Research Policy 51 (2022) 104398

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

A Framework and Databases for Measuring Entrepreneurial Ecosystems

Evan Johnson [1,*], Iman Hemmatian [2], Lauren Lanahan [3], Amol M. Joshi [4]

[1] University of North Carolina at Chapel Hill, Department of Public Policy, Abernethy Hall CB# 3435, Chapel Hill, NC 27599
[2] Cal Poly Pomona, College of Business Administration, 3801 West Temple Ave., Pomona, CA 91768
[3] University of Oregon, Lundquist College of Business, 1208 University St. Eugene, OR 97403
[4] Wake Forest University, School of Business, 1834 Wake Forest Rd. Farrell Hall, Building 60, Winston-Salem, NC 27109

ARTICLE INFO

JEL:
M13
O38
H57
H77

Keywords:
Government R&D
innovation systems
entrepreneurial ecosystems
entrepreneurship
firm-government interactions
federal programs

ABSTRACT

Scholarly literature on the concept of entrepreneurial ecosystems has increased sharply over the past five years. The surge in interest has also heightened the demand for robust empirical measures that capture the complexity of dynamic relationships among ecosystem constituents. We offer a framework for measurement that places collaborative relationships among entrepreneurs, firms, government agencies, and research institutions at the center of the ecosystem concept. We further emphasize the four roles of the federal government as a *catalyst, coordinator, certifier*, and *customer* in shaping these relationships. Despite the central importance of these firm-government interactions, there is surprisingly little research on suitable methodologies and appropriate data for systematically and reliably incorporating them into measures of ecosystem health. Our study aims to address this gap in the literature by first developing a conceptual framework for measuring entrepreneurial ecosystems and then describing an array of accompanying databases that provide rich and detailed information on firms and their relationships with government organizations, accelerators, and research institutions. A major advantage of our approach is that all the underlying databases are drawn from non-confidential, publicly available sources that are transparently disclosed and regularly updated. This greatly expands the potential community of scholars, managers, and policymakers that may independently use these databases to test theories, make decisions, and formulate policies related to innovation and entrepreneurship.

1. Introduction

This paper provides a measurement framework that highlights the role of the public sector in supporting entrepreneurial and innovative ecosystems. Questions of how to measure entrepreneurial ecosystems have only recently begun to gain attention as scholars and administrators aim to improve models of innovation and entrepreneurship policy (Stam, 2018). The ecosystem concept itself is still young, and efforts to build theory and robust analytical approaches leave much to be developed. Part of the challenge arises from the inherent complexity of entrepreneurial ecosystems and the need for both contextual sensitivity and attention to dynamic interactions among actors such as firms, governments, universities, and communities. We present an approach to ecosystem measurement that capitalizes on multiple rich sources of observational data. Critically, these data sources have key features that, we argue, make them particularly useful for ecosystem measurement: 1) they are updated frequently to reflect changes in economics and policy; 2) they contain links to one another that support the use of relational database tools; 3) they allow for detailed geographic analysis that captures local context; and 4) they afford measurement of nonlinear processes and outcomes among actors.

The idea that economic solutions to problems in innovation and entrepreneurship should be "place-based" has roots in Adam Smith's (1776) observation that productivity and wages were higher in areas with larger and denser populations. Marshall (1890) offered the first systematic explanation of agglomeration economies, emphasizing local supplier linkages and knowledge spillovers. The role of geographic clusters in shaping innovation and entrepreneurship outcomes began to receive serious attention in both economics and management scholarship in the early 1990s (Krugman, 1991; Porter, 1990). Efforts to conceptualize and measure economic and industrial clusters have emerged through the innovative combination of geographic data and industry classification codes (Porter, 2007). These data have allowed for econometric analysis that has shown the importance of cluster attributes in fostering entrepreneurship (Delgado et al., 2010).

Lacking from current economic, policy, and management scholarship

* Corresponding author
E-mail addresses: evanej@live.unc.edu (E. Johnson), ihemmatian@cpp.edu (I. Hemmatian), llanahan@uoregon.edu (L. Lanahan), joshia@wfu.edu (A.M. Joshi).

https://doi.org/10.1016/j.respol.2021.104398
Received 27 September 2021; Accepted 29 September 2021

is a corresponding attempt to measure the entrepreneurial ecosystem. In contrast to a geographic cluster, the ecosystem construct encompasses a greater degree of complexity, theorizing that entrepreneurs are both the products and producers of their surrounding contexts. A systems perspective is therefore required to account for nonlinear relationships and feedback effects that determine the health of these entrepreneurial contexts. As policy interest in entrepreneurial ecosystems grows (Feldman, 2020), so too does the need for metrics that capture ecosystem attributes and dynamics.

We use extant scholarship on agglomeration and geographic economies of scale as a starting point for our proposed approach for measuring entrepreneurial ecosystems. This paper defines our conception of the universe of relevant observational data, which includes firm and industry characteristics, revealed preferences of firms and government entities, collaborations among ecosystem actors, and numerous policy inputs and outputs. We use these data to build an argument for how best to capture and quantify the nuance and richness of entrepreneurial ecosystems. Our data illustrate the importance of both local ecosystem characteristics and broader determinants of ecosystem dynamics at higher levels. We thus create a database approach to measurement that connects local ecosystems to larger networks of policy and firm activity across multiple levels of jurisdiction (state, federal, and local) and institutional authority.

The COVID-19 pandemic has heightened urgency around the need for a multi-jurisdictional and multi-institutional approach for understanding the impacts of policy on local economies. For example, the United States has witnessed tremendous policy variation as states and municipalities have unleashed a patchwork of programs and restrictions that have affected firms and consumers differently and at different times. We argue that understanding these varied interactions among different economic and government actors is critical to measuring entrepreneurial ecosystems.

Further rationale for our focus on public/private sector interactions is rooted in market failure explanations of the under-provision of innovation by the private sector. Positive externalities of private innovation (Arrow, 1962) and capital market imperfections have led to the creation of government programs designed to help entrepreneurs capture the benefits of their work. The U.S. Small Business Innovation Research Program (SBIR), for example, is the largest of such programs, offering funding for both proof of concept and larger commercialization efforts by small high-tech firms. We feature such programs in our measurement approach because they represent a wide-scale source of data on the interactions among firms, universities, state governments, and federal mission agencies. We emphasize the role of the federal government in shaping these relationships in four main ways. First, the government is a major *catalyst* for cultivating human capital in STEM fields and seeding the formation of new high-tech ventures within this ecosystem (Joshi et al., 2018). Second, it serves as an overall *coordinator* of innovation policies that drive these ventures' inventive output in ways that stimulate economic development (Lanahan, Joshi, et al., 2021). Third, it acts as a *certifier* of the technical merit and/or commercial viability of the innovative output produced by these ventures (Lanahan, Armanios, et al., 2021). And fourth, it is a direct *customer* through the procurement of these ecosystem products and services by federal agencies (Hemmatian et al., 2021).

Firm-government interactions create large amounts of data, much of which is public, though underutilized. We use this information to provide a data-driven rationale for government support of entrepreneurial ecosystems. We illustrate an approach for the use of multiple large data sets that capture such interactions to better characterize and evaluate firms and ecosystems. As we detail later, these rich, overlapping data sets are a critical resource for evaluators in cases in which randomized evaluations are impractical or unethical.

Our data sources led us to construct a relational database in which firms are one type of unit nested within a system of partnerships, policy networks, and sectoral changes. We use this to offer an alternative to the status quo use of single consolidated data files for analysis. Relational Database Management Systems (RDBMS) offer a way to organize and manipulate data from many high-volume sources at once. This approach to measurement, absent from the bulk of policy and entrepreneurship literature, is well suited to establishing metrics for entrepreneurial ecosystems because of its ability to capture networks of relationships among ecosystem actors that emerge across many disparate sources of data.

Finally, our work is motivated by the need to harness the potential of publicly available data during a period of rapid data proliferation. Multiple recent projects have begun to illustrate the utility of public data through the construction of public visualization platforms designed for use by researchers, administrators, and the public. Prominent examples have so far focused on economic indicators such as consumer spending and employment rates (Chetty et al., 2020), as well as geographically specific health impacts of the COVID-19 pandemic (Fitzpatrick et al., 2020). These efforts have led to the creation of online dashboards to allow non-technical users to track geographic and temporal variation in their outcomes of interest. Our proposed measurement model and databases can enable similar dashboard projects that focus on innovation and entrepreneurship outcomes.

2. Conceptualization and Measurement of Entrepreneurial Ecosystems

The concept of entrepreneurial ecosystems has gained substantial attention from scholars and administrators seeking ways to increase economic prosperity. Under this view, entrepreneurs and their surrounding communities represent the primary actors in developing economies in response to constantly changing local conditions. Actors in these systems are highly sensitive to the context and history of their geographies. The importance of place-specific characteristics and nonlinear relationships among entities make large-scale evaluation and measurement of entrepreneurial strategy and policy extremely difficult. Indeed, one of the reasons for the focus on linear public-private sector relationships is that readily available data do not enable easy measurement of firms as pieces of a dynamic ecosystem rather than as isolated agents.

We build on previous definitions of entrepreneurial ecosystems in the following ways. First, we assemble data across 14 sources and identify tools and strategies to link these data sets together in a relational database management system. We call this system APPRISE, which stands for Accelerating Public Policy Research on Innovative Small Enterprises.[1] Second, we use these data to indicate revealed preferences of firms. Firms demonstrate intentions to engage with the government at different levels, indicating important characteristics of firm strategy that reflect on their surrounding ecosystems. Clear indicators of these preferences have grown in depth and number as government bodies have improved tracking and dissemination efforts. There are even greater opportunities to capture these preferences since the pandemic has given rise to more frequent updates to public data sources, especially those that feed into economic trackers such as those referenced above.

Entrepreneurs interact with the government in numerous ways that influence the condition of the surrounding entrepreneurial ecosystem. First, they are the recipients of government capital in the form of grants and awards that are intended to spur innovation and commercialization of new technologies. We thus treat capital flows from federal programs such as SBIR/STTR and ARPA-E as a key input to measuring the size and quality of an ecosystem. Second, entrepreneurs supply the government with products, research, and other services. Ecosystem dynamics, then, must include measurements of the density of procurement activity and

[1] For additional information on the scope and aims of APPRISE and links to available datasets integrated into the platform, see https://appriseplatform.org.

related supplier activity. Third, entrepreneurs are key collaborators for government bodies seeking to advance their own research agendas and agency mission needs. We include these collaborative relationships in our conceptualization of entrepreneurial ecosystems. Data at the level of federal grants, patents, and scholarly publications allow us to assemble measures of the extent of such collaborations at various geographic levels.

These private-public sector interactions make it even more critical that ecosystem measurement include data on geographic and temporal context, and the corresponding mixes of policy activities. Recent policy scholarship has emphasized the importance of analyzing research policy at multiple levels for a complete understanding of government funding and support for firms (Armanios et al., 2020) We, therefore, chart policy trajectories at the federal, state, and local levels by collecting information on policy change from the State Science and Technology Institute (SSTI) newsletter. We use data on policy change to denote the precise timing and geographic location of shifts in incentives and barriers for entrepreneurs and their collaborators in local settings. Our initial use of this collection procedure targets state-level policy change in response to the COVID-19 pandemic and related economic outcomes. We use these data to characterize the level of economic restriction in a given geographic area to better define the context for measuring local entrepreneurial ecosystem health.

We include several additional sets of actors as critical in defining the measurement space of entrepreneurial ecosystems. The APPRISE database system focuses on the interactions among individuals such as scientists and other academics, as well as both large and small firms. It also incorporates the behaviors of government bodies with time-sensitive indicators of state and local policy as well as federal action through various forms of support for businesses such as the U.S. Protection Payment Plan. A novel database called SEED-db provides detail on seed money startup accelerators around the world (Christiansen, 2009). These data allow us to track accelerator activity as well as acquisitions and venture capital funding for nearly nine thousand participating firms. Finally, we collect information on various research and entrepreneurship hubs such as universities and federal research labs using the Global Research Identifier (GRID) database. Each of these entities contribute to the formation and health of local entrepreneurial ecosystems. Fig. 1 illustrates some of these connections with causal arrows pointing toward outcomes at various levels.

As the diagram indicates, outcomes at one level feed back into other ecosystem dimensions, creating a dynamic system that serves as the basic structure for our measurement scheme. These feedback effects help us determine the metrics that are relevant for defining an entrepreneurial ecosystem. Such a conceptual model is in line with Arthur's work on complexity economics. Arthur (1999) encouraged policymakers to view the economy as an organic and constantly evolving system rather than as a static and deterministic one. Complexity economics is the idea that economic outcomes are not always the result of consistent and predictable patterns of behavior, but instead hinge on the organic formation of processes and systems that result in a constant evolution of novel phenomena. This lens has proven useful for scholars seeking to build the theory base around entrepreneurial ecosystems (Russell and Smorodinskaya, 2018). In such systems, innovation and commercial outputs are rarely created as the result of individual actions by producers or consumers. Rather, they arise from interactive co-creation that results from collaboration and partnership at multiple overlapping levels. Feedback effects are a critical component of these systems as actors in the entrepreneurial environment both shape and are shaped by the features of their contexts. We thus develop a measurement system that includes both individual and contextual data, as well as one that supports changes in system dynamics over time.

Our approach also emphasizes the linkages among distal ecosystems as a key component of local ecosystem health. To measure the processes and outcomes of local systems, analysts must also take into account the health of those systems with which it interacts. Recent scholarship has shown that interrelated ecosystems may not be geographically proximal. A Congressional report on the Department of Energy's SBIR program found that federally funded entrepreneurs will travel great distances to collaborate with specialists in niche technological areas (National Academies of Sciences and Medicine, 2020). A more obvious representation of this phenomenon occurs when entrepreneurs work with or from universities. These scientists may collaborate with other university faculty or staff who may be important colleagues in a given field, but who typically work at different institutions that may be quite distant.

There is very little empirical work to date that aims to establish hard metrics for the health and performance of entrepreneurial ecosystems. Stam (2018) proposed a multiplicative index that can be tailored to the ecosystem unit (e.g., region, state, country) relevant to the policy audience in question. These efforts represent a baseline means of conceptualizing ecosystem health as a composite of elements at multiple levels, mostly that of firms and surrounding communities. Stam's (2018) work relies heavily on survey data and local statistics on a sample of Dutch companies. Our approach, by contrast, focuses on a much larger universe of ecosystems across the entire United States. This allows us to measure the interconnectedness of ecosystems through broad, and often geographically dispersed, networks of actors.

These considerations have led us to build a RDBMS that we believe will vastly improve efforts to measure ecosystem dynamics. The primary aim of this project is to assemble a large network of interconnected data sources that capture important ecosystem elements, both directly through indicators like firm performance and indirectly through revealed preferences to engage with governing bodies in various ways. Public data on innovation and entrepreneurship are proliferating at an accelerating rate as governments and other organizations aim to quickly collect and publish economic metrics during the COVID-19 crisis. This rapid increase in data access is one of the chief motivations for this work and for constructing the APPRISE database system.

In addition, the empirical scholarship on entrepreneurial ecosystems, and entrepreneurship more broadly, is dominated by studies making use of privileged access to heavily protected Census data at federal Research Data Centers (RDCs). These data are heavily protected and require extensive application and approval processes that favor those with established credentials. Notwithstanding equity concerns with access to these data among scholars (particularly those who are young and members of underrepresented groups), such substantial restrictions make expedient scholarship and policy analysis very difficult. This is even more concerning in an era in which economic conditions are changing rapidly and unpredictably as the result of the global pandemic. Policy analysts and entrepreneurship scholars need ways of understanding rapid changes in entrepreneurial ecosystems as they unfold. We designed our measurement approach, and our resulting RDBMS, with these concerns in mind.

3. Data Description

We turn now to a basic description of our primary data sources before detailing the construction of our relational database management system (RDBMS.) Each data source contains a primary key, a unique identifying variable, that serves as a link between the original source and at least one other database by means of a foreign key in that database. For instance, we use DUNS[2] numbers to identify observations at the firm level and assemble a relational database structure across numerous locations. Table 1 and Table 2 below provide a snapshot of a sample of key measures from some of the major data sources in our RDBMS.

The U.S. General Services Administration System for Award

[2] A DUNS number is a unique 9-digit numerical identifier issued by Dun & Bradstreet and used by public and private sector entities to access firm-level records and transactions. See https://www.dnb.com/duns-number.html.

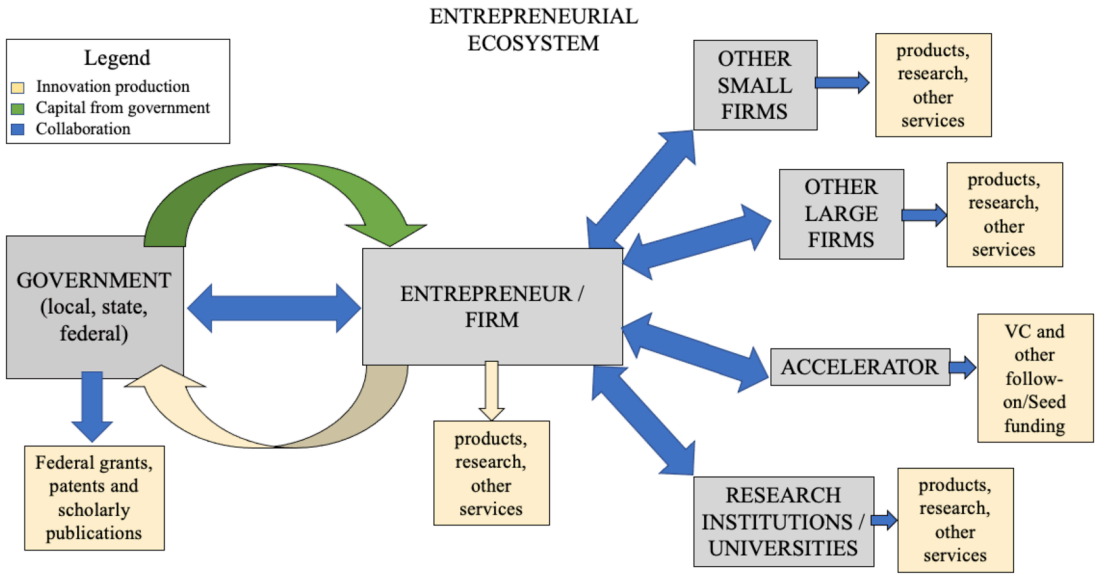

Fig. 1. Simple Conceptual Model of Entrepreneurial Ecosystem
Notes: Illustration depicts model of local entrepreneurial ecosystem with causal arrows pointing toward outcomes at various levels.

Management (SAM) is a central repository of data on any firm that has ever sought to transact with the U.S. federal government. It features transaction-level data for loans, grants, and contracts between firms and federal government agencies, providing us with rich data on firm-government interactions. Perhaps most importantly, the appearance (and disappearance) from SAM indicates important revealed preferences that distinguish some firms from others. We also use SAM to measure industry-level information through NAICS codes, as well as ownership status and intentions to support Federal Emergency Management Agency (FEMA) and other government efforts to promote public health and safety. This last measure serves as a key feature of firm- and ecosystem-level resources available to help governments respond to the COVID-19 pandemic.

The Dynamic Small Business Search (DSBS), managed by the Small Business Administration (SBA), provides a large subsample from SAM and consists of all firms that seek to participate in government outreach and commercial and/or technical assistance programs. It is also used as a tool to connect enrolled firms with potential customers, indicating another selection pathway that helps us measure ecosystem dynamics. (See Inouye et al. [2020] for an application example featuring immigrant entrepreneurs engaged in exporting.) These data also include critical firm information on employment and revenue.

USASpending provides detailed information on firm procurement relationships, as well as grants and loans. (See Hemmatian et al. [2021] for an application example focusing on the diversity of small businesses engaged in contracting with federal agencies.) Largely overlooked in the literature on firm strategy, these data represent a critical means of distinguishing firms that contract with the federal government. For instance, each contract file contains procurement codes that indicate the nature of the relationship, including codes for COVID-19-related transactions and products and services designed to aid the federal government in addressing natural disasters and supporting national interest actions.

Other key sources of data include local community characteristics at the county level, as well as geographically precise information on various industry activities from the U.S. Economic Development Administration (EDA) Cluster Map. We measure innovation outputs with the PatentsView database, as well as numerous other indicators at the firm, individual, state, and county levels. Secretary of State Offices are responsible for registering and maintaining records and status of businesses. They provide a searchable database of registered businesses in the state through their websites. As mentioned above, we track the number and characteristics of universities and other research institutions using GRID, while SEED-db provides information on accelerator activity. Finally, the ORCID database contains identifiers for scientists and scholars who may be involved in federally funded startup activity as well as collaborations with high-tech ventures through their own research institutions, which appear in the GRID database.

The relational database approach allows us to organize massive interconnected data sources simultaneously and to update those sources in an automated fashion. Accordingly, we are unable to detail each source here. Table 1 and Table 2 provide a list of a large sample of variables used in our initial construction along with primary and foreign keys for each source so that others may create their own strategies for data collection and analysis across numerous large public data sources.

4. Construction of the Database

Few entrepreneurship studies capitalize on relational database tools. One exception uses an RDBMS to manage large amounts of crowdfunding data from Kickstarter campaigns to estimate its effect on the level of angel investing in a geographic region (Yu et al., 2017). We applaud the authors' use of data science tools to help policy researchers. We build on their efforts by applying a similar set of tools to the task of entrepreneurial ecosystem measurement with public U.S. data.

Our conceptualization of entrepreneurial ecosystems implies numerous pathways to data creation, which we illustrate in Fig. 2. This figure represents the key components of the APPRISE database system. Firm- and government-level decisions both derive from, and lead to, outcomes at numerous levels of entrepreneurial activity. As an example, increased government spending on COVID-19-related research and services leads to more firm-level decisions to supply the federal government with research and products to combat viral outbreaks. These firm decisions are revealed in the USASpending database at the transaction level. Increased output from such firms feeds back into additional government grants and procurement relationships that can be traced to the SAM database through the DUNS number. Fig. 2 maps the data generation process for our RDBMS from our key data sources. The SAM and USASpending databases combine to form our initial platform for ecosystem measurement and serve as our intermediate RDBMS. This database is updated monthly by the federal government, allowing us to chart entrepreneurial activity across months for firms, cities, counties,

Table 1
Mapping the Ecosystem Measurement Framework to Relevant Publicly Available U.S. Federal Databases for APPRISE (Accelerating Public Policy Research on Innovative Small Enterprises)

Database(s) and Source(s)	Role of the U.S. Federal Government in U.S. Entrepreneurial Ecosystem			
	Catalyst	*Coordinator*	*Certifier*	*Customer*
Description of Role	Cultivates human capital in STEM fields and seeds formation of new high-tech ventures.	Guides innovation policies that drive high-tech ventures' inventive output to boost economic development.	Ascertains the technical merit and/or commercial viability of the innovative output produced by high-tech ventures.	Procures ecosystem products and services from high-tech ventures.
Government-Wide Databases				
General Services Administration (GSA) System for Award Management (SAM)	Lists firm-level preferences and interest in transacting business with the federal government since registration is required before applying for any contracts, grants, or loans.	Lists firm-level indicators for compliance with federal acquisition regulations and use of set-asides and preferences for small, disadvantaged businesses. Confirms venture eligibility for contracting purposes.	Lists firm-level indicators verifying demographics and socioeconomic characteristics of firm owners and any formal procurement certifications obtained from the SBA for small, disadvantaged businesses.	Reports agency-level contract data, upcoming contract opportunities, and wage determinations for estimating prevailing wages across labor categories.
General Services Administration (GSA) USASpending.gov	Stores firm-level and agency-level transaction records of all federal loans and grants awarded by the Small Business Administration (SBA) and other federal agencies.	Lists firm-level indicators for compliance with federal acquisition regulations and use of set-asides and preferences for small, disadvantaged businesses.	Lists firm-level indicators verifying any formal procurement certifications obtained from the SBA for small, disadvantaged businesses. Includes self-reported number of employees and annual revenues.	Stores firm-level and agency-level transaction records of all non-classified federal procurement contracts FY2000-present.
General Services Administration (GSA) Grants.gov	Provides a common website for federal agencies to post discretionary funding opportunities and for grantees to find and apply to them.	Increases awareness of federal grants, while simplifying and streamlining the grant-application and grant-evaluation processes.	Specifies any resources or capabilities as well as financial or size criteria that prospective grant applicants must satisfy to be eligible for the grants.	Searchable index of grant solicitations provides indicators of agencies' technology roadmaps that preview potential demand for products and services.
General Services Administration (GSA) COVID-19 Procurement Activities	Establishes the application process and eligibility criteria for ventures that seek to supply products and services for federal COVID-19 response.	Identifies which agencies are responsible for managing different aspects of the COVID-19 response.	Specifies the eligibility criteria, including geographic location(s), time period(s), and quality standards for responding to COVID-19 procurement requests.	Provides directory of buying guides and purchasing programs to connect public sector buyers with private sector sellers.
Federal Emergency Management Agency (FEMA)	Establishes the application process for receiving federal disaster aid for rebuilding ventures.	Identifies which agencies are responsible for managing different aspects of the disaster response.	Specifies the eligibility criteria, including geographic location(s) and time period(s) for any official federal disaster declaration.	Provides government-wide lists of products and services to be procured in response to declared disasters.
U.S. Economic Development Administration (EDA) Cluster Mapping Project	Provides dataset on the presence of clusters across geographies, based on variables including employment, wage, mobilization, labor, innovation, R&D expenditure, venture capital, scientific degrees, etc.			
Dynamic Small Business Search (DSBS)	Lists firm-level preferences for interest in participating in SBA training programs (technical assistance and exporting).	Provides searchable directory of SBA national, regional, and local offices for small business support programs.	Lists firm-level indicators verifying any formal quality certifications obtained for military and civilian applications and any formal procurement certifications for small, disadvantaged businesses.	Provides searchable directory of U. S. small businesses interested in selling their products and services to federal agencies.
SBIR.gov	Provides searchable index of grant funding opportunities.	Reports number and dollar amount of funds awarded at the firm-, agency-, and state-levels including awards made to ventures owned by members of underrepresented groups.	Tibbets Award and Hall of Fame recognizes commercially successful ventures that have leveraged their SBIR/STTR funding to generate a measurable impact and demonstrate the spirit and mission of the programs.	
Procurement Scorecard				Publishes annual agency-level scorecard with standardized metrics to (1) measure small business and socio-economic prime- and sub-contracting goals, (2) provide accurate and transparent data and (3) report progress.
Integrated Postsecondary Education Data System (IPEDS)	Provides a searchable index of institution-level indicators for all federally-funded non-profit educational institutions.	Publishes large-scale evaluations of the effectiveness of federal education programs and policies using standardized metrics.		
U.S. Patent and Trademark Office (USPTO) PatentsView		Encourages disclosure and commercialization of intellectual property specified in patent claims.	Granted patents signify an invention that is useful, novel, and non-obvious based on the examiners' evaluation of the claims.	

Notes: Table lists U.S. government information sources integrated into APPRISE.

Table 2
Mapping the Ecosystem Measurement Framework to Other Relevant Publicly Available Databases for APPRISE (Accelerating Public Policy Research on Innovative Small Enterprises)

Database(s) and Source(s)	Description
Other Databases	
Global Research Identifier Database (GRID)	GRID is a free and openly accessible database of educational and research organizations worldwide. Each organization is assigned a unique GRID ID and contains information such as institution's type, geo-coordinates, official website, email address, and Wikipedia page.
Open Researcher and Contributor ID (ORCID)	The ORCID provides a unique and persistent identifier (ORCID ID) for researchers.
Seed-DB	Seed-DB is a database of seed accelerators and the companies that have gone through them.

Notes: Table lists non-U.S. federal government integrated into APPRISE.

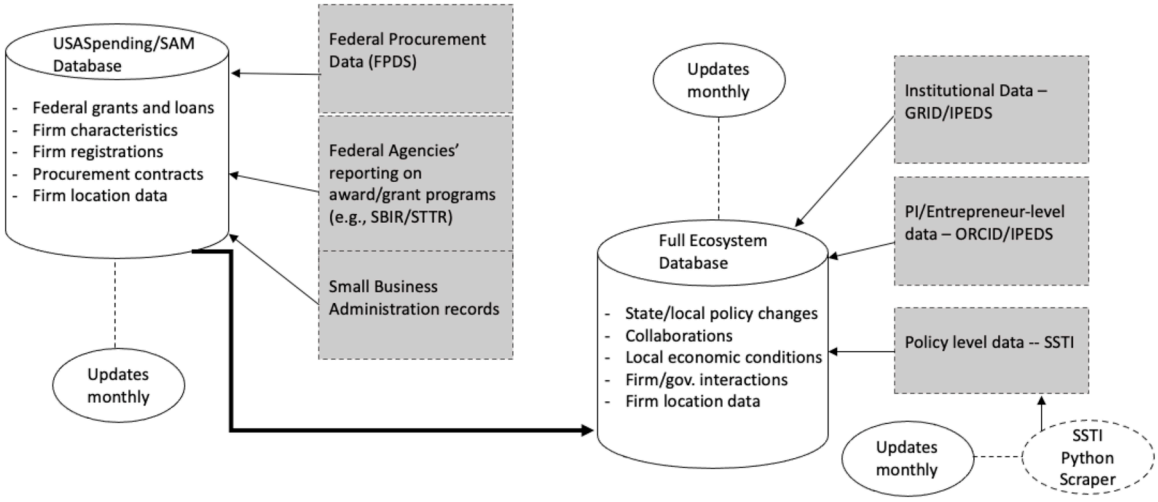

Fig. 2. Data Generating Process
Notes: Figure maps the data-generation process for the RDBMS from the key data sources listed in Table 1 and Table 2. The SAM and USASpending databases combine to form the initial platform for ecosystem measurement and serve as our intermediate RDBMS.

and U.S. states. SAM also includes detail on industry NAICS codes and physical addresses of firms, allowing for measurement of broader ecosystem effects at local geographic levels as well as fluctuations across industry types. We also incorporate detail on scientists, research institutions, and state-level policy detail into our final RDBMS, as depicted in the diagram.

The example of federally funded COVID-19 research described above represents one of many possible feedback loops within an entrepreneurial ecosystem that is contained within our database system. The goal of the present paper is not to theorize about and empirically model one such feedback process, but to explain the data management process and methodology that enable a great many such models.

In Fig. 3, we present the basic structure of our relational database. Each major data source, or entity, is represented by a table in the diagram. Relationships among tables are represented by pathways from an original entity's primary key to a foreign key in a separate table. For instance, the company DUNS number links a firm's activity in SAM to that same firm's activity in USASpending and other databases. (See Lanahan, Armanios, et al. [2021] for an application example highlighting the grants and contracting activities of SBIR awardee firms.) Structured query language (SQL) allows firm data from either of those two sources to be compiled at the county level and then linked in a separate resulting table to other county-level characteristics using the FIPS code as a primary key. These queries can then form the basis of software applications, including data visualizations and statistical models that explain the behavior of entrepreneurial ecosystems.

An RDBMS is a program or system designed for creating and managing relational databases like those discussed above. We use PostgreSQL 13 software to build and manage our database. The advantages of using an RDBMS are numerous. For one, the approach is essential for organizing unprocessed data from many disparate sources into a single, organized structure. Once in place, the RDBMS structure allows for easy collection, storage, maintenance, and retrieval of massive amounts of data. Unlike opening a single flat-file database, such as a STATA .dta file or excel spreadsheet, an RDBMS allows for access and processing of data drawn from multiple sources at once. Further, the software enables regular updates to the database as new data come in from various sources. Monthly updates to both SAM and USASpending allow for more detailed observation of the dynamic forces that shape local entrepreneurial ecosystems, including responses to state and local policy.

As mentioned earlier, each separate data source results in a table, or entity, that relates to another entity through a primary key, or unique identifier, that matches a foreign key in the connected table. This relational structure allows for easy retrieval and storage of large amounts of information across many sources. We illustrate this application in Fig. 4. Here, we map the geographic landscape of entrepreneurial activity across four dimensions. We define the sample of entrepreneurial ventures as SAM registrants that were founded between October 2014 and June 2021. In turn, this sample reflects the population of young firms with an expressed interest in transacting with the federal government. Moreover, we designate geography by U.S. county (FIPS).

Panel A in Fig. 4 reports the total count of young firm registrations in SAM; Panel B reports the count of SAM registrations from underrepresented minority-owned firms; Panel C reports the count of SAM registrations for firms operating in high-tech industries;[3] and Panel D reports the total procurement activity for the full sample of young firms. Perhaps the most striking feature is the fact that all four panels report

[3] Firms with a primary NAICS classification of 54- (Professional, Scientific, and Technical Services) or 33- (Manufacturing) comprise the high-tech sample.

Fig. 3. Excerpt from Relational Database
Notes: Figure illustrates the basic structure of APPRISE. Each major data source, or entity, is represented by a table in the diagram. Relationships among tables are represented by pathways from an original entity's primary key to a foreign key in a separate table.

entrepreneurial activity across a large portion of the contiguous U.S. While we observe a greater concentration of activity along the U.S. border and in metropolitan areas, these illustrations highlight not only the scale of activity from the data but also the potential application for future research.

5. Future Applications

The RDBMS approach enables countless specific applications in the measurement and analysis of entrepreneurial ecosystems, which we capture in the APPRISE system explained above. Our massive data collection effort over the past several years, along with the use of database management tools, improves measurement on several fronts. For one, our proposed system allows users to capture the effects of policy mixes more clearly across multiple overlapping levels of government activity. Lanahan and Feldman (2015) demonstrated the importance of state-federal policy interactions in determining the outcomes of funding for entrepreneurs. The APPRISE RDBMS facilitates further inquiry along these lines by incorporating policy activity at several levels, as well as frequent data updates to capture state and municipal policy changes that affect local economies. These changes have occurred much more frequently since state and local governments began to implement responses to the pandemic. The heterogeneity and intermittency of these responses allow users to measure ecosystem changes in response to policy at a much more granular level than what previous studies have managed.

Furthermore, our approach enables analysts to capitalize on large volumes of underutilized data. The policy sciences, and social science scholarship in general, have been slow to grasp the difference between big data sets and "big data." The former amounts to large consolidated data sets, which may result from merging several files together. "Big data," by contrast, typically refers to extremely large volumes of data from many sources, linked in a manner that makes them amenable to machine learning and semantic querying (Janssen and Kuk, 2016). The use of machine learning tools to solve problems in entrepreneurial strategy and policy has recently taken root in prominent policy and management journals (Guzman, 2017; Guzman and Stern, 2020; von Hippel and Cann, 2020). Our approach fosters this trend by providing a system for measuring ecosystem dynamics using multiple high-volume data sources at once.

Additional applications include improvements to classic research designs for policy analysis. Absent sources of exogenous variation in policy inputs, scholars often rely on regression and matching techniques to craft a plausible representation of what a firm or ecosystem would look like without receiving the relevant intervention (Lanahan, Joshi, et al., 2021; Lerner, 2000). The SAM database alone contains over 200

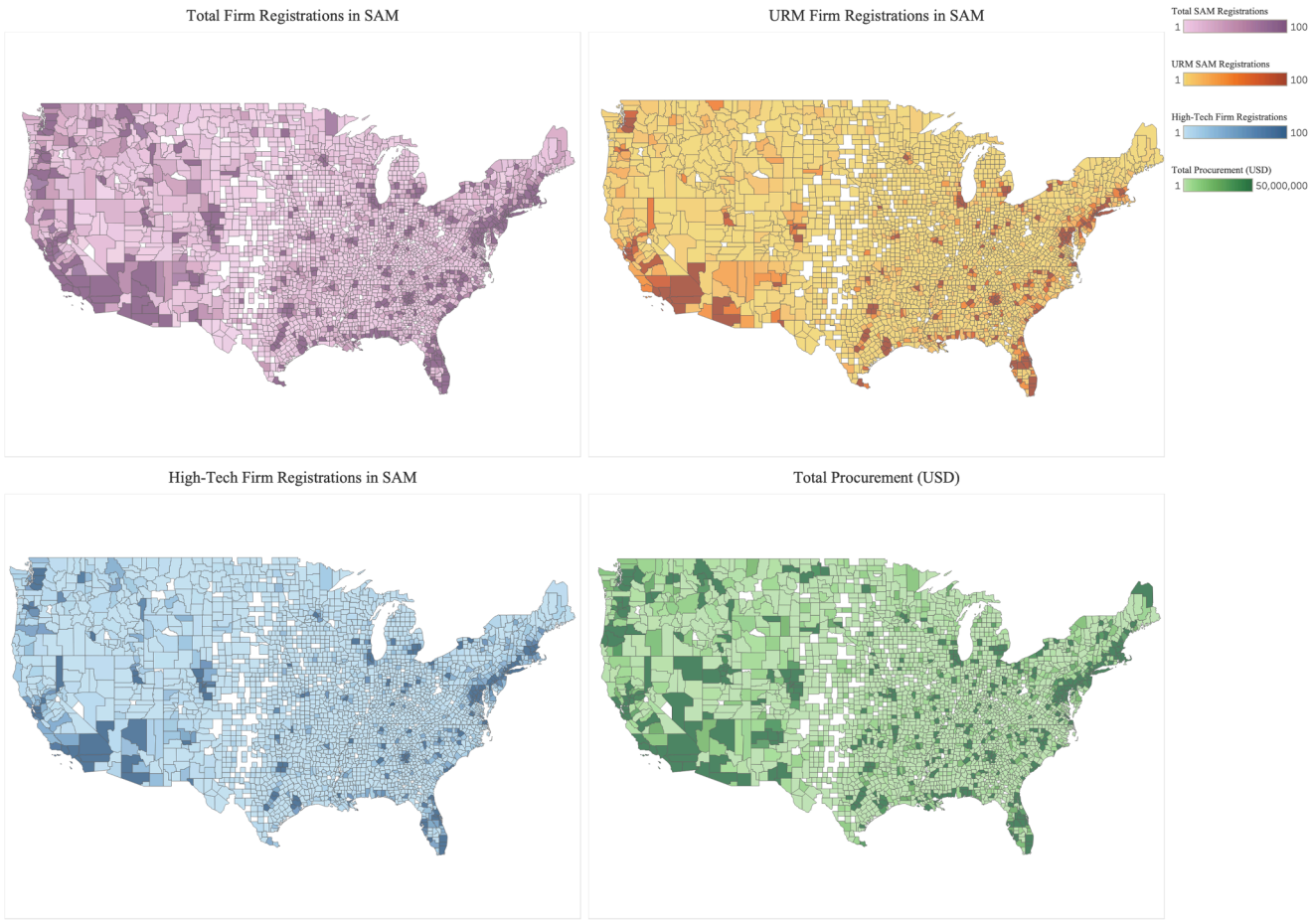

Fig. 4. Geographic Landscape of Entrepreneurial Activity
Notes: Figures report activity among firms registered in SAM that were founded between October 2014 and June 2021. Geographic designation is by U.S. county (FIPS). Panel A reports Total Firm Registrations in SAM; Panel B reports Underrepresented Minority-owned (URM) Firm Registrations in SAM; Panel C reports High-Tech Firm Registrations in SAM; and Panel D Total Procurement (USD) for young firms.

fields for potential use as covariates in causal policy models.

Since the RDBMS tracks policy changes at multiple levels of government and points in time, it provides an ideal platform for exploiting natural variation in policy variables to examine the determinants of ecosystem health. The use of multi-level policy analysis to design natural experiments has been established with regard to the SBIR program (Lanahan and Feldman, 2018). Scholars have also begun to unearth sources of exogenous variation in COVID-19 policies, capitalizing on abrupt Supreme Court decisions (Dave et al., 2020) and natural variation in the timing and extent of mask and stay-at-home orders (Janssen and Kuk, 2016; Lyu and Wehby, 2020). The use of relational database tools affords a higher degree of organization and precision in measuring exogenous sources of policy variation. This allows for easier access to natural experiments for program evaluation and causal research.

6. Conclusion

Scholarship on measuring entrepreneurial ecosystems contains few serious efforts to establish an empirical basis for translating theory to practice. Part of the explanation for the difficulty is that there are obstacles to organizing and combining data at all relevant levels and dimensions in order to capture the complexity of an entrepreneurial ecosystem. We scoured all public U.S. sources of data on firm-government interactions and augmented them heavily with rich detail on demographic, economic, and policy characteristics at multiple levels of geographic specificity to construct the APPRISE database system. Our relational database structure enables analysis of real-time data drawn from a multitude of high-volume sources at once. We highlight just a few of the many applications that this measurement system can support.

We encourage further use of our data by highlighting database methodologies as a key tool for measuring the dynamic relationships that comprise entrepreneurial ecosystems. We create a clear platform for the organization and use of rich sources of underutilized public data to measure ecosystem performance. We hope this adds value to a literature that is heavily tilted toward reliance on protected confidential data that require privileged access. We conducted a systematic review of Research Policy (RP) and found that over eighty articles from 2010-2020 made use of protected data at special access research data centers at the U.S. Census Bureau, the Bureau of Labor Statistics, and the Bureau of Economic Analysis. Over the same period only six RP papers used data from USASpending, and none made use of SAM or DSBS. Some of that deficit admittedly owes to a lack of precedent and guidance as to how best to use these high-volume public data sources. We hope our measurement system clarifies that process and perhaps levels the playing field a bit for scholars without special access to protected data.

Complex relationships among networks of actors that define an ecosystem are difficult to conceptualize, let alone measure. Our approach takes seriously the nonlinearities and feedback effects that define an ecosystem. Accordingly, we collect data that affords measurement of nonlinear processes through which firms, communities, universities, labs, and governments interact. An RDBMS is a logical tool for managing the data required for such measurement. As we have demonstrated, it enables analysis and data manipulation across a far wider array of sources than what is possible under conventional

statistical analysis. It also provides a means of handling high-volume and frequently updated databases and of feeding them directly into a number of applications for policy analysis. In this way, we have tried to push policy and entrepreneurship scholarship from "big" data to "big data," by orienting researchers toward both a set of tools and data sources that are underrepresented in the literature. We believe that the APPRISE RDBMS will help advance scholarly inquiry and generate actionable insights for policymakers, managers, and entrepreneurs to improve firm performance and enhance innovation outcomes along multiple dimensions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research is funded by the National Science Foundation's Science of Science Program (NSF 2032914) and a Kauffman Foundation Junior Faculty Fellowship grant. This project was enriched by extended discussions with panel members on the National Academies' *Review of the SBIR and STTR at Department of Energy, National Institutes of Health,* and *National Science Foundation.* These Study Committees were co-chaired by Maryann Feldman and Scott Stern. We thank Luke Coury and Natalie Short for their research assistance. We thank John Owen and Tom Newby for their IT support. We thank Amy Nuetzman for her copy-editing assistance.

References

Armanios, D.E., Lanahan, L., Yu, D., 2020. Varieties of local government experimentation: U.S. state-led technology-based economic development policies, 2000–2015. Academy of Management Discoveries 6, 266–299.
Arrow, K., 1962. Economic welfare and the allocation of resources for invention. Universities-National Bureau Committee for Economic and Council The Rate and Direction of Inventive Activity: Economic and Social Factors, pp. 609–626.
Arthur, W.B., 1999. Complexity and the economy. Science 284, 107–109.
Chetty, R., Friedman, J., Hendren, N., Stepner, M., 2020. The economic impacts of COVID-19: Evidence from a new public database built from private sector data. Opportunity Insights.
Christiansen, J., 2009. A Framework for Developing Seed Accelerator Programmes. Cambridge Master Thesis. Cambridge University.
Dave, D.M., Friedson, A.I., Matsuzawa, K., McNichols, D., Sabia, J.J., 2020. Did the Wisconsin Supreme Court restart a COVID-19 epidemic? Evidence from a natural experiment. National Bureau of Economic Research.
Delgado, M., Porter, M.E., Stern, S., 2010. Clusters and entrepreneurship. Journal of Economic Geography 10, 495–518.
Feldman, M.P., 2020. Entrepreneurial ecosystems. American Enterprise Institute Studies 1–8.
Fitzpatrick, K.M., Harris, C., Drawve, G., 2020. Fear of COVID-19 and the mental health consequences in America. Psychological Trauma: Theory, Research, Practice, and Policy 12, S17.
Guzman, J., 2017. Go west young firm: The value of entrepreneurial migration for startups and founders. MIT and National Bureau of Economic Research.
Guzman, J., Stern, S., 2020. The state of American entrepreneurship: New estimates of the quantity and quality of entrepreneurship for 32 US states, 1988–2014. American Economic Journal: Economic Policy 12, 212–243.
Hemmatian, I., Joshi, A.M., Inouye, T.M., Robinson, J.A., 2021. Exploring the effects of discretion, discrimination, and oversight on the inclusiveness of small business contracting. Entrepreneurship for Social Change. Emerald Publishing Limited.
Inouye, T.M., Joshi, A.M., Hemmatian, I., Robinson, J.A., 2020. Counteracting globalization's skeptics: How diasporas influence the internationalization preferences of minority entrepreneurs' firms. Global Strategy Journal 10, 123–173.
Janssen, M., Kuk, G., 2016. The challenges and limits of big data algorithms in technocratic governance. Elsevier.
Joshi, A.M., Inouye, T.M., Robinson, J.A., 2018. How does agency workforce diversity influence federal R&D funding of minority and women technology entrepreneurs? An analysis of the SBIR and STTR programs, 2001–2011. Small Business Economics 50, 499–519.
Krugman, P.R., 1991. Geography and Trade. MIT Press.
Lanahan, L., Armanios, D.E., Joshi, A., 2021. Inappropriateness penalty, desirability premium: What do more certifications actually signal? Organization Science. Forthcoming.
Lanahan, L., Feldman, M.P., 2015. Multilevel innovation policy mix: A closer look at state policies that augment the federal SBIR program. Research Policy 44, 1387–1402.
Lanahan, L., Feldman, M.P., 2018. Approximating exogenous variation in R&D: Evidence from the Kentucky and North Carolina SBIR State Match programs. Review of Economics and Statistics 100, 740–752.
Lanahan, L., Joshi, A.M., Johnson, E., 2021. Do public R&D subsidies produce jobs? Evidence from the SBIR/STTR program. Research Policy 50, 104286.
Lerner, J., 2000. The government as venture capitalist: The long-run impact of the SBIR program. The Journal of Private Equity 3, 55–78.
Lyu, W., Wehby, G.L., 2020. Comparison of estimated rates of coronavirus disease 2019 (COVID-19) in border counties in Iowa without a stay-at-home order and border counties in Illinois with a stay-at-home order. JAMA Network Open 3, e2011102.
Marshall, A., 1890. "Some Aspects of Competition." The Address of the President of Section F–Economic Science and Statistics–of the British Association, at the Sixtieth Meeting, held at Leeds, in September, 1890. Journal of the Royal Statistical Society 53, 612–643.
National Academies of Sciences, Engineering, Medicine, 2020. Review of the SBIR and STTR Programs at the Department of Energy. National Academies Press.
Porter, M.E., 1990. The competitive advantage of nations. Harvard Business Review 68, 73–93.
Porter, M.E., 2007. Clusters and economic policy: Aligning public policy with the new economics of competition. Institute for Strategy and Competitiveness at Harvard Business School.
Russell, M.G., Smorodinskaya, N.V., 2018. Leveraging complexity for ecosystemic innovation. Technological Forecasting and Social Change 136, 114–131.
Smith, A., 1776. The Wealth of Nations. In: Cannan, E. (Ed.), An Inquiry into the Nature and Causes of the Wealth of Nations. Oxford University Press, United States.
Stam, E., 2018. Measuring entrepreneurial ecosystems. Entrepreneurial Ecosystems. Springer, pp. 173–197.
von Hippel, C.D., Cann, A.B., 2020. Behavioral innovation: Pilot study and new big data analysis approach in household sector user innovation. Research Policy, 103992.
Yu, S., Johnson, S., Lai, C., Cricelli, A., Fleming, L., 2017. Crowdfunding and regional entrepreneurial investment: An application of the CrowdBerkeley database. Research Policy 46, 1723–1737.