

Hunting for Data Sources: How Improving Data Can Increase Capital for Emerging Domestic Markets

Glenn Yago, Betsy Zeidman, and Jill Manning

Milken Institute

It is well known that the U.S. population is rapidly diversifying. Ethnic groups now constitute majorities in four states (California, Hawaii, New Mexico, and Texas) and the District of Columbia. Within twenty years, this will be the case in nine states, including the electoral battleground, Florida.¹ Major shifts in the composition of U.S. business ownership flow from these demographic changes, with a dramatic increase in firms in emerging domestic markets (EDMs). EDM firms are ethnic- and women-owned businesses serving low- to moderate-income (LMI) populations and those located in urban and rural areas. Multiplying far faster than other businesses, they represent a growing share of entrepreneurs and job-creating small- and medium-sized enterprises (SMEs).² Given this trend, the degree to which investors recognize and respond to these changes will have significant consequences for the national economy in the years ahead.

Unfortunately, investors face challenges in tapping the potential opportunities. For financial markets to thrive, decision-makers (investors, lenders, and funders), business leaders, and public policy officials must be able to price risk and consider options effectively. To do so, they need comprehensive, reliable demographic and financial information and a well-developed infrastructure to obtain that information. The data on emerging domestic markets is fragmented, nonstandardized, and not widely accessible. One way to remedy this information gap would be to create a “data consortium” that leverages existing resources and uses them to build an integrated database. Increasing the amount and quality of data available on EDM firms would provide opportunities for improved analysis, policymaking, capital flow, and product development.

Recognizing the need for more and better EDM data, the Milken Institute’s (MI) Center for Emerging Domestic Markets undertook an in-depth review of the issue. The report, “Emerging Domestic Markets: Increasing Capital by Improving Data,” includes detailed information on our methodology, findings, and recommendations, a summary of which we provide in this article.³

1 U.S. Census Bureau, “Texas Becomes Nation’s Newest ‘Majority-Minority’ State, Census Bureau Announces,” August 11, 2005. William H. Frey, e-mail to authors, October 17, 2007.

2 For an in-depth examination of emerging domestic markets and their difficulty in accessing capital, see “A History of Emerging Domestic Markets” (Yago, Zeidman, and Abuyuan) in the Federal Reserve Bank of San Francisco’s *Community Development Investment Review* 3, no. 1 (2007).

3 Glenn Yago, Betsy Zeidman, Teresa Magula, and Jon Sederstrom, “Emerging Domestic Markets: Increasing Capital by Improving Data,” Santa Monica, CA, 2007.

Methodology and Summary of Findings

To better understand the EDM data gap, MI studied the full range of data collection efforts. In doing so, we:

- Conducted an extensive survey of the current literature on EDM data concerns;
- Interviewed more than one hundred experts, including lenders, both traditional finance and EDM experts, in the public, private, and nonprofit sectors, and researchers on the EDM market;
- Reviewed nearly seventy databases with information on EDM business and investment performance;
- Consulted with database managers;
- Explored potential products that could be developed from improved data in the course of two Financial Innovations Laboratories sponsored by the Institute.

Among our key findings:

- There are, as anticipated, numerous existing databases. This is particularly true when “EDM” is broadly defined to include people (minorities and women) and places (urban areas, LMI communities), and when the data of interest include demographic information and financial transactions. We chose the broad definition to capture as many sources as possible.
- Data are collected by a variety of types of organizations, with potentially divergent data needs. Compared to other types of organizations, government agencies maintain the highest number of EDM databases.
- Databases use a variety of survey units (making comparison difficult). Of those we surveyed, forty-five use “businesses” as their survey unit; nine use “lenders,” eight use “individuals,” and six use “transactions.”
- While significant data overlaps exist, so do substantial holes, in terms of quantity and quality of the data.
- There is little information covering the financing of EDM businesses at the transaction level, except in proprietary databases of financial institutions or information management companies.
- Differences in definitions and terminology hinder comparability, even among databases with the same survey unit.
- Our proposed data consortium differs from current efforts to improve data in that it attempts to capture the full range of EDM businesses and would include both demographic and financial data. It would be tailored to investors’ needs.
- There is definite interest in forming a data consortium and gaining access to an integrated EDM database, but concern exists regarding privacy of the EDM business, as well as the privacy of the data source’s information. We feel there are adequate legal and technological solutions to these privacy concerns.

The Data Matrix – Organization of Findings

We grouped the organizations collecting data into four key groups and identified them as: financial institutions/funds (12 EDM databases); government agencies (21); trade associations, nonprofit organizations, and research groups (26); and information management companies (8).⁴ In an effort to compare the different data pools, and to clarify overlaps and gaps, we created a matrix of data providers. An abbreviated version of this matrix appears at the end of this article. A discussion of organization type and the data provided by each follows.

Financial Institutions/Funds

Financial institutions, specifically banks, generally have robust EDM databases. Either within a larger data set or separated for reporting on Community Reinvestment Act (CRA) activity, banks collect and store demographic data and information on small-business owners, small businesses, loan details, and loan performance. Because of privacy concerns, financial institutions do not release information on the size of their databases, but we believe they are among the largest, second only to credit-reporting agencies (categorized as “information management companies” and detailed in a later section). They are also the least willing to share data, as they are proprietary and central to their business. In general, banks collect information at the company/business level rather than by transaction.

Nonbanks also collect data on borrowers. For example, Allied Capital’s small-business lender, Business Loan Express (BLX), collects data on each transaction performed. While its databases are smaller because of the size of the population served, they still rank among the largest databases we identified. Like banks, nonbanks in general are not willing to share proprietary information.

On a similar note, the EDM-targeted investment funds (National Association of Investment Companies members, for example) track performance and, frequently, demographics for their own purposes. Funds in which public pension funds have invested increasingly release performance data because of Freedom of Information Act requests, but not on a company level. With the increasing interest in double-bottom-line investments (sometimes known as blended-value or hybrid investments), there are efforts to track the investments’ financial returns and collateral benefits, such as job creation, environmental mitigation, workforce benefits, and urban redevelopment. One source of such data is the Research Initiative in Social Enterprise (RISE). RISE’s annual Social Investor Survey provides a national database of investments whose products, services, or business structures can be considered to have positive social or environmental impacts. The data are available in the RISE “Double Bottom Line Investor Directory,” a searchable public database of these funds.

4 Note that the organization type “financial institutions/funds” includes both debt and equity. “Information management companies” collect and sell data related to all types of businesses. Such companies include Dun & Bradstreet, Fair Isaac, VentureOne, and the three major credit-reporting agencies: Equifax, Experian, and TransUnion.

Government Agencies

Government agencies, led by the U.S. Census Bureau, have abundant demographic data on business owners. Studies such as the “Survey of Business Owners and Self-Employed Persons,” “National Longitudinal Surveys,” and the “Panel Study of Income Dynamics” capture the gender and ethnicity of business owners, as well as the location and age of their businesses. The Small Business Administration (SBA) has a number of databases covering its lending programs. These databases are potentially important because they would enable the creation of proxies for comparisons by income level or business type.

For example, Standard & Poor’s analyzed default data from the SBA 504 loan program to generate a risk model and develop ratings for the Community Reinvestment Fund’s Series 17 and Series 18 note sales (community development small-business loans). These ratings were critical in enabling institutional investors to purchase the notes, a first in the community development field. In addition to demographic data, the SBA has amassed information on small-business cash flows, financing, and repayment performance. The data are generally publicly available, although access may be limited by regulation or process. Sole proprietorships, for instance, report revenues on the owner’s personal income tax form, which would not be accessible to researchers.

Trade Associations, Nonprofit Organizations, and Research Groups

Trade associations, nonprofit organizations, and research groups tend to have mission-specific databases that vary greatly in size and content. Detailed data are generally available by request, although access may be restricted to members. In most cases, aggregate data are available online or in the form of reports or white papers. Three general subgroups exist among these databases:

- Databases that collect information only on the business and the business owner (location, number of employees, revenues, ethnicity, gender), including the Kauffman Financial and Business Research Database and the four databases from the Initiative for a Competitive Inner City.
- Databases like the Brookings Institution’s Urban RPM Investor and the PRI Makers Network, which collect information on enterprise financing but do not survey the characteristics of the business or owner.
- Databases like the Business Consortium Fund of the National Minority Supplier Development Council, which collect data on both demographics and business financing.

Information Management Companies

Information management companies, such as Dun & Bradstreet, Fair Isaac, VentureOne, and the three major credit-reporting agencies (Equifax, Experian, and TransUnion), collect and sell data related to businesses of all types and have some of the most extensive databases. Dun & Bradstreet markets general business information—ownership, location, size, age, and cash flows—used by its clients to decrease risk exposure and increase sales. VentureOne sells venture capital data: firm location, cash flows, and financing. Although these databases are

not specific to small or EDM businesses, many EDM firms are captured. Fair Isaac captures hundreds of thousands of individual transactions and uses them to model predictive risk. Its databases are solely for the use of its clients.

Among the largest databases, holding millions of records each, are those of the three major credit-reporting agencies. These records are proprietary, strictly regulated by the Federal Trade Commission, and are not available for general pooling. Although they cannot share their data directly, credit agencies are interested in reviewing other pooled data as a means to explore alternative credit-scoring mechanisms.

Overall Data Quantity and Quality

Our review of the EDM databases made it clear that substantial holes exist in terms of the quantity and quality of the collected data. Many data sets have a narrow focus, capturing only information on demographics, specific industries, or individual cities. Although the scope of these databases meets the needs of the groups maintaining them, few effectively cover EDM businesses as a market. For example, the U.S. Census (the source of the Survey of Business Owners) is the best source for demographic information, but it does not capture data pertaining to business financing. In addition, few databases cover the financing of EDM businesses at the transaction level, except in proprietary content of financial institutions or information management companies. Yet this information is the most critical for assessing business opportunities and improving capital flows.

Furthermore, a large number of EDM databases contain relatively few survey units. The Venture Capital Fund Database, owned by the National Association of Investment Companies, has information on twenty-four venture capital funds. The Aspen Institute's MicroTest Database and the Calvert Foundation's Profiles Databases likewise consist of seventy-five and seventy-one survey units, respectively. Alone, they do not provide sufficient information on EDM businesses to allow potential investors to analyze the markets effectively.

In terms of quality, existing EDM data are self-reported and their accuracy is unverified. Dun & Bradstreet, for instance, manages an extensive database covering 366,000 minority-owned businesses and 1.4 million women-owned firms.⁵ However, identification of a firm as minority- or women-owned depends solely on voluntary disclosure by the business.

Proposed EDM Data Solutions

A review of key literature uncovers several potential solutions to the need for improved EDM data. The Information Policy Institute analyzes nontraditional information sources that could be used to bring minorities, low-income individuals, and others with insufficient credit information into the credit system. It suggests that consumer information from utility, child-care, and health-care providers, along with data from auto insurance companies and

⁵ Interview with Darren Elsner, Dun & Bradstreet, July 29, 2005.

rental agencies (housing, furniture, and consumer durables) could be used to evaluate “thin-file to no-file” (TFNF) loan applicants.⁶

An obstacle to the use of these nontraditional data sources could be the limitations inherent in voluntary reporting. To that end, Afshar recommends that the public sector offer incentives to potential data providers (such as the utility companies) to overcome economic costs and possible regulatory barriers.⁷ Increased reliance on this alternative transaction data has the potential to provide many TFNF individuals with expanded access to credit. To date, most activity with alternative data involves consumer, as opposed to small-business, credit. However, personal credit history is a significant predictor of small-business credit risk.⁸

The Minority Business Development Agency, the Small Business Administration, Weissbourd, and Weissbourd and Berry all suggest that government, nonprofit organizations and for-profit companies collaborate to improve EDM data.⁹ The federal government and for-profit companies could make existing databases more accessible and augment their value by disaggregating the data.¹⁰ Similarly, Hawke recommends combining U.S. Census, private marketing, and nontraditional sources to better understand the economic importance of EDM.¹¹

Creating a Data Consortium

Our report, “Emerging Domestic Markets: Increasing Capital by Improving Data,” recommends taking this collaborative approach of combining existing data sources. A central database, populated by multiple entities and managed by a third party, would make the data accessible to a number of people. Contributors would agree to a standard set of definitions and reformat their data to these standards to enable comparison across databases. In exchange for contributing data, members of this “data consortium” would gain access to the contents of the entire database. They would be free to run searches, download records, and analyze data. Thus, they would dramatically increase their understanding of EDM business and investment performance, and they would be better able to identify EDM investment and lending opportunities.

6 Information Policy Institute. “Giving Underserved Consumers Better Access to the Credit System: The Promise of Non-Traditional Data,” 2005.

7 Anna Afshar, “Use of Alternative Credit Data Offers Promise, Raises Issues.” Federal Reserve Bank of Boston, 2005.

8 Interview with Thomas C. Wise, Fair Isaac, May 30, 2006.

9 U.S. Minority Business Development Agency, “Accelerating Job Creation and Economic Productivity: Expanding Financing Opportunities for Minority Businesses,” 40; U.S. Department of Commerce, 2004; U.S. Small Business Administration, Office of Advocacy, “The Small Business Economy” (Washington, DC: U.S. Government Printing Office, 2004); Robert Weissbourd, “Banking on Technology: Expanding Financial Markets and Economic Opportunity” (Washington, DC: Brookings Institution, 2002); and Robert Weissbourd and Christopher Berry, “The Market Potential of Inner City Neighborhoods: Filling the Information Gap” (Washington, D.C: Brookings Institution, 1999).

10 U.S. Small Business Administration, Office of Advocacy, 2004.

11 John Hawke, “Growing Diverse Banking Markets: Going Beyond Traditional Measures,” Comptroller of the Currency Administrator of National Banks Community Development Newsletter, 2001.

The primary value of a data consortium is its ability to motivate increased interest and investment in EDM. Using its data, members could model new structures, develop and refine products tailored to EDMs, generate data, increase understanding of the market, and deploy additional capital. Product development would become an ongoing activity, fed by the information contained within the consortium database. Historically, financial innovation has significantly broadened access to capital, for example, the increase in small-business lending by large banks upon adoption of small-business credit scoring. Of particular interest to the individuals interviewed by the Milken Institute were EDM-tailored credit-scoring models (similar to those being developed for consumers, but reflective of business lending data) and structured finance vehicles leveraging multiple sources of capital, such as private funds, philanthropic contributions, and government guarantees. In addition, the data consortium, with a robust database, would significantly enhance the ability to securitize EDM loans by providing methods and data for risk assessment and the establishment of proxies and synthetic structures as needed. The richness of the data would inform investors and lenders and also help establish pricing.

Several notable efforts are under way to pool and cross-reference data within subsectors of the EDM field: the Community Development Financial Institutions (CDFI) Data Project, the National Association of Investment Companies (NAIC) project research on returns, and the Research Initiative in Social Enterprise (RISE) surveys of community impact double-bottom-line investment funds, for example. However, current pools of data remain fragmented, and many lack sufficient size, scope, or format. Our work differs in that it attempts to capture the full range of EDM businesses and includes both demographic and financial data. The database would be tailored to the needs of investors.

Data Content Considerations

Because each organization currently employs a unique data format, submissions would have to be standardized. Each consortium member—the organization supplying the data—should be responsible for reformatting its own data, for the following reasons:

- *Greater data integrity.* Each organization knows its data well and is better equipped to work with its own information. Placing the onus on each organization to ensure properly formatted data will reduce errors and increase the accuracy and integrity of the data.
- *Reduced costs for the consortium.* The consortium would be relieved of the time and labor needed to ensure that data are properly formatted, which would greatly reduce costs.
- *Simple data submission.* The consortium would determine what information to collect and would define a standard data format, a schema using XML, a common markup language for documents containing structured information. Consortium members would then either periodically send XML files to the database administrator or have a web-based interface for submitting data themselves.

- *Access to data.* In exchange for contributing data, members would gain access to the contents of the entire database. They would be free to run searches, download records, and analyze data. Thus, they would dramatically increase their understanding of EDM business and investment performance, and they would be better able to identify EDM investment and lending opportunities.

Consortium members would also have to adopt common definitions. EDM data consistency is undermined by a lack of common definitions. Concepts such as “LMI” and “ethnic-owned business” vary, as do calculations of the rates of financial and social return. Differences in definitions and terminology hinder comparability, even among databases with the same survey unit. For example, the Survey of Business Owners (SBO) defines a business as any nonfarm or nongovernmental business that filed a tax form as an individual proprietorship, partnership, or any type of corporation, and with receipts of \$1,000 or more. The Kauffman Index of Entrepreneurial Activity uses household survey data and measures individual business owners, defined as individuals from the age of twenty to sixty-four who own a business as their main job with fifteen or more hours worked per week.

Possible uses of consortium data are numerous. They would likely vary with user type and would require enhanced database functionality. Profit-driven entities might want to mine the data to evaluate investment and lending opportunities, while nonprofits might seek to track social returns. Researchers might want to explore relationships, such as those between total loans disbursed and geographic location. Thus, the database would have to have significant flexibility to meet each user’s needs.

To ensure the privacy of the borrower or equity recipient, the name, address, and contact information would be replaced with a unique identification number. Multiple loans to one borrower could then be linked to a single customer identification number, but the identity of the customer would remain anonymous.

The identity of the data provider could be hidden through two mechanisms:

- *A standardized format.* Converting all data records to a common format would, at a minimum, make it harder to attribute a data record to a specific organization.
- *Recruiting a significant number of organizations of each type.* Even with a standardized format, it might still be possible to guess what type of organization, financial institution, nonprofit, or trade association supplied a given record. Another layer of masking could be accomplished by recruiting a significant number of organizations of each type into the consortium.¹²

¹² Take the example of a consortium comprised of the CDFI Data Project and two banks, Bank A and Bank B. The data collected by the CDFI Data Project would be recognizably different from bank data because it would track social returns. Therefore, Bank A should be able to attribute any record that did not track social returns, and was not its own, to Bank B. This could give Bank A a competitive advantage. However, adding more contributors removes this risk. Take the example of a consortium comprised of ten nonprofits and ten banks. Bank A should be able to determine if a data record belonged to a competing bank, but it would not be able to identify which one.

The consortium would need to monitor data accuracy. While the cost of an independent data watchdog would likely be prohibitive, consortium members could be asked to watch for suspicious records and report abnormalities. Penalizing members with expulsion from the consortium for providing false data could encourage them to scrutinize their own data before submitting it.

Additional issues to consider include ownership and funding of the database. The initial establishment of a consortium would require philanthropic support. However, long-term ownership, financial responsibility, administration, and maintenance should be shifted to members. They would provide the data, and they should retain ownership of that data. Any transfer of ownership would present a conflict of interest related to potential resale opportunities. Furthermore, because members would directly benefit from the consortium, they should pay for it. If they chose, the membership could offset expenses by selling access subscriptions to nonmember financial institutions and research organizations. Alternatively, database users could pay a fee based on the number of queries or downloads.

Recommended Format: Relational Database

The database could take several forms: relational, dimensional, or object. For a variety of reasons, the recommended format is a relational database. A relational database uses a two-dimensional structure of rows (records) and columns (fields) to store data. All data are stored inside tables and operations are performed on the tables. Operations include the retrieval of a subset of columns, a subset of rows, or selected records and columns from multiple tables to create a new table based on their intersections. Relational databases must adhere to basic rules. Each record in each table must be unique and correspond to a primary key, a field that provides record identification. Each column must have a unique name. Entries in the same column must be of the same kind, and no column value can depend on any other column value other than the primary key. The database also must be normalized—each table should include unique fields that are not redundant with other fields within that table.

A relational database has several advantages over other models. First, it provides a description of the data using its natural structure only; no additional programming is required to structure the database.¹³ Second, it allows for easy normalization “to ensure data consistency and stability, to minimize data redundancy, and to ensure updatability and maintainability of the data.”¹⁴ Finally, and perhaps most important for the consortium, the relational model is mature and widely used. Most organizations have experience with such models and are comfortable using common query languages like SQL to manipulate data. Thus, the use of a relational database should increase the ability of members to create their own programs to search and export data from the database, thereby lessening the responsibilities of the database administrator and reducing overall costs.

13 E. F. Codd, “A Relational Model of Data for Large Shared Data Banks,” *Communications of the ACM* 13, no. 6 (June 1970): 377–87.

14 Tore Bostrup, “Introduction to Relational Databases—Part 1: Theoretical Foundation.” 15 Seconds. <http://www.15seconds.com/issue/020522.htm>.

Conclusion

In this issue of the *Review*, Federal Reserve Chairman Ben S. Bernanke stresses the importance of community economic development data. “By making companies, entrepreneurs, and investors aware of the new opportunities,” he said, “and by promoting competition in underserved areas, such information helps put market forces in the service of community development.”¹⁵ A well-constructed data consortium could help eliminate information barriers and unleash the dynamism of the financial markets through knowledge building and product development. Ultimately, both the emerging domestic markets and the national economy would benefit.

References

- Afshar, Anna. “Use of Alternative Credit Data Offers Promise, Raises Issues.” Federal Reserve Bank of Boston, 2005.
- Bostrup, Tore. “Introduction to Relational Databases, Part 1: Theoretical Foundation.” 15 Seconds, <http://www.15seconds.com/issue/020522.htm>.
- Clark, Catherine H., and Josie Taylor Gaillard. “RISE Capital Market Report: The Double Bottom Line Private Equity Landscape in 2002–2003.” Research Initiative on Social Enterprise (August 2003).
- Codd, E. F. “A Relational Model of Data for Large Shared Data Banks.” *Communications of the ACM* 13, no. 6 (June 1970): 377–87.
- Hawke, John. “Growing Diverse Banking Markets: Going Beyond Traditional Measures.” Comptroller of the Currency Administrator of National Banks Community Development Newsletter (2001).
- Information Policy Institute. “Giving Underserved Consumers Better Access to the Credit System: The Promise of Non-Traditional Data.” 2005.
- Robins, Charles, and Robert Toomey. “Keeping the ‘Private’ in Private Equity: Dealing with FOIA Concerns.” Weil, Gotshal & Manges, March 2004.
- U.S. Department of Commerce, Minority Business Development Agency. “Minority Purchasing Power: 2000 to 2045.” September 2000.
- U.S. Minority Business Development Agency. “Accelerating Job Creation and Economic Productivity: Expanding Financing Opportunities for Minority Businesses.” 40: U.S. Department of Commerce, 2004.
- U.S. Small Business Administration, Office of Advocacy. “The Small Business Economy.” Washington, D.C.: U.S. Government Printing Office, 2004.
- Weissbourd, Robert. “Banking on Technology: Expanding Financial Markets and Economic Opportunity.” Washington, D.C.: Brookings Institution, 2002.
- Weissbourd, Robert, and Christopher Berry. “The Market Potential of Inner-City Neighborhoods: Filling the Information Gap.” Washington, D.C.: Brookings Institution, 1999.
- Yago, Glenn, Betsy Zeidman, and Alethea Abuyuan. “A History of Emerging Domestic Markets.” *Community Development Investment Review* 3, no. 1, 2007.
- Yago, Glenn, Betsy Zeidman, Teresa Magula, and Jon Sederstrom. “Emerging Domestic Markets: Increasing Capital by Improving Data.” Santa Monica, CA, 2007.

¹⁵ Remarks by Chairman Ben S. Bernanke, Greenlining Institute’s Thirteenth Annual Economic Development Summit, Los Angeles, April 20, 2006.

Community Development INVESTMENT REVIEW

Exhibit A. Matrix of Emerging Domestic Markets Databases

Data source	Database overview				Characteristics of individual			Characteristics of business ^a					Performance of financing				Interest in data consortium ^b				
	Survey unit	Survey unit details	Number of units surveyed	Availability	Earliest year and frequency	Race / ethnicity	Gender	Geographic location	LMI location	Age of business	Number of employees	Financial information	Type of financing	Financing amount	Cost of financing	Other terms of financing	Rate of return	Rate of default	Social return	Willing to share data	Interest in consortium data
Board of Governors of the Federal Reserve System Survey of Small Business Finances (SSBF)	Business	Firms with fewer than 500 full and part-time employees (including owners working in the firm)	3,200 to 4,600 (exact number varies with year)	Free online for 1987, 1993, 1996, and 2003	1987, About every 5 years (1987, 1993, 1996, 2003)	Y	Y	Y, not on public file other than census division	N	Y	Y, including owners	Y	Y	Y, limited to most recent loan only	Y	Y, return rate on assets or equity can be calculated from balance sheet	N, know if rate on assets in past 3 years and if firm / owner filed for bankruptcy in past 7 years	N	Y	Y, available online	Y
Federal Reserve Board Flow of Funds Accounts	Business	Nonfirm, noncorporate businesses	Aggregate of all corporations' IRS returns	Available online for download	1945, Quarterly	N	N	N	N	N	Y, aggregate	Y, aggregate	Y, aggregate	N	N	N	N	N	Y, available online	N	
SBA & National Community Reinvestment Coalition (NCRC) Community Express Loans	Business	Community borrowers - minority, women, and veteran-owned businesses in lower income areas	Approximately 17,000	Captured in 7(a) database; Can obtain separate by request to NCRC or SBA	1999, Quarterly	Y	Y	Y, city, state, or Census tract	Y	N	N	Y, For recent loans	Y	Y	Y	Y	Y	N	Y, not all	Y	
SBA 504 Loan Program	Business	Business	Approximately 200,000 (7(a) & 504 combined)	Free online (aggregated by city), more detailed available by request	1996, Annual	Y	Y	Y, city or county	N	Y, not consistent	Y, for recent loans	Y	Y	Y	Y, individual status or aggregate return	Y	Y	N	Y, not all	Y	
SBA 7(a) Loan Program	Business	7(a) loan recipients - Small, for-profit businesses (size limit varies with industry)	Approximately 200,000 (7(a) & 504 combined)	Free online (aggregated by city), more detailed available by request	1996, Annual	Y	Y	Y, city or county	N	Y, not consistent	Y, for recent loans	Y	Y	Y	Y, individual status or aggregate return	Y	Y	N	Y, not all	Y	
SBA Microloan Program	Business	Microloan recipients - For Profit businesses (generally less than 5 employees)	Approximately 24,000	Available by request	1992, Annual	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N	Y, not all	Y	
SBA New Markets Venture Capital	Business	Small businesses invested in by a New Market Venture Capital Company	Approximately 55 (there are 6 New Market Venture Capital Companies)	Generally not available to the public	2001	Y	Y	Y	Y, beginning in 1998	Y, incomplete and unreliable	Y, incomplete and unreliable	Y	Y	Y	N	N	N	N	Y, but depends on the situation	Y, depending on content	
SBA Small Business Investment Company (SBIC)	Business	Small businesses invested in by an SBIC	Approximately 35,000	Generally not available to the public	1997, Annual	Y	Y	Y	Y, beginning in 1998	Y, incomplete and unreliable	Y, incomplete and unreliable	Y	Y	Y	N	N	N	N	Y, but depends on the situation	Y, depending on content	

Exhibit A. Matrix of Emerging Domestic Markets Databases

Data source	Database overview				Characteristics of individual?				Characteristics of business?				Characteristics of financing				Performance of financing		Interest in data consortium?		
	Survey unit	Survey unit details	Number of units surveyed	Availability	Earliest year and frequency	Race / ethnicity	Gender	Geographic location	LMI location	Age of business	Number of employees	Financial information	Type of financing	Financing amount	Cost of financing	Other terms of financing	Rate of return	Rate of default	Social return	Willing to share data	Interest in consortium data
CDFI Fund Community Investment Impact System (CIS)	Lender	CDIs and CDFIs	Approximately 300	Data not currently available to the public; plans to make data available in the future	2004, Annual	Y	Y	Y	Y	Y	Y, revenues	Y	Y	Y	Y	N	Y	Y, jobs created, housing developed, and technical services provided	Y, in the future	Y	
Federal Deposit Insurance Company (FDIC)/Call Reports	Lender	Banks	Approximately 8,000	Free online; Data aggregated by bank	1998, Quarterly (electronic version)	N	N	Y	N	N	Y	Y	Y	N	Y	N	Y	N	Y	N	
Federal Financial Institution / Funds Examination Council Community Reinvestment Act (CRA) Reports	Lender	State banks, national banks, and large saving associations (\$250M+)	Approximately 2,000	Free online; Data aggregated by census tract level	1996, Annual	N	N	Y, state, county, and MSA	Y	N	Y, census tract	Y	Y	N	N	N	N	N	Y	N	
Information Management Companies																					
Don & Braintree (DOB)	Business	Companies	7.9 million	For purchase; Price based on the number of records	1941; 1969 electronic records, Monthly	Y, if offered by owner	Y	N	Y	Y	Y	N	N	N	N	N	N	N	Y, available for purchase	N	
VentureOne	Business	Venture-backed firms	Approximately 18,000 U.S. firms	By subscription only	1987 Quarterly	Y	Y, city	N	Y	Y	Y	Y	Y	N	N	N	N	Y, available for purchase	N		
Equifax	Individual	Not disclosed	A file for every credit active adult in the country	For sale based on permissible purpose as governed by the FTC	Not disclosed	N	Y	N	Y	N	Y	Y	Y	N	Y	N	Y	N, restricted by FTC regulations on sale of data	Not disclosed		
Experian	Individual	Not disclosed	A file for every credit active adult in the country	For sale based on permissible purpose as governed by the FTC	Not disclosed	N	Y	N	Y	N	Y	Y	Y	N	Y	N	Y	N, restricted by FTC regulations on sale of data	N		
TransUnion	Individual	Credit tradeline history, public records	A file for every credit active adult in the country	For sale based on permissible purpose as governed by the FTC	Online file updated daily, only the current version is available	N	Y, street address	N	Y	N	Y	Y	Y	N	Y	N	Y	N, restricted by FTC regulations on sale of data	N		
BankLab DataBanc	Transaction	Small business loans	Over 700,000	For purchase (with analysis)	1983, Annual	N	Y, varies	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y, data and analysis available for purchase	N		

Exhibit A. Matrix of Emerging Domestic Markets Databases

Data source	Database overview				Characteristics of individual				Characteristics of business				Characteristics of financing				Performance of financing		Interest in data consortium?	
	Survey unit	Survey unit details	Number of units surveyed	Availability	Earliest year and frequency	Race / ethnicity	Gender	Geographic location	LMI location	Age of business	Number of employees	Financial information	Type of financing	Financing amount	Cost of financing	Other terms of financing	Rate of return	Rate of default		Social return
Fair Isaac Small Business Scoring Service	Transaction	Individual transactions	Approximately 1 million	Private	1990s, Regularly	N	N	Y	N	Y	Y, if transaction has business listing from DDB or Experian	Y	Y	N	N	N	Y	N	N	N
VentureXpert	Transaction	Venture capital funds	Approximately 7,500	For purchase	1978, Quarterly	N	N	Y, ZIP code and MSA	N	Y	Y	Y	Y	Y, management fee	Y, average investment size, round, etc.	Y	N	N	N	Y, available for purchase
Xigi.net	Transaction	High-performing social enterprises	Various	In development; Available online	2006	In development	In development	In development	In development	In development	In development	In development	In development	In development	In development	In development	In development	In development	In development	Y
Nonprofit Organizations and Research Groups																				
Aspen Institute MicroTest Outcomes Data Set	Business	Micro-enterprises in underprivileged communities	Sample of approximately 800 each year	Private	2003, Annual	Y	Y	N	N	Y	Y, revenues	Y	Y, aggregate loans from participants	N	N	N	N	N	Y, effect on owner's household	Y, only if approval and privacy concerns are met
Brookings Institution Urban RPM	Business	Urban retailers	Undetermined	Still in development	The dataset will include 10-15 years prior to the current year	Y	Y	Y, ZIP code	Y	Y	Y	N	N	N	N	N	N	N	N	Y, only with approval of members
Brookings Institution Urban RPM Investor	Business	Urban retail investors and lenders	Undetermined	Still in development	The dataset will include 10-15 years prior to the current year	N	N	Y, ZIP code	Y	N	Y	Y	Y	Y	Y	Y	Y	N	Y, only with approval of members	Y
Community Development Technologies Center/ Merril Lynch Southern California Minority Business Atlas	Business	Business owners	1,200	Aggregated data available online	1999-2000	Y	Y, county	N	Y	Y	Y, sales	Y, financing source	N	N	Y, types of financing needed	N	N	N	Y	Y
Initiative for a Competitive Inner City (ICIC) Inner City 100	Business	Companies	445 companies	Limited data made public online and in Inc. Magazine. Remainder is private	1999, Annual	Y	Y	Y	N	Y	Y	Y	N	N	N	N	N	N	Y, but only data already public (growth rate, most recent FY revenue, number of employees)	Y

Exhibit A. Matrix of Emerging Domestic Markets Databases

Data source	Database overview				Characteristics of individual				Characteristics of business ²				Characteristics of financing				Interest in data consortium?			
	Survey unit	Survey unit details	Number of units surveyed	Availability	Earliest year and frequency	Race / ethnicity	Gender	Geographic location	LMI location	Age of business	Number of employees	Financial information	Type of financing	Financing amount	Cost of financing	Other terms of financing	Rate of return	Rate of default	Social return	Willing to share data
Kauffman Financial and Business Research Database	Business	Companies	Approximately 900,000 for demographic information, 300,000 for financial information	Up to 15,000 records available upon request	1983, Annual	Y, minority owned/ind. owned/CEO/Owner	Y	Y, addresses	N	Y, as of current management	Y	Y	N	N	N	N	N	N	Y	Y
Kauffman Firm Survey	Business	New businesses - Started operations in 2004	5,000	Still in development	2004, Annual for 4 years	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y
Kauffman Index of Entrepreneurial Activity	Business	Individual business owners	7,500,000 total sample size	Available by request	1986, Annual	Y	Y	Y, regions only	N	Y, measures number of newly created businesses	N	N	N	N	N	N	N	N	Y	Y
National Minority Supplier Development Council (NMSDC) and Business Response Panel (BGR Capital)	Business	Minority-owned business	15,000	Some data available free online with most by approved request only	1985, Annual	Y	Y	Y, city	N	Y	Y, financial statements and five-year worth projections	Y, three years worth	Y	Y	Y	Y, but depends on the situation	Y	Y, jobs created	Y	Y
NMSDC	Business	Minority-owned businesses	Approximately 15,000	Available only to dues-paying national corporate members	1974, Monthly	Y	Y	Y, ZIP code	Y	Y	Y, limited	N	N	N	N	N	N	N	Y	Y
RIS/Investors Circle/Social Venture Network/Social Venture Survey	Business	For-profit social ventures less than 30 years old	212	Not yet available	2003	N	N	Y, ZIP code	N	Y	Y, revenues and profits	Y	Y	N	N	N	N	Y	Y, depending on costs and benefits	Y
Kauffman Foundation Panel Study of Entrepreneurial Dynamics (PSED)	Individual	Nascent Entrepreneurs	830 (there was some attrition in follow-up surveys)	Free online	1998, Four years of data collected over five years	Y	Y	Y	N	Y	Y, Y	Y, Y	Y, Y	N, N	N, N	N, N	N, N	N, N	Y, Y	N, N
Social Compact Neighbor Market DrillDown	Individual	Low-income neighborhoods	101 low-income neighborhoods over 6 years	Free online (in PDF report format, some information disaggregated by geographic group level) more detailed information available by request	Snapshot years of data profile but can be updated annually.	Y	Y, where possible	Y, down to neighborhood level analysis	Y	Y, not consistent	Y, in terms of buying power	N	N	N	N	N	N	N	Y, not all	Y

Exhibit A. Matrix of Emerging Domestic Markets Databases

Data source	Database overview				Characteristics of individual				Characteristics of business ²				Characteristics of financing				Performance of financing		Interest in data consortium ³	
	Survey unit	Survey unit details	Number of units surveyed	Availability	Earliest year and frequency	Race / ethnicity	Gender	Geographic location	LMI location	Age of business	Number of employees	Financial information	Type of financing	Financing amount	Cost of financing	Other terms of financing	Rate of return	Rate of default	Social return	Willing to share data
Calvert Foundation Profiles Database	Lender	Community investment Organization	Approximately 71 (investing in small businesses and affordable housing)	Some data available free online; more robust data for sale	2000, Annual	Y, staff & board of the fund plus percent of fund's assets underlying investments	Y, staff & board of the fund plus percent of fund's assets underlying investments	Y, address and targeted region	N	Y	Y	Y, debt	Y	N	N	Y, performance of fund and investment in fund	Y, performance of fund and investment in fund	Y, not collected across all funds	Y, with some doubts	Y, if it fits their needs
CDFI Data Project (CDFI) by the Ford Foundation and MacArthur Foundation	Lender	Community Development Finance Institutions (CDFIs)	Approximately 500	750; select aggregate data available online	2001, Annual	Y	Y	Y	N, year financing started	Y	Y	Y	Y	N, occasionally	Y	Y	Y	Y, jobs created, housing units created, etc.	Y	Y
RISE Double Bottom Line Investor Survey	Transaction	Private equity funds	59	Not yet available	2002	Y	Y	Y, ZIP code	N	Y	Y, managed capital	Y	Y, managed capital	Y, target IRR and return hurdle	Y, average deal size	Y, IRR	Y	Y	Y, depending on costs and benefits	Y
Trade Associations																				
Community Development Venture Capital Alliance (CDVCA) Transaction Database	Business	Transaction-level data from CDVC fund investments	696	Summary information in CDVC report on the industry	2005, Annual	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y, in aggregate	Y
National Federation of Independent Business (NFIB) and Gallup Organization Small Business Poll	Business	Nationally representative sample of small firms	Varies	Information available on website; publications; data sets available for purchase	8 reports annually since 2001	N	Y	Y	Y, extend varies with survey	Y, extend varies with survey	Y, extend varies with survey	Y, extend varies with survey	Y, extend varies with survey	Y, extend varies with survey	Y, extend varies with survey	N	N	N	Y, available for purchase	Y
National Federation of Independent Business (NFIB) Economic Trends (SEET)	Business	Members of NFIB organization	600,000	Private	1973, Quarterly and monthly	N	N	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	Y, available on request	Y
Community Development Venture Capital Alliance (CDVCA) ROI Project	Lender	CDVC Funds	20	To be shared with CDVC industry	2004-05, One time study	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y, in aggregate	Y
National Association of Investment Companies (NAIC) and Wayne State University Venture Capital Fund Database	Lender	Venture capital investments monthly-solicited investments	24	Available for purchase; more robust data for research purposes	2000 & 2003	Y	Y	Y, ZIP code	N	Y	Y	Y	Y	Y	N	N	N	N	Y	Y

Exhibit A. Matrix of Emerging Domestic Markets Databases

Data source	Database overview				Characteristics of individual ¹				Characteristics of business ²				Characteristics of financing				Performance of financing		Interest in data consortium ³	
	Survey unit	Survey unit details	Number of units surveyed	Availability	Earliest year and frequency	Race / ethnicity	Gender	Geographic location	LMI location	Age of business	Number of employees	Financial information	Type of financing	Financing amount	Cost of financing	Other terms of financing	Rate of return	Rate of default	Social return	Willing to share data
Neighborhood Funders Group PRI Makers Network	Lender	Project-Related Investment (PRI)	Approximately 200	Available to PRI Makers members only (PRI Funders and grantmakers only)	Spring 2006	N	N	Y, state	Y	N	N	Y	Y	Y	Y, collateral, guarantee, etc.	Y	N	Y, wide variety of reported outcomes	Y, limited	Y
Opportunity Finance Network CDFI Asset Based Lending Rating System (CAFRS)	Lender	Community Development Institutions (CDFIs)	Currently 12; plans for more than 30 by YE 2006	\$15,000 for full subscription, \$2,500 per CDFI	2004, Annual (for cases by CDFI)	Y, for some	Y, for some	Y, for some	Y	Y, for some	Y	Y	Y, described in analysis	Y, described in analysis	Y, described in analysis	Y	Y	Y, jobs created, housing units created, etc. provided community conditions	Y, available for purchase	Y

¹ This table is organized by Type of institution, which refers to the data source, not the survey unit.

² When the survey unit is "transaction," data are available at the transaction level. When the survey unit is not transaction, data may be aggregated (e.g., percent of portfolio that is a certain race/ethnicity) or at the individual level.

³ Providing this information in no way obligates the organization to participate in later phases of the consortium or to share data.

⁴ ICIC maintains several emerging domestic markets databases.