingest. Given the very high proportion of volunteerism, the actual cost of data ingest would be expected to remain low. Additionally, any incremental revenue generated would be regarded as highly beneficial—particularly if it also provided additional

intangible benefits, such as acknowledgement of volunteer contribution to society, identification of social injustice issues, and enhancing a feeling of community with global peers.

Survey respondents were quite entrepreneurial in offering additional practical mechanisms for motivating CS organizations to contribute data, and these concepts deserve further study. Recommendations spanned sophisticated online analytical tools, incentive payments via crypto funds, formal and informal acknowledgement of contributions to society, and open communication facilitation with peers.

Application of Findings and Conclusions to the Problem Statement

The findings of this study aligned with the literature review in that there was a significant and global unmet need to facilitate science-data sharing. Additionally, CS organizations were used as a proxy for the broader science community. Various researchers have tried to quantify the extent of the problem of the lack of science-data sharing and have arrived at nominal values of USD billions per year and increasing. This study confirmed that barriers to science-data sharing are not due to lack of motivation; on the contrary, data owners are very willing to share their data for reuse. Further, findings suggested a common consensus that a technology platform with parameters that are FAIR/O compliant should be required to facilitate science-data sharing. There are significant costs to establish a standardized global facility and data take-up process to reach the critical mass of data sets and provide ongoing support to make such a platform of utility. This high-cost threshold and risk has traditionally only been acceptable to large government-funded projects, with very specific goals that have led to siloed data sets that only meet the needs of the government program and not necessarily the broader research

community. Consequently, there is a commercial opportunity for an organization willing to undertake the risk of systems development and substantial data ingest costs.

CS organizations have been identified as having numerous unmet needs, and as a proxy for the broader science community, it was concluded this situation will continue indefinitely for the science community until one or more market-based solutions can be established. There is clearly a market failure and, consequently, an opportunity to satisfy this global market that continues to expend time—costs of billions per year in resources in data gathering. Moreover, there are significant unmet needs ranging from lack of data sharing and reuse to an absence of recognition and lack of even token remuneration for the efforts of a vast volunteer resource. Despite these unmet needs, there is very strong comradery and civic mindfulness that goes unrecognized. Intangible losses from this market breakdown include the difficulty in communicating project findings to impacted communities, along with lost benefits from data reuse and the lost opportunity from releasing value between pools of isolated data sets that modern artificial intelligence (AI) techniques could potentially correlate and leverage.

Any science-data sharing platform that is not directly impacted by market forces—which provide a feedback mechanism to both data providers and data consumers—will continue to yield inefficient results in both sharing, and, by extension, satisfying the primary unmet needs of these market participants. It was also concluded the lack of science-data sharing is not due to a technical barrier, as these requirements for data sharing and data reuse are well defined and agreed upon; rather, it is a marketing and business model barrier that needs to be overcome.

Application to Business

Through this study, based on successful outcomes for establishing other technology platform business models, it was identified that the definition of a platform business distinctly differs from the traditional definitions of a linear business. Consequently, different growth strategies and organizational emphasis must be applied. As evidenced through the literature review, platform businesses have the potential to be significant market disrupters. The analysis helped formulate a theory for the establishment of a market-driven, open-access science data sharing business model which, if implemented, would present a paradigm shift for the science community—particularly for those communities that remain disproportionately marginalized and underfunded. Such a platform would provide market-demand signals that would reward and compensate data providers' time-costs while responding more effectively to the data needs of researchers. This platform would be a market-disrupting event impacting the science-data-gathering industry and its multibillion-dollar operations per year, and would instead create a commercial market for science-data sharing where one does not currently exist.

Recommendations for Action

Once a robust technology platform that is FAIR/O compliant has been established, there are essential strategies of the business model that need further compliance to ensure broad acceptance by the global science community and enable the platform to flourish. These strategies differ significantly from those typically experienced by linear businesses.

Establish Frictionless Market Participation

The platform needs to be open to both data consumers and data contributors (i.e., zero direct cost of market participation). Platform monetization should not incur friction to market participants. This action will facilitate maximum rate of growth of participation.

Focus on the Core Transaction

During the platform launch phase, systems development resources should prioritize the discrete core transaction of value to both parties to the two-sided platform of data contributors and data consumers. Sophisticated functions for both sides can be added subsequently. For a science-data sharing platform, the core transactions that provide primary value creation between data content owners and data consumers include the ability to efficiently publish and search respectively.

Acknowledge Emerging Web 3.0 Concepts

Acknowledgment of emerging Web 3.0 concepts of both revenue sharing and enhanced social value are the two most important concepts for a science-data sharing platform to succeed, but they are conspicuously absent in government-funded and controlled data-exchange projects. This acknowledgement involves establishing an equitable revenue-sharing model with data contributors based on the relative popularity of their data sets, as well as establishing effective communication channels to enable the platform community to efficiently interact. Examples include bringing awareness of social justice issues to the directly impacted communities, and transparency on transaction volume trends as a metric of market acceptance.

Focus on Unmet Needs

For broad acceptance by the global science community, there needs to be a focus on unmet needs of market participants of a nonmonetary nature. This focus may take the form of providing recognition to individuals, projects, or organizations that heavily rely on volunteers to collect data in the field, or who take considerable personal risk in collating hazardous community health data, such as Ebola virus outbreak data.

Solve the “Chicken or the Egg” Conundrum

The chicken or the egg conundrum is the most significant and overlooked barrier to overcome to achieve market-building success. Initially, focusing on a segment of the market where a critical mass can be cost-effectively attained that rapidly satisfies both sides of the platform, then incrementally seek out other distinct segments of the market to satisfy, broadening the market offering. There are three distinct elements of this strategy: user sequencing, monetary subsidy, and product feature subsidy. Each platform business implements these strategies differently.

User Sequencing

Where the platform strategist identifies there is minimal market resistance to establishing a new business, they should also initially seek out areas of the market where there are both considerable unsatisfied needs and where the platform is not perceived as a threat by legacy market incumbent participants. For a global science-data sharing platform, an example of a priority region with which to commence may include the continent of Africa, where there has been a historic paucity of science-data sharing for various reasons.

Monetary Subsidy

To establish a critical mass of data content, data providers may have to be paid to collect and ingest data sets to the platform until a critical mass of data are achieved that it then becomes valuable to the data consumers. To take the science-data sharing platform example further, qualified university students and graduates can be paid by the hour to search for peer-reviewed data sets concerning community health, ingest that data into the platform, and publish that data under their own accounts.

Product Feature Subsidy

For a platform that wishes to provide an equitable share of revenue to data providers based on popularity of searches—and prior to the platform reaching critical mass that would attract sponsorship revenue—a strategy could be to provide an internally funded sponsorship stream that could emulate future sponsorship levels. Continuing with the example, the paid-data ingest teams could receive further incremental revenue from their published data sets based on an algorithm that detects the popularity of the data sets they have published, as measured by unique online views, downloads, and other metrics. This product feature of revenue sharing for data contributors would initially be subsidized internally until such time as the volume of platform participants reaches a critical mass where it becomes attractive for independent sponsorship.

Platform Organizational Management

Platform businesses have different organizational emphasis to that of traditional linear businesses. For a science-data sharing platform, this different emphasis would manifest itself in establishing heightened governance functions within the management

organization, particularly in data curation, adherence to data standards, and community communication support.

Viral Marketing and Platform Valuation

The value of a platform business is directly proportional to the volume of platform users. All platforms need to be cognizant of Metcalfe's power law to grow their user base—namely, that the valuation of an entity is proportional to the square of the number of users. Viral marketing techniques have been identified in the literature as being the most cost-effective strategy for promoting platform technology-based businesses (Chakravorti, 2004; Fouad, 2017; Garrigos-Simon, 2012; Kennedy, 2020; Moazed & Johnson, 2016; Parker et al., 2016; Reillier & Reillier, 2017).

Facilitate Market Entrepreneurship

This study identified technology platforms have the propensity to become market disrupters that are significantly more efficient and cost effective than linear businesses and government-fiat-based platforms. To closely emulate independent market forces, the science-data sharing platform should encourage participation by science entrepreneurs who can potentially create incremental value. An example would be by contributing to the long-tail effect of having esoteric data sets that create marginal initial value, but aggregate recurring value over time and fill market niches that otherwise would be overlooked by mainstream science and satisfy the data needs of niche consumers. This market emulation would be particularly valuable to those who can provide data from underdeveloped countries, where project funding is weak or nonexistent and where a market-based science-data sharing business platform would provide monetary incentives

and feedback signals for market participation appropriate to the global market value of their published contribution.

Recommendations for Further Research

Issues that require further research identified in this dissertation are explained as follows: survey participant ideas, data set copyright, decentralized finance and global payments.

Further Expansion of Platform Concepts to Facilitate Data Sharing

Increasingly, it became clear that the principles discussed in this dissertation can apply to many evidence-based science-data projects (e.g., the National Institute of Health's Africa Open Science Data Platform Project; Glass & Tromberg, 2019).

Survey Participant Ideas

Survey respondents were quite entrepreneurial in offering additional practical mechanisms for motivating CS organizations to contribute data, and these concepts deserve further study. Recommendations ranged from having sophisticated online analytical tools, incentive payments via crypto funds, formal and informal acknowledgement of contributions to society, and open communication facilitation with peers.

Data Set Copyright

Proof of copyright of data may become an increasingly contentious issue once the owner becomes aware that someone else who has published their data onto the science-data sharing platform has generated revenue from their work. This intellectual property

issue may require arbitration by platform management. Platform proof of original ownership of data sets may ultimately be resolved through the use of non-fungible tokens (NFT).

Decentralized Finance and Global Payments

For a global platform, the cost of labor in widely varying economic regimes becomes significant. Similar to how manufacturing has been outsourced from developed countries to underdeveloped countries with low-cost labor, it is anticipated that a global science-data sharing platform will see a similar contribution by educated people from underdeveloped countries, where costs of living and subsequent labor costs are relatively low. The value of the published data sets will be uniform globally, but traditional banking transaction costs vary in different economic regimes. An ability to disburse funds globally outside the traditional banking infrastructure with very low or zero transaction fees using decentralized finance of crypto currency payments would be highly advantageous, particularly for those from underdeveloped countries.

Concluding Statement

Using platform technology to address the need to efficiently and globally share science data for reuse is surely one of the last significant market opportunities ripe for breakthrough disruption. Science-data collection has time-costs of many billions of USD per year and growing, the market is highly fragmented, and there are diverse unmet needs. AI techniques have made breakthroughs in many industries, but their prerequisite is a standardized pool of big data from which to analyze and identify data correlations and anomalies to make their projections. This application of AI cannot readily be

undertaken on data sets that are siloed. Consequently, the potential for this market is significant; yet, why has it not been attempted?

The answer is a number of organizations have commenced to use cloud computing techniques to enable the sharing of science data, but their strategies to establish a workable business model have been mediocre and the costs are prohibitive for many nongovernment entities. The U.S. government has tried to encourage these efforts; for example, the U.S. National Institutes of Health (NIH) launched an initiative in early 2019 to acquire a cloud-based solution to store and share health data from all countries in Africa, and has a working budget of approximately \$62 million over the initial 6 years. This example gives some idea of the costs of establishing a government facility for just one continent. It also gives an insight into the lack of entrepreneurial market initiative that big government possesses, as it has completely failed to use free-market business constructs, as referenced in this dissertation.

Consequently, there exists a market opportunity to use platform technology to satisfy global unmet needs for sharing and reuse of science data, thereby creating an industry for market-driven commercial exchange of science data where one does not currently exist, independent of government fiat. Whether any new market entrants make the same strategic business model mistakes to those of early platform adopters remains to be seen; however, the potential business opportunity for a successful technology platform for open-science sharing is significant, and may be worth the commercial risk.

REFERENCES

- Abdel-Fattah, M. A. (2015). Grounded theory and action research as pillars for interpretive information systems research: A comparative study. *Egyptian Informatics Journal*, 16(3), 309–327. <https://doi.org/10.1016/j.eij.2015.07.002>
- Akerlof, G. A. (1970). The market for “lemons”: Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3), 488–500. <https://doi.org/10.2307/1879431>
- Baines, A., Ittefaq, M., & Abwao, M. (2021). Scamdemic, #Plandemic, or #Scaredemic: What Parler social media platform tells us about COVID-19 vaccine. *Vaccines*, 9(5), Article 421. <https://doi.org/10.3390/vaccines9050421>
- Belous, Y. (2014). Why do we share information online? A study of motivations. *International Journal of Arts & Sciences*, 7(4), 383–400. <https://www.proquest.com/docview/1644513235>
- Bonney, R., Phillips, T. B., Ballard, H. L., & Enck, J. W. (2016). Can citizen science enhance public understanding of science? *Public Understanding of Science*, 25(1), 2–16. <https://doi.org/10.1177/0963662515607406>
- Chakravorti, B. (2004). The new rules for bringing innovations to market. *Harvard Business Review*, 82(3), 58–126. <https://hbr.org/2004/03/the-new-rules-for-bringing-innovations-to-market>
- Chamanara, J., Gaikwad, J., Gerlach, R., Algergawy, A., Ostrowski, A., & König-Ries, B. (2021). BEXIS2: A FAIR-aligned data management system for biodiversity, ecology and environmental data. *Biodiversity Data Journal*, 9, e72901-e72901. <https://doi.org/10.3897/BDJ.9.e72901>

- Chan, N. K. (2019). The rating game: The discipline of Uber's user-generated ratings. *Surveillance & Society*, 17(1/2), 183–190.
<https://doi.org/10.24908/ss.v17i1/2.12911>
- Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). SAGE Publications.
- Christensen, C. M. (2016). *The innovator's dilemma: When new technologies cause great firms to fail* (3rd ed.). Harvard Business Review Press.
- Cici, C., & D'Isanto, D. (2017). Integrating sustainability into core business. *Symphonya*, (1), 50–65. <https://doi.org/10.4468/2017.1.05cici.disanto>
- Clemens, J., & Veuger, S. (2020). Implications of the COVID-19 pandemic for state government tax revenues. *National Tax Journal*, 73(3), 619–644.
<https://doi.org/10.17310/ntj.2020.3.01>
- Cooperrider, D. L., & McQuaid, M. (2012). The positive arc of systemic strengths: How appreciative inquiry and sustainable designing can bring out the best in human systems. *Journal of Corporate Citizenship*, 2012(46), 71–102.
<https://doi.org/10.9774/gleaf.4700.2012.su.00006>
- Costa, D. G. R. (2016). From pipelines to networks: The intellectual disruptions of online platforms. *Economic Affairs*, 36(3), 379–389. <https://doi.org/10.1111/ecaf.12207>
- Cowell, J. M., Lee, K., Malcolm-Smith, S., Selcuk, B., Zhou, X., & Decety, J. (2017). The development of generosity and moral cognition across five cultures. *Developmental Science*, 20(4), 1–12. <https://doi.org/10.1111/desc.12403>
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed method approaches* (3rd ed.). SAGE Publications.

- De Filippo, D., Lascurain, M. L., Pandiella-Dominique, A., & Sanz-Casado, E. (2020). Scientometric analysis of research in energy efficiency and citizen science through projects and publications. *Sustainability*, *12*(12), Article 5175. <https://doi.org/10.3390/su12125175>
- de Soto, H. (2000). *The mystery of capital: why capitalism triumphs in the west and fails everywhere else*. Perseus Books Group.
- DiNucci, D. (1999). Fragmented future. *Print*, *53*(4), 32–34. http://darcy.d.com/fragmented_future.pdf
- Driver, M., & Porter, M. E. (2012). An interview with Michael Porter: Social entrepreneurship and the transformation of capitalism. *Academy of Management Learning & Education*, *11*(3), 421–431. <https://doi.org/10.5465/amle.2011.0002a>
- Drucker, P. F. (1955). *The practice of management* (Rev. ed.). Routledge.
- Echezona, R. I., & Ugwuanyi, C. F. (2010). African university libraries and internet connectivity: Challenges and the way forward. *Library Philosophy and Practice*, 1–13. <https://digitalcommons.unl.edu/libphilprac/421/>
- Financial Times. (2020, July). *TikTok/Hong Kong: Worlds apart*. <https://www.ft.com/content/bdf76bd9-897b-47df-87b1-14c7ae662331>
- Fouad, N. (2017). *Viral marketing effect on digital knowledge acquisition: WhatsApp as a model*. SAGE Publications.
- Fung, I. C. H. (2008). Open access for the non-English-speaking world: Overcoming the language barrier. *Emerging Themes in Epidemiology*, *5*(1), 1–6. <https://doi.org/10.1186/1742-7622-5-1>

- Garrigos-Simon, F., Rafael, L. A., & Ribera, T. B. (2012). Social networks and web 3.0: Their impact on the management and marketing of organizations. *Management Decision*, 50(10), 1880–1890. <https://doi.org/10.1108/00251741211279657>
- Gerhardt, P. (2007). *Diversity at work: What every leader should know to be more effective*. Xlibris.
- Gilder, G. (2018). *Life after Google: The fall of big data and the rise of the blockchain economy*. Salem Media Group.
- Glass, R., & Tromberg, B. (2019). *Harnessing data science for health discovery and innovation in Africa*. National Institute for Health, Common Fund. <https://commonfund.nih.gov/AfricaData>
- Gold, A. E. (2019). Community consequences of AirBnB. *Washington Law Review*, 94(4), 1577–1637. <https://digitalcommons.law.uw.edu/wlr/vol94/iss4/2/>
- Hart, H. (2018). *Doing a literature review* (2nd ed.). SAGE Publications.
- Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a theory of ecosystems. *Strategic Management Journal*, 39(8), 2255–2276. <https://doi.org/10.1002/smj.2904>
- Jahnke, A. (2015, July 5). Who picks up the tab for science? *The Brink*. <http://www.bu.edu/articles/2015/funding-for-scientific-research>
- Jetter, M., Satzger, G., & Neus, A. (2009). Technological innovation and its impact on business model, organization and corporate culture – IBM’s transformation into a globally integrated, service-oriented enterprise. *Business & Information Systems Engineering*, 1(1), 37–45. <https://doi.org/10.1007/s12599-008-0002-7>

- Kennedy, M. (2020). If the rise of the TikTok dance and e-girl aesthetic has taught us anything, it's that teenage girls rule the internet right now: TikTok celebrity, girls and the coronavirus crisis. *European Journal of Cultural Studies*, 23(6), 1069–1076. <https://doi.org/10.1177/1367549420945341>
- Kirkpatrick, D. (2010). *The Facebook effect*. Simon and Schuster.
- Kobori, H., Dickinson, J. L., Washitani, I., Sakurai, R., Amano, T., Komatsu, N., Kitamura, W., Takagawa, S., Koyama, K., Ogawara, T., & Miller-Rushing, A. J. (2016). Citizen science: A new approach to advance ecology, education, and conservation. *Ecological Research*, 31(1), 1–19. <https://doi.org/10.1007/s11284-015-1314-y>
- Lee, D. J., & Stvilia, B. (2017). Practices of research data curation in institutional repositories: A qualitative view from repository staff. *PloS One*, 12(3), 3–44. <https://doi.org/10.1371/journal.pone.0173987>
- Malgonde, O., Zhang, H., Padmanabhan, B., & Limayem, M. (2020). Taming the complexity in search matching: Two-sided recommender systems on digital platforms. *MIS Quarterly*, 44(1), 48–84. <https://doi.org/10.25300/MISQ/2020/14424>
- Marshall, C., & Rossman, G. B. (2016). *Designing qualitative research* (6th ed.). SAGE Publications.
- Metcalfe, R., & Boggs, D. (1976). Ethernet: Distributed packet switching for local computer networks. *Communications of the ACM*, 19(7), 395–404. <https://doi.org/10.1145/360248.360253>

- Mickey, R. (2010). The impact of a seller's eBay reputation on price. *The American Economist*, 55(2), 162–169. <https://doi.org/10.1177/056943451005500217>
- Minot, J. R., Arnold, M. V., Alshaabi, T., Danforth, C. M., & Dodds, P. S. (2021). Ratioing the president: An exploration of public engagement with Obama and Trump on twitter. *PloS One*, 16(4), Article e0248880. <https://doi.org/10.1371/journal.pone.0248880>
- Mishra, M., Yan, P., & Schroeder, R. (2022). TikTok politics: Tit for tat on the India–China cyberspace frontier. *International Journal of Communication*, 16, 814–339. <https://ijoc.org/index.php/ijoc/article/view/17771>
- Moazed, A., & Johnson, N. (2016). *Modern monopolies: What it takes to dominate the 21st-century economy* (1st ed.). St. Martin's Press.
- Morello, L., & Guglielmi, G. (2018). US science agencies reap budget windfall. *Nature*, 555(7698), 572. <https://doi.org/10.1038/d41586-018-03700-9>
- Navia, P., Perelló, L., & Masek, V. (2022). Demand without supply? Mass partisanship, ideological attachments, and the puzzle of Guatemala's electoral market failure. *International Area Studies Review*, 25(2), 99–120. <https://doi.org/10.1177/22338659211072939>
- Nasdaq. (n.d.). *Stocks*. Retrieved December 15, 2021, from <https://www.nasdaq.com/market-activity/stocks/>
- Nash, J. F. (1950). The bargaining problem. *Econometrica*, 18(2), 155–162. <https://doi.org/10.2307/1907266>
- Nesi, J., Choukas-Bradley, S., & Prinstein, M. J. (2018). Transformation of adolescent peer relations in the social media context: Part 1-A theoretical framework and

- application to dyadic peer relationships. *Clinical Child and Family Psychology Review*, 21(3), 267–294. <https://doi.org/10.1007/s10567-018-0261-x>
- Núñez, M. A. & Pauchard, A. (2011). Biological invasions in developing and developed countries: Does one model fit all? *Biological Invasions*, 12(4), 707–714. <https://doi.org/10.1007/s10530-009-9517-1>
- Page, W. H., & Childers, S. J. (2012). Antitrust, innovation, and product design in platform markets: “Microsoft” and “intel.” *Antitrust Law Journal*, 78(2), 363–395. <https://doi.org/10.2139/ssrn.1914737>
- Parker, G. G., Van Alstyne, M. W., and Choudary, S. P. (2016). *Platform revolution: How networked markets are transforming the economy and how to make them work for you* (1st ed.). W.W. Norton & Company.
- Perren, R., & Kozinets, R. V. (2018). Lateral exchange markets: How social platforms operate in a networked economy. *Journal of Marketing*, 82(1), 20–36. <https://doi.org/10.1509/jm.14.0250>
- Phillips, W., Lee, H., Ghobadian, A., O’Regan, N., & James, P. (2015). Social innovation and social entrepreneurship: A systematic review. *Group & Organization Management*, 40(3), 428–461. <https://doi.org/10.1177/1059601114560063>
- Plantin, J.-C., Carl, L., & Edwards, P. N. (2018). Re-integrating scholarly infrastructure: The ambiguous role of data sharing platforms. *Big Data & Society*, 5(1), 1–14. <https://doi.org/10.1177/2053951718756683>
- Porter, M. E. (1985). *Competitive advantage: Creating and sustaining superior performance*. The Free Press.

- Pounds, A., & Bostock, J. (2019). Suitable economic models for open educational resources initiative in aquaculture higher education. *Aquaculture International*, 27(5), 1553–1563. <https://doi.org/10.1007/s10499-019-00406-1>
- Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and opportunities of open data in ecology. *Science*, 331(6018), 703–705. <https://doi.org/10.1126/science.1197962>
- Reillier, B., & Reillier, L. C. (2017). *Platform strategy: How to unlock the power of communities and networks to grow your business*. Routledge.
- Rochet, J., & Tirole, J. (2014). Platform competition in two-sided markets. *Competition Policy International*, 10(2), 79–96. <https://doi.org/10.1162/154247603322493212>
- Rudman, R., & Bruwer, R. (2016). Defining web 3.0: Opportunities and challenges. *Electronic Library*, 34(1), 132–154. <https://doi.org/10.1108/EL-08-2014-0140>
- Rufai, S. R., & Bunce, C. (2020). World leaders' usage of twitter in response to the COVID-19 pandemic: A content analysis. *Journal of Public Health*, 42(3), 510–516. <https://doi.org/10.1093/pubmed/fdaa049>
- Simon, M. K., & Goes, J. (2018). *Dissertation and scholarly research: Recipes for success*. Dissertation Success. <https://doi.org/10.13140/RG.2.1.5089.0960>
- Spulber, D. F. (2019). The economics of markets and platforms. *Journal of Economics & Management Strategy*, 28(1), 159–172. <https://doi.org/10.2139/ssrn.3227135>
- Suarez, F. F., & Kirtley, J. (2012). Dethroning an established platform. *MIT Sloan Management Review*, 53(4), 35–41. <https://sloanreview.mit.edu/article/dethroning-an-established-platform/>

- Sun, M., & Zhu, F. (2013). Ad revenue and content commercialization: Evidence from blogs. *Management Science*, *59*(10), 2314–2331.
<https://doi.org/10.1287/mnsc.1120.1704>
- Theobald, E. J., Ettinger, A. K., Burgess, H. K., DeBey, L. B., Schmidt, N. R., Froehlich, H. E., Wagner, C., HilleRisLambers, J., Tewksbury, J., Harsch, M. A., & Parrish, J. K. (2015). Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biological Conservation*, *181*, 236–244. <https://doi.org/10.1016/j.biocon.2014.10.021>
- Thiel, P., & Masters, B. (2014). *Zero to one: Notes on startups, or how to build the future* (1st ed.). Random House US.
- Tirole, J. (2015). Market failures and public policy. *The American Economic Review*, *105*(6), 1665–1682. <https://doi.org/10.1257/aer.15000024>
- Van Alstyne, M. W., Parker, G. G., & Choudary, S. P. (2016). Pipelines, platforms, and the new rules of strategy. *Harvard Business Review*, *94*(4), Article 54.
<https://hbr.org/2016/04/pipelines-platforms-and-the-new-rules-of-strategy>
- Vijay, D., & Gekker, A. (2021). Playing politics: How Sabarimala played out on TikTok. *The American Behavioral Scientist*, *65*(5), 712–734.
<https://doi.org/10.1177/0002764221989769>
- Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS One*, *8*(7), Article e67332. <https://doi.org/10.1371/journal.pone.0067332>

Wan, X., Cenamor, J., Parker, G., & Van Alstyne, M. (2017). Unraveling platform strategies: A review from an organizational ambidexterity perspective.

Sustainability, 9(5), Article 734. <https://doi.org/10.3390/su9050734>

Wee, V. (2017). Youth audiences and the media in the digital era: The intensification of multimedia engagement and interaction. *Cinema Journal*, 57(1), 133–139.

<https://doi.org/10.1353/cj.2017.0064>

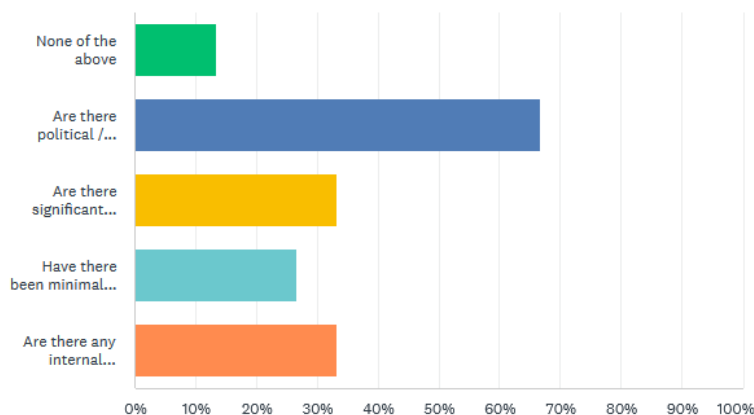
APPENDIX A

Survey Questions

The following questions were asked of CEO of Citizen Science (CS) organizations drawn from both the developed and underdeveloped countries.

Question 1. In your experience, what are the barriers that impede science data from being shared with the global science community? Check all that apply:

Answered: 15 Skipped: 0



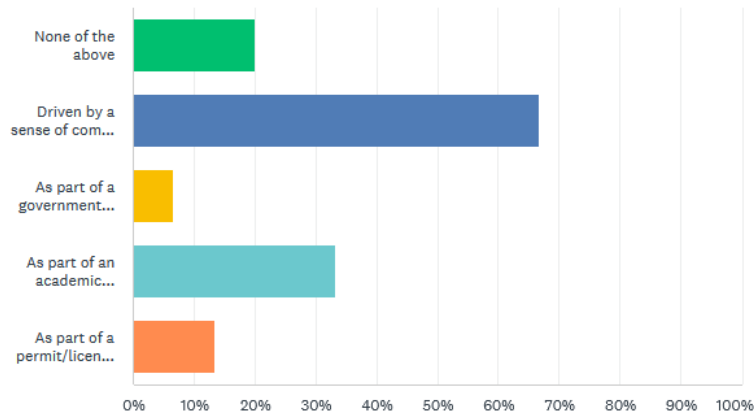
ANSWER CHOICES	RESPONSES
None of the above	13.33% 2
Are there political / administrative boundaries of the jurisdiction that is funding the project that creates any barriers to sharing the data outside of those boundaries.	66.67% 10
Are there significant technical barriers to effectively sharing data sets with external entities?	33.33% 5
Have there been minimal external requests for data sharing?	26.67% 4
Are there any internal reasons for not wanting to share your data sets, such as confidentiality.	33.33% 5
Total Respondents: 15	

Survey respondent comments:

- We have encountered data hoarding from government officials that fall into two categories: a. They don't trust the conclusions you will draw from any data released. (and don't want you second guessing their assessments using their own data). b. Their database archives are in such a mess, they are embarrassed to release any of it, for fear their mismanagement will be exposed.
- Information is gathered by outside entities but not shared locally.

Question 2. What initially motivated you to gather data? Check all that apply:

Answered: 15 Skipped: 0



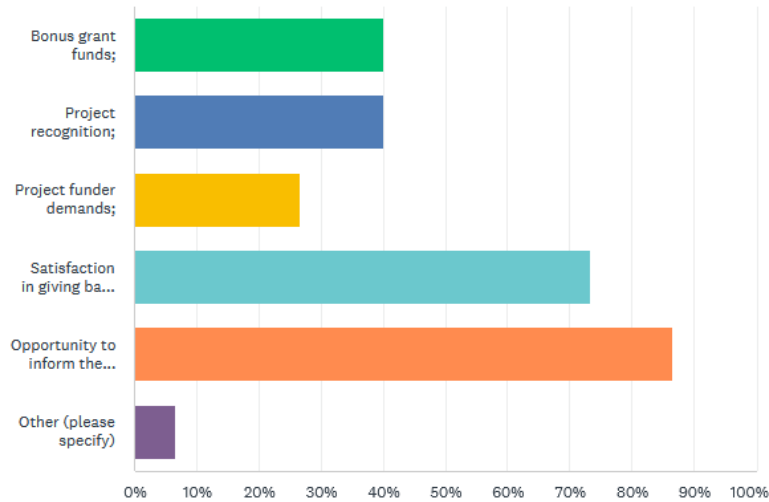
ANSWER CHOICES	RESPONSES
None of the above	20.00% 3
Driven by a sense of common interest by your data gathering community of volunteers.	66.67% 10
As part of a government funded project.	6.67% 1
As part of an academic requirement.	33.33% 5
As part of a permit/license requirement.	13.33% 2
Total Respondents: 15	

Survey respondent comments:

- Our State environmental program is underfunded and lacks motivation to study our natural resources. We started collecting data to fill in the gaps.
- To collect data that is not otherwise being collected by government entities; To engage members of the community in monitoring work to develop a stronger sense of environmental awareness

Question 3. What other incentives exist to share the data gathered outside of the immediate community. Check all that apply:

Answered: 15 Skipped: 0



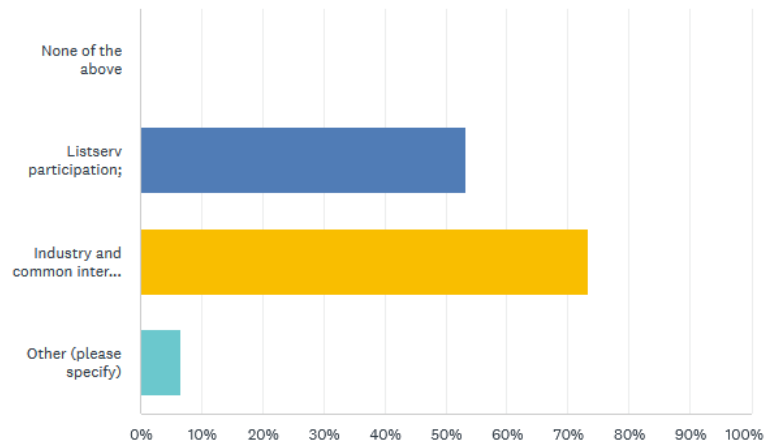
ANSWER CHOICES	RESPONSES	
▼ Bonus grant funds;	40.00%	6
▼ Project recognition;	40.00%	6
▼ Project funder demands;	26.67%	4
▼ Satisfaction in giving back to the community;	73.33%	11
▼ Opportunity to inform the community and share issues;	86.67%	13
▼ Other (please specify) Responses	6.67%	1
Total Respondents: 15		

Survey respondent comments:

- Crowdsourcing data analysis to come up with more detailed and interesting findings.

Question 4. Identify what mechanisms exist for coordination of environmental data gathering projects with the broader community. Check all that apply:

Answered: 15 Skipped: 0



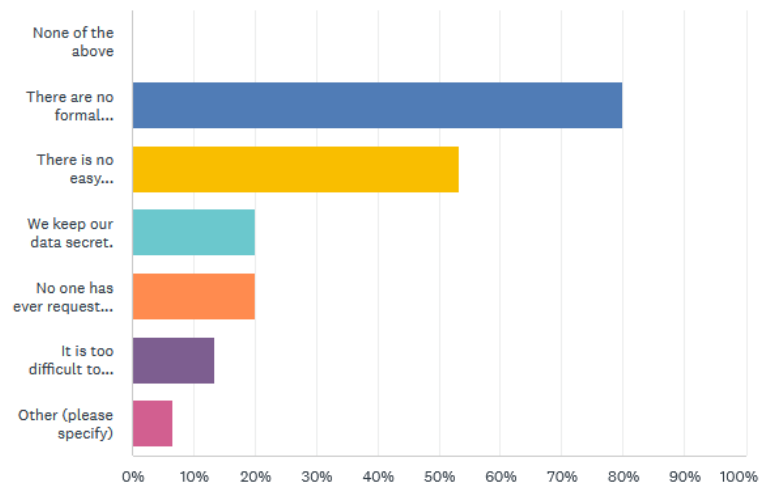
ANSWER CHOICES	RESPONSES
None of the above	0.00% 0
Listserv participation;	53.33% 8
Industry and common interest conferences;	73.33% 11
Other (please specify)	Responses 6.67% 1
Total Respondents: 15	

Survey respondent comments:

- To provide critical information necessary for environmental protection.

Question 5. Identify what mechanisms, if any, there are for the public or professional community to gain access to project data sets; and identify whether these mechanisms tend to be effective and efficient. Check all that apply:

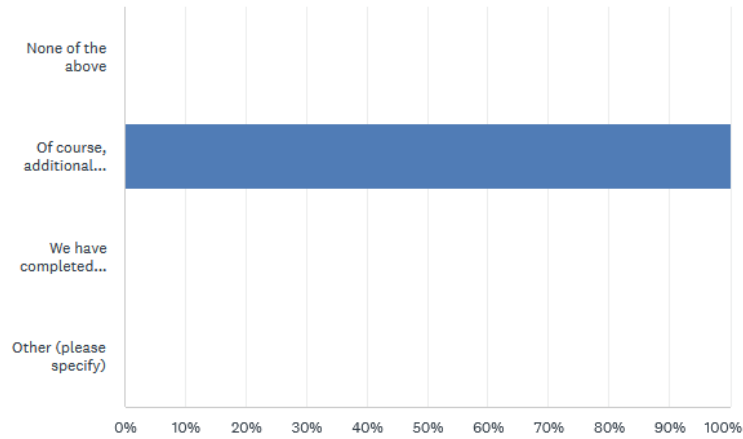
Answered: 15 Skipped: 0



ANSWER CHOICES	RESPONSES	
None of the above	0.00%	0
There are no formal mechanisms to share data sets;	80.00%	12
There is no easy distribution channel/mechanism;	53.33%	8
We keep our data secret.	20.00%	3
No one has ever requested access to our data sets.	20.00%	3
It is too difficult to satisfy requests for access for a variety of reasons.	13.33%	2
Other (please specify)	Responses 6.67%	1
Total Respondents: 15		

Question 6. Identify whether additional funding would induce projects to continue and expand their data collection efforts further. Check the box that applies:

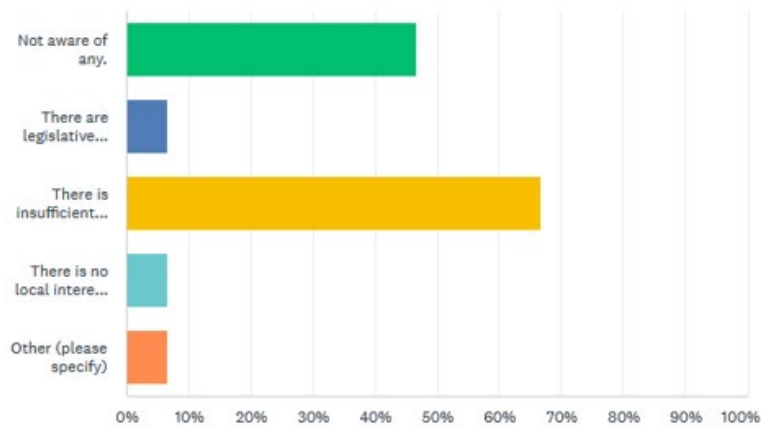
Answered: 15 Skipped: 0



ANSWER CHOICES	RESPONSES
None of the above	0.00% 0
Of course, additional funds would enable us to expand the data collection project.	100.00% 15
We have completed gathering all potential data and additional funding is not relevant.	0.00% 0
Other (please specify)	Responses 0.00% 0
TOTAL	15

Question 7. Identify if you are aware of any unique barriers to citizen science participation by developing countries. Check all that apply:

Answered: 15 Skipped: 0



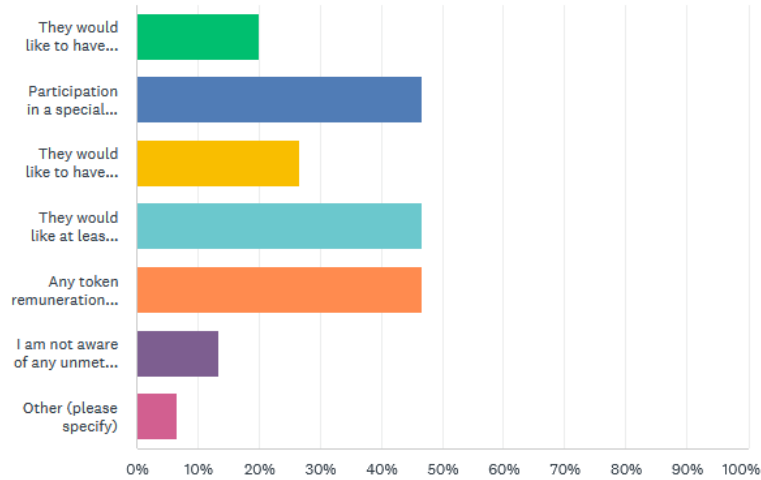
ANSWER CHOICES	RESPONSES
▼ Not aware of any.	46.67% 7
▼ There are legislative barriers that prohibit sharing of data with the general public.	6.67% 1
▼ There is insufficient funding available to facilitate data gathering in developing countries.	66.67% 10
▼ There is no local interest in volunteering to gather data in developing countries.	6.67% 1
▼ Other (please specify)	Responses 6.67% 1
Total Respondents: 15	

Survey respondent comments:

- Citizens of totalitarian regimes may be wary of collecting data and submitting it to a central authority. Reminds them of being spied on by the government.

Question 8. What are the “unmet needs” of your data collecting group? Check all that apply.

Answered: 15 Skipped: 0



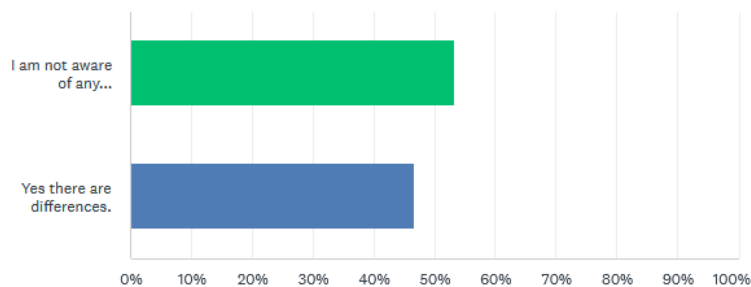
ANSWER CHOICES	RESPONSES
▼ They would like to have greater professional recognition by external peers.	20.00% 3
▼ Participation in a special interest group, i.e. sense of community.	46.67% 7
▼ They would like to have at least some additional remuneration above and beyond reimbursement of their direct costs.	26.67% 4
▼ They would like at least their direct costs to be reimbursed in full.	46.67% 7
▼ Any token remuneration would be gratefully received and would contribute to recognizing their work effort.	46.67% 7
▼ I am not aware of any unmet needs within our group.	13.33% 2
▼ Other (please specify)	Responses 6.67% 1
Total Respondents: 15	

Survey respondent comments:

- Clearer data entry mechanisms, such as smart phone apps that work, and immediate feedback on results submitted.

Question 9. Are you aware of any differences between the unmet needs of underdeveloped countries compared to the developed countries?

Answered: 15 Skipped: 0



ANSWER CHOICES	RESPONSES	
▼ I am not aware of any differences.	53.33%	8
▼ Yes there are differences.	46.67%	7
TOTAL		15

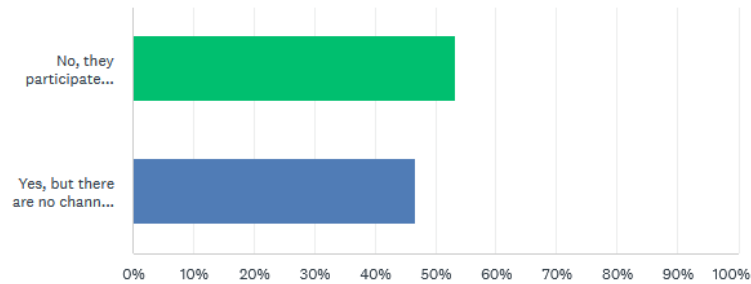
[Comments \(5\)](#)

Survey respondent comments:

- More people are engaged in data gathering, because allocated resources are considerable, also a wider network of interest.
- Community awareness level is lower and lack of access to information in underdeveloped countries.
- There is additional funding and access to monitoring resources in developed countries.
- Funding, expertise and equipment availability.
- Generally undeveloped countries are usually not even aware of research projects and do not benefit from them.

Question 10. Is there a motivation by citizen scientists to obtain recognition for their work?

Answered: 15 Skipped: 0



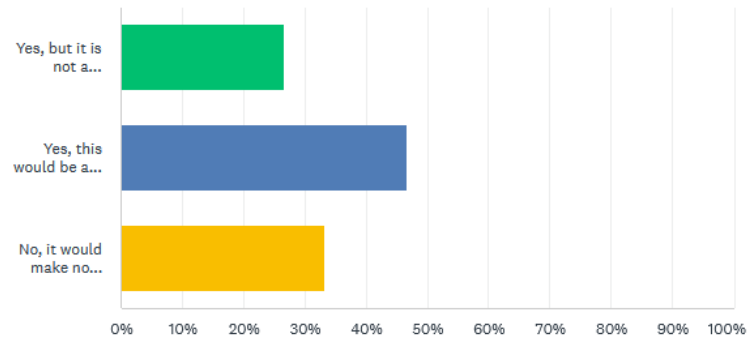
ANSWER CHOICES	RESPONSES	
▼ No, they participate only to satisfy their personal scientific interests.	53.33%	8
▼ Yes, but there are no channels available to readily do that.	46.67%	7
TOTAL		15

Survey respondent comments:

- Though there are some volunteers who participate for community service or who may be interested in sharing their data more widely in their own community.

Question 11. Is the ability of citizen scientists to generate a modest revenue stream to offset their data collection time–cost a motivation for sharing their data content?

Answered: 15 Skipped: 0

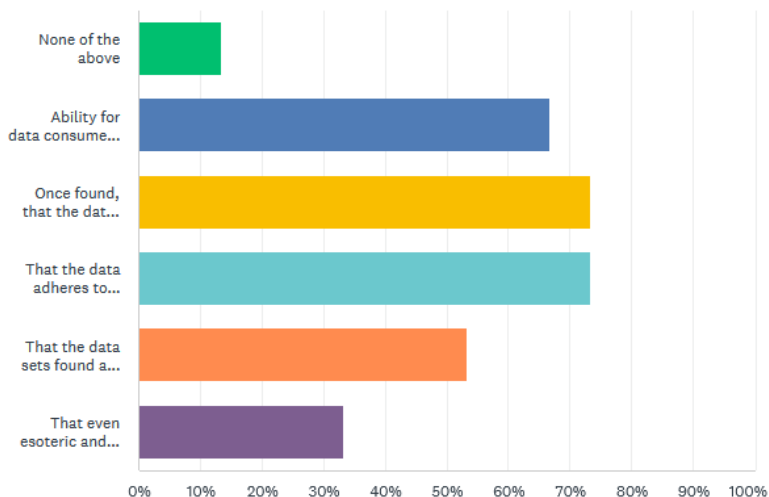


ANSWER CHOICES	RESPONSES	
▼ Yes, but it is not a significant motivator.	26.67%	4
▼ Yes, this would be a welcome motivator for a variety of reasons.	46.67%	7
▼ No, it would make no significant difference.	33.33%	5
Total Respondents: 15		

Comments (0)

Question 12. What are the critical attributes of a platform business model whose mission is to facilitate science data-sharing and satisfy unmet science community needs? Check all that apply:

Answered: 15 Skipped: 0



ANSWER CHOICES	RESPONSES
None of the above	13.33% 2
Ability for data consumers to easily search for and identify the availability of data sets of interest (Find-ability).	66.67% 10
Once found, that the data sets are readily accessible and easy to download in acceptable formats (Accessible);	73.33% 11
That the data adheres to recognized scientific standards (Interoperable);	73.33% 11
That the data sets found are 'open source' and that anyone can access and have access to the same data sets (Reusable);	53.33% 8
That even esoteric and difficult to find data sets are readily available (long tail effect).	33.33% 5
Total Respondents: 15	

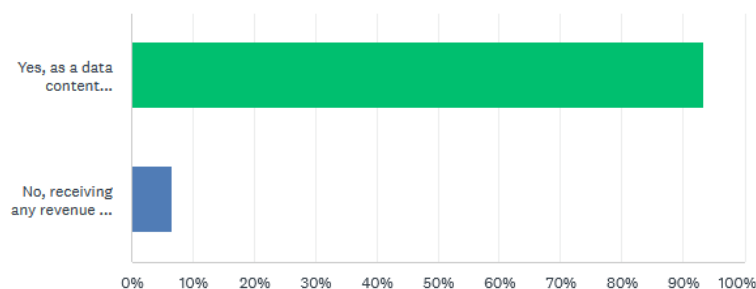
[Comments \(1\)](#)

Survey respondent comments:

- Value added analysis and comparisons between datasets, for example, how Orbitz compares data from American Airlines and Delta on flight schedules and costs so the data consumer can make better choices and get a better idea of market conditions/fluctuations.

Question 13. Would the citizen science community in particular support the concept of a market-driven platform for sharing science data? Definition: Market-driven is an approach to business that prioritizes identifying the needs and desires of consumers and creating products and services that satisfy them (with or without a cost to the consumer).

Answered: 15 Skipped: 0



ANSWER CHOICES	RESPONSES
▼ Yes, as a data content provider I would welcome the opportunity to obtain a series of modest payments to my organization for providing access to data sets that I decide to publish.	93.33% 14
▼ No, receiving any revenue at all from sharing our data sets is irrelevant to me and my organization.	6.67% 1
TOTAL	15

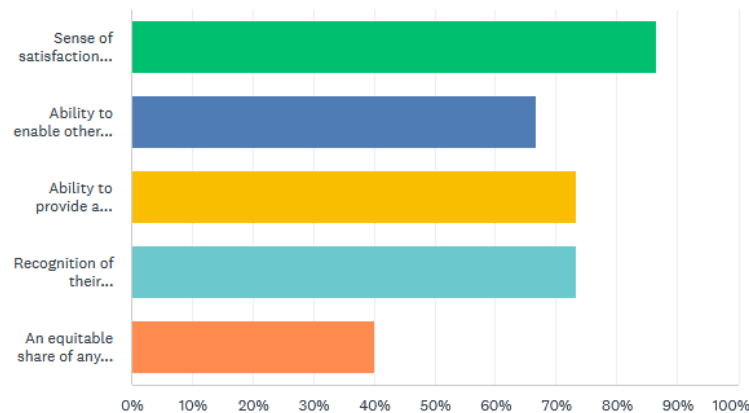
[Comments \(1\)](#)

Survey respondent comments:

- Well, of course we will sell you our data for the right price as long as we don't have to jump through a bunch of hoops and massive re-formatting to submit it. Your scooper should sort, categorize and organize based on your data model, not make us do that for you.

Question 14. What motivational factors would encourage the citizen science community volunteers to participate in a market-driven platform technology. Check all that apply:

Answered: 15 Skipped: 0



ANSWER CHOICES	RESPONSES
▼ Sense of satisfaction in contributing to science and the global community;	86.67% 13
▼ Ability to enable others to satisfy a social injustice and/or potentially avert a community health issue.	66.67% 10
▼ Ability to provide a teaching opportunity to others.	73.33% 11
▼ Recognition of their contribution, even when this is modest.	73.33% 11
▼ An equitable share of any remuneration resulting from the sharing of the data collected.	40.00% 6
Total Respondents: 15	

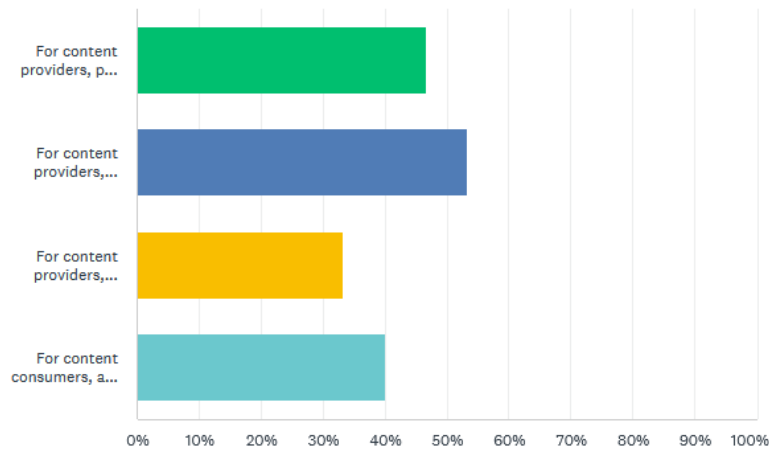
[Comments \(1\)](#)

Survey respondent comments:

- Recruitment of new data collection volunteers if properly promoted through your service.

Question 15. The ‘chicken or the egg’ problem refers to the barrier experienced by growing platforms of surpassing the critical threshold of content that encourages others to participate in a two-sided market of content providers and content consumers. What mechanisms can you suggest to overcome that barrier? Check all that apply:

Answered: 15 Skipped: 0



ANSWER CHOICES	RESPONSES
▼ For content providers, pay them their agreed time-costs to upload their selected data sets to the platform;	46.67% 7
▼ For content providers, expect them to upload their selected data sets at their own cost in the expectation by them that there will be adequate future platform revenue to cover their time-cost investment;	53.33% 8
▼ For content providers, there is no need to expect remuneration as long as they perceive value in being able to store their data sets privately and securely on the platform without obligation to share any of their data sets;	33.33% 5
▼ For content consumers, as long as the platform remains open source that is FAIR compliant and has a reasonable ability to satisfy their data requirements then there is no need for further incentives to have them utilize the platform;	40.00% 6
Total Respondents: 15	

Comments (0)

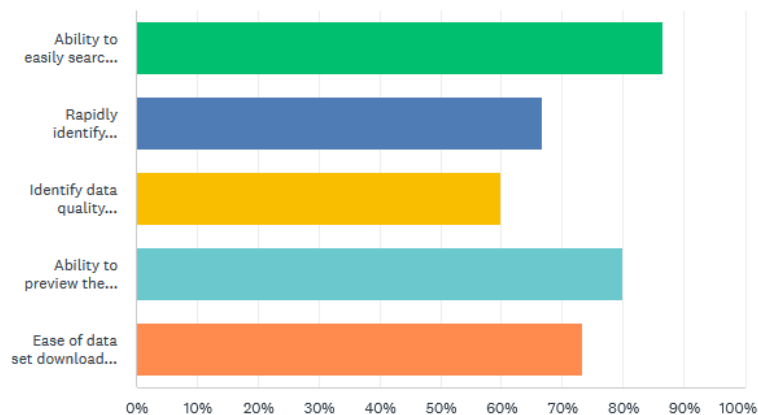
Question 16. Describe other mechanisms to provide incentives that will encourage ongoing platform data set contributions: Comment box.

Survey respondent comments:

- Availability to have analyzing tools, and dashboard functionalities.
- Incentives payment via digital money account based on total data sets uploaded.
- a. Motivation b. Authentic discussion c. Time with cost payment distribution.
- Larger scale projects and data collection requesting ability
- Communication meeting, certificate of honor
- Sharing of data between groups. One group provides their data to another group whose data they need (a type of reciprocity).
- Make people feel important and needed
- Environmental, health and economic interest for their own interest, it generates motivation.
- Open records requests
- Having open dialogue with government agencies and a recognition by them for such work

Question 17. For data consumers, what are the most important features of an open science data sharing platform? Please check all that apply:

Answered: 15 Skipped: 0

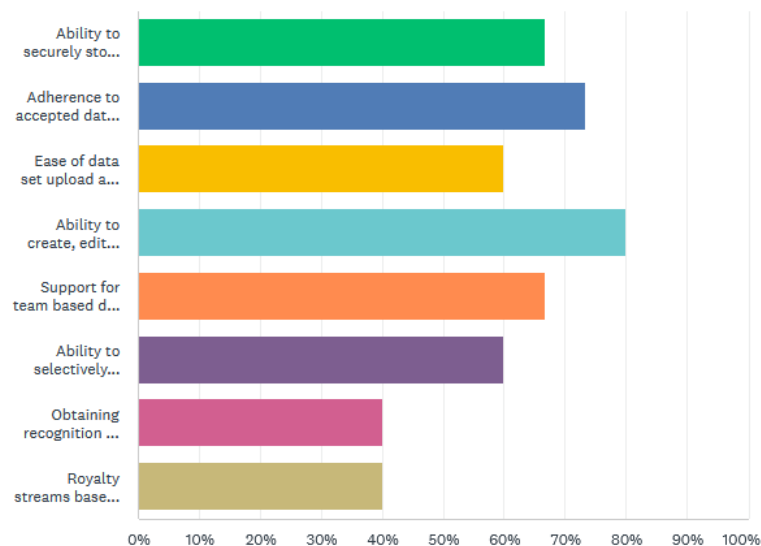


ANSWER CHOICES	RESPONSES
Ability to easily search and find published and peer-reviewed data sets (find-ability);	86.67% 13
Rapidly identify project background information;	66.67% 10
Identify data quality certification;	60.00% 9
Ability to preview the selected data sets online in graphical form;	80.00% 12
Ease of data set download into a consumable format;	73.33% 11
Total Respondents: 15	

Comments (0)

Question 18. For data contributors, what are the most important features of an open science data sharing platform? Check all that apply:

Answered: 15 Skipped: 0

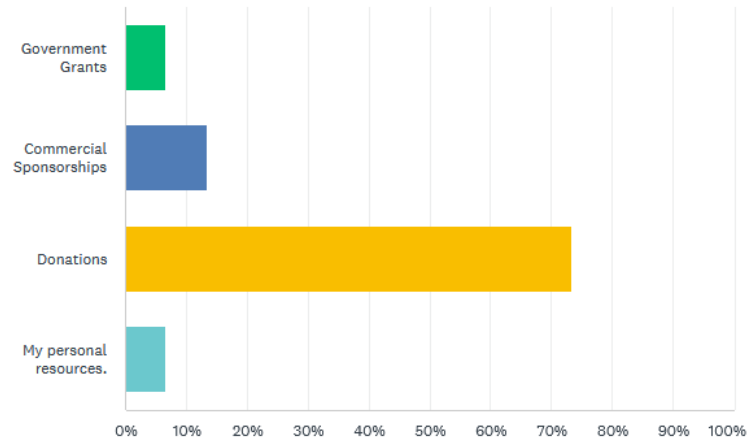


ANSWER CHOICES	RESPONSES
Ability to securely store data sets in the cloud at no cost;	66.67% 10
Adherence to accepted data standards;	73.33% 11
Ease of data set upload and download of selected data sets;	60.00% 9
Ability to create, edit and visualize data sets online;	80.00% 12
Support for team based data set contributions;	66.67% 10
Ability to selectively publish data sets to the public domain;	60.00% 9
Obtaining recognition for data set publication;	40.00% 6
Royalty streams based on data set views and downloads by the public;	40.00% 6
Total Respondents: 15	

Comments (0)

Question 19. What is the source of your data collection funding?

Answered: 15 Skipped: 0

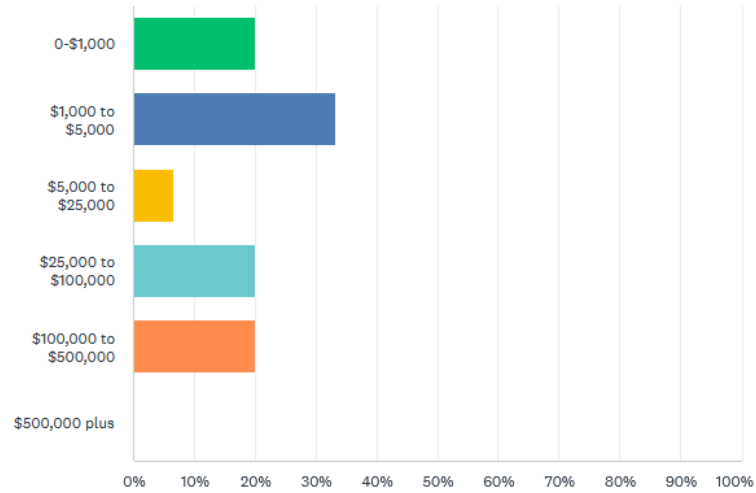


ANSWER CHOICES	RESPONSES
▼ Government Grants	6.67% 1
▼ Commercial Sponsorships	13.33% 2
▼ Donations	73.33% 11
▼ My personal resources.	6.67% 1
TOTAL	15

[Comments \(1\)](#)

Question 20. What is your typical annual funding for citizen science projects?

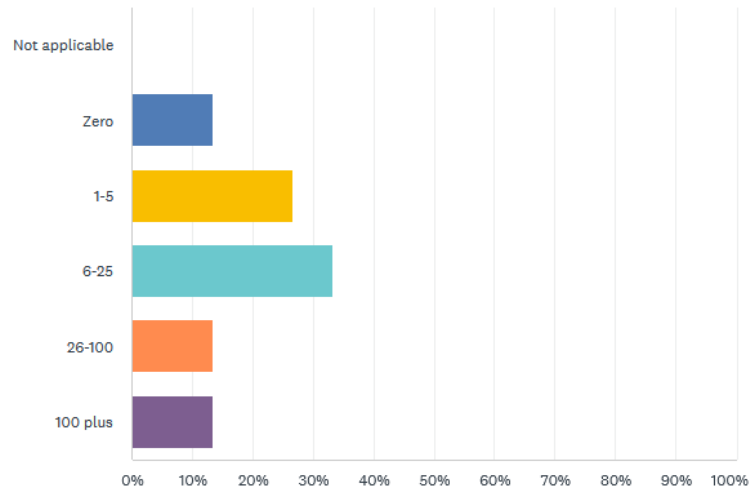
Answered: 15 Skipped: 0



ANSWER CHOICES	RESPONSES
▼ 0-\$1,000	20.00% 3
▼ \$1,000 to \$5,000	33.33% 5
▼ \$5,000 to \$25,000	6.67% 1
▼ \$25,000 to \$100,000	20.00% 3
▼ \$100,000 to \$500,000	20.00% 3
▼ \$500,000 plus	0.00% 0
TOTAL	15

Question 21. For data collectors, how many volunteers would you use in a typical year of data collection?

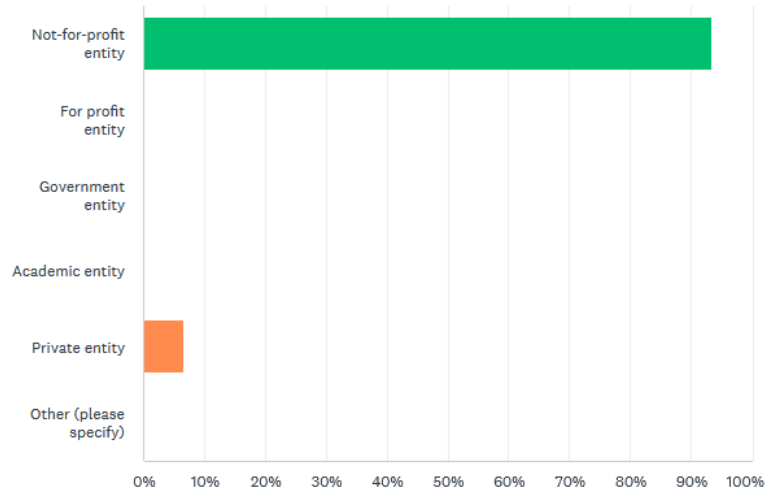
Answered: 15 Skipped: 0



ANSWER CHOICES	RESPONSES	
▼ Not applicable	0.00%	0
▼ Zero	13.33%	2
▼ 1-5	26.67%	4
▼ 6-25	33.33%	5
▼ 26-100	13.33%	2
▼ 100 plus	13.33%	2
TOTAL		15

Question 22. Please identify your organization type.

Answered: 15 Skipped: 0



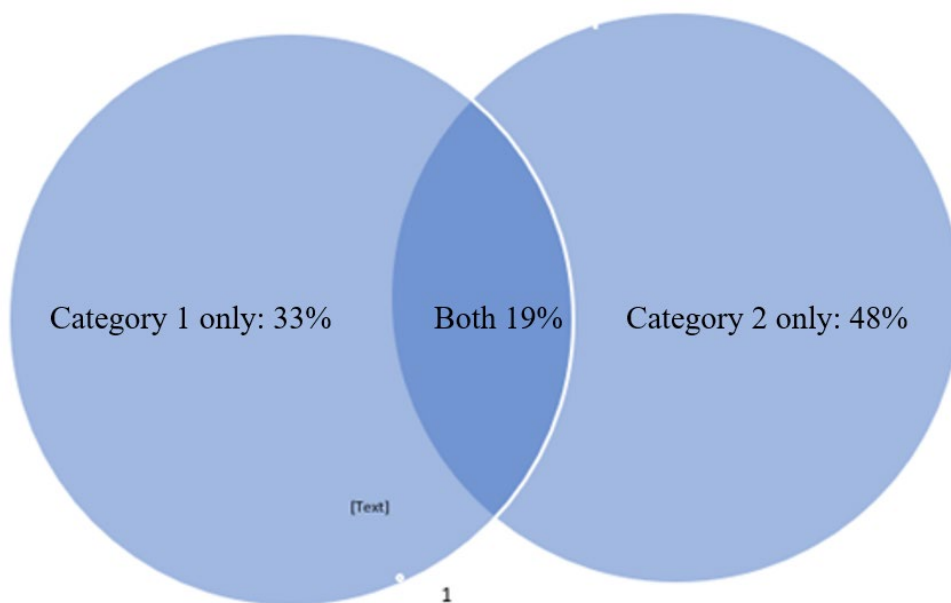
ANSWER CHOICES	RESPONSES
▼ Not-for-profit entity	93.33% 14
▼ For profit entity	0.00% 0
▼ Government entity	0.00% 0
▼ Academic entity	0.00% 0
▼ Private entity	6.67% 1
▼ Other (please specify)	Responses 0.00% 0
TOTAL	15

APPENDIX B

Venn Diagram of Themes by Category

Category 1. Barriers that impede science data from being shared with the global science community.

Category 2. Critical attributes of a market-driven platform technology business model whose mission is to facilitate global science-data sharing and satisfy unmet science community needs.



Themes by Category:

- Political and administrative barriers to data sharing (1)
- Camaraderie and altruism with the global CS community (2)
- Civic responsibility to share collected data with the global community (2)
- Mechanisms for communication for informing the public (2)

- Existing data sharing mechanisms absent and/or inadequate (1)
- CS funding is currently totally inadequate (1)
- Funding in underdeveloped countries is acutely inadequate (1)
- Desire to belong to a greater (global) community (2)
- Nominal reimbursement of volunteer personal expenditures (1/2)
- Widespread frustration at not satisfying unmet needs (1)
- Deficiency of resources particularly acute in underdeveloped countries (1)
- Inability to be able to communicate with the global professional community (1)
- Desire to be recognized for CS contribution (2)
- Motivation for CS to satisfy personal research needs (2)
- Recognition motivator if receiving reimbursement for data gathering cost (2)
- Consensus on the FAIR/O requirements of a data sharing platform (2)
- Overwhelming support for monetizing selected published data sets (2)
- Civic-mindedness and desire to give back to the community (2)
- Informing the public of social injustice (2)
- High degree of trust in future compensation from platform participation (2)
- Availability of free data management tools (1/2)
- Current absence of industry recognition and motivation factors (1)

- Core transaction for data consumers: findability, online graphical review, ease of download (1/2)
- Core transaction for data providers: editing tools, data visualization, zero-cost hosting, ease of publication (1/2)
- Need for diversity of funding sources (1/2)
- Annual budget constraints (1)
- Volunteer data collector recognition (2)