

Texas Federal Statistical Research Data Center: Opportunities for Research Using Restricted Data

Proposal Development Workshop

University of Houston

May 9, 2017

Mark Fossett

**Executive Director, Texas Research Data Center
College of Liberal Arts, Texas A&M University
College Station, Texas**

Overview of Presentation

Welcome everyone! Thank you for attending.

1. Overview of FSRDCS and TXRDC Resources

2. Some Example Projects and Scouting RDC Projects and Data

3. Preparing Proposals for Projects Using Restricted Data

Starting a Conversation

We appreciate the opportunity to share information about the resources and services the Texas Research Data Center (TXRDC) can offer to researchers.

Please consider this the beginning of an extended conversation.

We cannot anticipate and answer every question a prospective researcher might have today. But we hope to give folks a better sense of the research opportunities in research data centers (RDCs) and of the nature of RDC projects.

So, let's start the conversation ...

Early Milestones in TXRDC History

The Texas Research Data Center (TXRDC) came into official existence in August 2011 with US Census Bureau approval and National Science Foundation funding support.

The next twelve months were devoted to implementing the TAMU-Census contract and building the Texas RDC's secure computing lab.

In September 2012 our secure lab went "live" and joined the RDC network managed by the Center for Economic Studies, U.S. Census.

Our official grand opening was held October 2012.

In spring 2015 we changed from being a "Census RDC" to being a "Federal Statistical RDC." The change signaled expanded data holdings and capabilities and a broader mission.

Grand Opening Ribbon Cutting – October 19, 2012



José Luis Bermúdez, Dean of Liberal Arts; Mark Fossett, Director TXRDC; Thomas L. Messenbourg - Acting Director of the US Census Bureau; R. Bowen Loftin, President Texas A&M University; Theresa W. Fossum, Interim Vice President for Research, Texas A&M University

Recent Milestones in TXRDC History

2015 – RDCs officially change from being “Census Research Data Centers” to being “Federal Statistical Research Data Centers”

Spring 2016 – UT-Austin and Texas A&M win approval to establish an RDC “branch” on the UT-Austin Campus

Fall 2016 – TAMU hosted the Annual FSRDC Business meeting and Annual FSRDC Research Conference

Spring 2017 – TXRDC project activity equals or exceeds that seen in the oldest and most established RDCs

Summer 2017 – The UT-Austin Branch of TXRDC will come online

Winter 2017/2018 – The University of Houston will join TXRDC Consortium

“Core” and “Branch” Distinctions in Brief

Core and Branch RDC locations are functionally near-identical for researchers with active projects. There are a few differences.

Core locations:

- **Are approved by a more demanding and competitive joint review by Census and by NSF (only 1-2 locations per year)**
- **Typically have a more senior RDC Administrator**
- **Represent the core-and-branch configuration to Census**

Branch locations:

- **Are approved by review by Census (no NSF involvement)**
- **Must be affiliated with a Core location which coordinates with the Branch to submit the proposal to Census**
- **Coordinates activities and policies with the Core location**
- **Can be more or less autonomous depending**

Current TXRDC Activities

We focus on (a) promoting awareness of the Texas RDC and its resources and (b) assisting researchers in moving forward with research projects that draw on restricted-access data.

We advance these goals with the following activities:

- **Hosting Proposal Development Workshops (such as this)**
- **Consulting with individual researchers and research teams**
- **Conducting a Seed Grant program to support proposal efforts**
- **Hosting speakers presenting research from RDC projects**
- **Hosting workshops on data and methods of interest to RDC researchers**
- **Making promotional presentations around TAMU and Texas**

RDC Research Conference 2016

Each year since 2002 RDCs have been hosting a research conference that brings together RDC researchers and administrators from around the country.

The TXRDC Consortium hosted the event in 2016

- **The main events were held September 14-15, 2016**
- **Related workshops continued September 16-17, 2016**
- **Approximately 60 Census Bureau Leaders, RDC Directors, and RDC Administrators from 24 RDCs**
- **Approximately 48 research presentations**
- **Keynote presentation by Steve Ruggles on the history of big data (see TXRDC website for link)**
- **Approximately 250 attendees at all events combined**

The 2017 RDC Research Conference is hosted by UC- Berkeley

Overview of RDC Network Today

24 RDCs & Branches are in operation; 6 came online this year.



The Texas RDC is a part of a growing national network of RDCs located at elite research institutions

Acknowledging Support – I

The Texas RDC is supported by a consortium of institutions led by Texas A&M University.

The “founding” institutions of the TXRDC Consortium:

- **Texas A&M University (lead institution)**
- **The Texas A&M University System**
- **The University of Texas at Austin**
- **Baylor University**

More recent additions:

- **Rice University**
- **The University of Texas at San Antonio**

We welcome other institutions to join our consortium.

- **The University of Houston ??**

Acknowledging Support – II

RDCs are costly to implement and operate.

The consortium member institutions provide commitments to cover these costs and establish this important resource for research.

The consortium model:

- **Provides long-term stability needed for planning and conducting RDC projects.**
- **Maximizes researcher access to this valuable resource.**
- **Produces more cost-efficient research (i.e., more research projects and products per total investment).**
- **This approach casts RDCs as “general research infrastructure” similar to a research libraries.**

Acknowledging Support – III

Texas RDC Consortium Institutions



TEXAS A&M
UNIVERSITY®



THE TEXAS A&M
UNIVERSITY SYSTEM



BAYLOR
UNIVERSITY



THE UNIVERSITY OF
TEXAS
— AT AUSTIN —



RICE®

UTSA®

Administrative Home



LIBERAL ARTS
TEXAS A&M UNIVERSITY

Growth of the National FSRDC Network

Established Before 2010

Chicago-FRB	CES-Census HQ	Berkeley/Stanford
NY-Baruch/CUNY	Boston/NBER	Duke/Res. Triangle
NY-Cornell	UCLA	Michigan

Established 2010-2014

Atlanta-FRB	Texas	Minnesota
Washington	Pennsylvania State	

Recent 2015-2016 and Under Construction (2017)

Nebraska	Kansas City-FRB	Yale (NY)
USC (UCLA)	UC-Irvine (UCLA)	Missouri (KC)
Maryland (CES)	Wisconsin (MN)	
Colorado	Kentucky	Illinois (CFRB)
Georgetown (CES)	UT-Austin (TXRDC)	Pennsylvania (PA)

TXRDC was number 12 in 2010. There are 24 in 2016; 30 in 2017.

What is a Federal Statistical Research Data Center (RDC)?

Federal Statistical Research Data Centers (RDCs) are unique research facilities based on contractual relationships – Joint Statistical Partnerships (JSP) – between the Census Bureau and leading research institutions around the nation.

In essence, the RDC is a census “outpost” at hosted at TAMU.

The RDC has a secure computing lab that is linked with the U.S. Census Bureau’s internal computing network in Washington DC.

The RDC network is managed by the Center for Economic Studies (CES) at Census Headquarters which hosts restricted-access data sets from federal statistical agencies on their servers.

Local RDCs are managed by an on-site Census Bureau employee – the RDC Administrator – who serves as a liaison with the research community. Dr. Bethany DeSalvo is the “RDC Admin” at TXRDC.

The Claim to Fame for RDCs

The contractual relationship (JSP) with the Census Bureau makes it possible for “qualified” researchers who have approved projects to access non-public data from within the RDC facility.

Based on this, RDCs make it possible for researchers to use restricted-access data maintained by Census and other agencies (e.g., NCHS) in the federal statistical system.

Note that researcher access requires meeting specific conditions.

- **Researchers must be “qualified” based on professional credentials and, importantly, having “special sworn status” which makes them an (unpaid) Census Bureau researcher.**
- **Research projects must undergo under an external review.**
- **Projects must provide benefits to Census or to other federal agencies in keeping with regulations regarding data access.**
- **Data access, data analysis, and disclosure of results must follow strict protocols to protect confidentiality of the data.**

Why are RDCs Needed?

Federal laws and regulations protect confidential data in the federal statistical system

- **It is illegal to disclose confidential federal data**
- **The restrictions reflect a variety of legal, ethical, & practical concerns**

Federal employees can access restricted data but only under specific guidelines and only as needed to fulfill their institutional mission

RDCs extend the possibility of accessing restricted data to researchers via the mechanism of Special Sworn Status (SSS)

- **In effect, the researcher becomes as an unpaid Census Bureau employee whose efforts benefit the Census Bureau**
- **SSS is conferred under a formal, mandatory review process**

What does the Texas RDC Offer?

Primary Benefits and Resources

- **ACCESS** – Texas RDC provides qualifying researchers access to restricted data in the federal statistical system.
- **SECURE LAB** – Access is only possible on site from within the physical facility of the secure computing lab housed at TXRDC.
- **PROPOSAL DEVELOPMENT** – TXRDC provides proposal development assistance, seed grant funding, research analyst support, and other relevant assistance
- **A SENIOR RDC ADMINISTRATOR** – TXRDC has a senior RDC Admin, a big advantage for scouting and developing proposals

Other nice things

- Easy driving distance – San Antonio/Austin/Houston/Waco
- Easy airport access
- Relatively inexpensive accommodations for extended stays
- Workspace for team meetings and out-of-town researchers

Recent Developments – Federal Statistical RDC

The role of RDCs is evolving.

Many federal agencies permit researchers to use restricted data sets “on-site” (e.g., BLS, NCES, NIJ). It is inconvenient & expensive. It precludes options of linking with other restricted data sets.

Many of them are exploring the option of Census hosting their restricted data sets on the RDC network to facilitate research. (The Bureau of Labor Statistics is a recent example.)

This is the wave of the future.

Federal Statistical Research Data Centers (FS RDC)

The new name reflects these developments and signals future directions in access to restricted federal data.

The national FSRDC network will make it possible for investments in federal statistical and administrative data sets to benefit society through research conducted in RDCs.

What Research Communities Do RDCs Serve Best?

RDCs Serve Well ...

- **Basic Science Research**
 - discipline-based substantive research
 - statistical and methodological research
- **Planning and Policy Science Research**
 - research evaluating program effectiveness
 - research assessing program impacts
- **Research programs with long time horizons**

RDCs Serve Less Well ...

- **Research with short time horizons**
- **Research producing detailed lists, descriptive statistics, special tabulations, and maps**
- **Projects to produce data products for public distribution**

Advantages of RDC Access to Researchers

Microdata not available publicly

Key variables not available in public versions of data sets (e.g., low level geography, sensitive health information, etc.)

Large data sets – full population counts and/or larger samples

Original responses/items prior to “editing” and processing (e.g., detailed race answers, income is not top-coded, etc.)

Finer codings of variables (e.g., 5 digit industry codes)

Ability to link with external data (e.g., via geocodes, establishment ID, etc.)

Ability to merge multiple *internal restricted* data sets via non-public link keys

RDC-Based Research is Common in Many Fields

Leading researchers in many fields conduct research in RDCs

- **Business, Trade, Finance, and Management**
- **Crime and Crime Victimization**
- **Demography and Population Studies**
- **Economics, Labor Markets, Entrepreneurship, Employment and Industry**
- **Education and Education Policy**
- **Hazard Reduction, Environmental Impact Assessment, Pollution Abatement**
- **Health and Well-Being, Health Insurance, and Health Policy**
- **Housing, Housing Markets, and Residential Patterns**
- **Poverty and Social Welfare Policy**
- **Transportation Analysis and Planning**
- **Urban and Regional Economics and Planning**

Available Data Sets Number in the 100's

1. **AHRQX MEPS Extract**
2. **American Community Survey**
3. **American Housing Survey**
4. **Annual Capital Expenditures Survey**
5. **Annual Retail Trade Survey**
6. **Annual Survey of Manufactures**
7. **Auxiliary Establishment – ES9200**
8. **Business Expenditures Survey**
9. **Business Register Bridge**
10. **Census of Construction Industries**
11. **Census of Finance, Insurance, and Real Estate**
12. **Census of Manufactures**
13. **Census of Mining**
14. **Census of Retail Trade**
15. **Census of Services**
16. **Census of Transportation, Communications, and Utilities**
17. **Census of Wholesale Trade**
18. **Commodity Flow Survey**
19. **Compustat-SSEL Bridge**
20. **Current Industrial Reports**

Examples of Available Data Sets – Continued

- 21. Current Population Survey March Supplement**
- 22. Decennial Census Long Form Sample**
- 23. Decennial Employer-Employee Database**
- 24. Economic Census of Puerto Rico**
- 25. Employer Characteristics File**
- 26. Employment History Files**
- 27. Enterprise Summary Report – ES9100 (large company)**
- 28. Exporter Database**
- 29. Foreign Trade Data – Export**
- 30. Foreign Trade Data – Import**
- 31. Form 5500 Bridge File**
- 32. Geocoded Address List**
- 33. Individual Characteristics File**
- 34. Integrated Longitudinal Business Database**
- 35. Longitudinal Business Database**
- 36. Manufacturing Energy Consumption Survey**
- 37. Medical Expenditure Panel Survey (MEPS) – Insurance Component**
- 38. National Center for Health Statistics Data Extract**
- 39. National Employer Survey**
- 40. National Longitudinal Survey**

Examples of Available Data Sets – Continued

41. Ownership Change Database
42. Quarterly Financial Report
43. Quarterly Survey of Plant Capacity Utilization
44. Quarterly Workforce Indicators
45. Services Annual Survey
46. Standard Statistical Establishment Listing – non Name and Address File
46. Standard Statistical Establishment Listing – Name & Address File
47. Survey of Business Owners
48. Survey of Income and Program Participation (SIPP) Panels
49. Survey of Income and Program Participation – Longitudinal
50. Survey of Industrial Research and Development
51. Survey of Manufacturing Technology
52. Survey of Plant Capacity Utilization
53. Survey of Pollution Abatement Costs and Expenditures
54. Unit-to-Worker

...

... and on and on and on ...

OK, you get the point

New Data Sets Become Available Regularly

New Data Sets (and versions of data sets) are added regularly

See the CES for detailed compilations

- **Link to CES [Census for Economic Studies \(home page\)](#)**
- **Link to CES [Notes on Restricted Data](#)**

Consult with RDC Administrator Bethany DeSalvo to inquire about data sets

- **Some data sets are available but are not publicized**
- **Important information about data sets often is not public**

Health Data Sets?

Health-related research in RDCs has recently grown dramatically. Currently about 50% of RDC projects are health related, up from under 10% five years ago.

The website for the National Center for Health Statistics (NCHS) provides descriptions of the restricted variables that are available in RDCs by NCHS data set.

[NCHS Restricted Data Sets & Information](#)

The website for the US Department of Health and Human Services Agency for Healthcare Research and Quality (AHRQ) RDC provides a description of the restricted variables that are available in RDCs by AHRQ data set at the following link.

[AHRQ Restricted Data Sets & Information](#)

Trends in RDC Scope and Mission

The Number of RDCs is Growing

- **24 RDCs are now operating**
- **30 RDCs will be operating by 2017 (more soon to follow)**
- **As of January 2017 about 60% of Research I universities are affiliated with RDCs**

Mission and Data Sets are Expanding

- **More statistical agencies are partnering with Census to make restricted data available in RDCs**
- **The number of available data sets is growing**

The Role of RDCs is on Track to Increase in Importance

- **Public microdata are going to be increasingly limited**
- **RDCs will increasingly be the only option for detailed microdata and data linking possibilities**
- **RDCs are envisioned as crucial for Evidence Based Policy Making” (per “Ryan Commission” and related initiatives)**

Trends in RDC Activity

Researcher Usage is Increasing

- **Consortium funding models makes RDCs more feasible and cost effective; this has increased access and usage**
- **The review process is improving**
- **In many fields, access to restricted data has become a crucial resource for research (journal reviews increasingly request RDC analysis)**

Patterns and Trends

- **Traditionally, most research was in economics/business**
- **In recent years, health research has risen dramatically**
- **Demographic research and other research has held steady**

University Benefits of TXRDC

TXRDC Benefits to Universities and Research Institutions

- **Enables researchers to undertake advanced research projects that would not be feasible otherwise**
- **Expands basic knowledge via ability to analyze superior data**
- **Enhances a university's ability to attract and retain top researchers, scholars, and doctoral students**
- **Serves as a "hub" that attracts leading researchers and scholars to visit and collaborate**
- **Fosters a vibrant Texas region research community**

Benefits to Federal Statistical Agencies

RDC Benefits to the Census Bureau

- Ability to engage leading researchers and increase the pool of expertise the Census can draw on
- Identifying methodological issues and potential solutions
- Deriving new estimates from existing data
- Enhance utility of data sets
- Improving documentation

The Issue of Benefits is IMPORTANT for Many Projects

- The National RDC program is justified based on Benefits it provides to the Census Bureau
- The RDC program is NOT justified based on Benefits to the research community

Accordingly, Projects are Justified in Terms of Census Benefit

Texas RDC Director and RDC Administrator

Executive Director – Mark Fossett

**Oversees contracts with U.S. Census Bureau;
coordinates with the RDC Admin to oversee the facility**

**Promote awareness, share information, and encourage and assist
with proposal development**

**Oversee seed grant programs, establish and interpret policies,
and nurture/grow the TXRDC consortium**

Approves lab access for projects

RDC Administrator – Bethany DeSalvo

Manages the secure computing lab

Assists with proposal development and project planning

Serves as direct liaison to Census Bureau

Oversees disclosure review

UT-Austin Branch Director and RDC Administrator

Formally, the UT-Austin Branch does not yet exist and a local (Austin-based) RDC Administrator is not in place.

Future Arrangements are TBD and Will Evolve

Presumptions are:

- **The local RDC Director will be a point of contact for information, guidance, and local policies**
- **The TXRDC Director will often be involved for setting and interpreting broader policies**
- **The local RDC Administrator will be a facilitator for researchers**
- **The TXRDC RDC Administrator will continue to be centrally involved in proposal development**

Prediction – this is going to work very well!

Notes on Access and Usage for Approved Projects

Costs for Researchers at Consortium Institutions

- **Access is at no cost or at low cost compared to charges for non-consortium users**
- **Scheduling will be prioritized, if needed, on several criteria. (Our lab is large so this will not be an issue in the near term.)**

Costs for Non-Consortium Researchers

- **Researcher access will be on a per project cost schedule in keeping with practices in the national RDC network**
- **Usual terms are \$15,000-\$20,000 per project per year**
- **Collaborating with researchers at consortium institutions can bring more favorable terms.**

Consortium membership is the most cost effective way for institutions to give researchers stable, predictable access

Important Practical Issues to Consider

Obtaining Census Project Approval

- **Census Bureau approval of projects is required.**
- **The main criteria for Census Bureau approval is not scientific merit (this is mostly presumed). It is “benefits to Census”.**
- **Special Sworn Status requirement for PI and all “in-lab” researchers (3-year residency requirement for non-citizens)**

The RDC Computing Environment – Software and Hardware

- **Computing is conducted in the secure lab**
- **The data and software are hosted on Census Bureau servers**
- **Census Bureau workstation hardware/software limitations**

Getting Research Products Out Involves a Review Process

- **Project results are reviewed on confidentiality criteria; NOTHING leaves the secure lab without approval**
- **It is crucial to plan ahead for getting project approval and for getting project output (e.g., regression results) out of the lab**

Stage of Career and Planning RDC Research

Researchers at different stages of their career should think about how RDC projects fit in their plans and research agenda.

The key issues are the trade-offs between

Pro - the potential for performing cutting edge research

Con – the time investments associated with first projects and the extended time line to completed research projects

Researchers at Different Stages

Doctoral Students – must be especially careful & thoughtful

Assistant Professors – must use informed judgment

Associate and Full Professors – time to learn “new tricks” and bring junior collaborators along with you

I Want In! What Do I Do? When?

Contact the TXRDC Director and the RDC Administrator to learn about next steps

Attend TXRDC Workshops and Speaker presentations

Apply to our Proposal Development (Seed) Grant program

Review materials at our website, the CES website, the NCHS website, and other sites listed at our website.

When? *Start as Soon As Possible*

Census Bureau approval is needed for RCD access. The process is not fast, so do not delay unnecessarily.

Also, the RDC needs to demonstrate its value to justify the consortium commitments.

If the RDC is a potential resource for you, USE IT OR LOSE IT.

How Will the TXRDC Help?

Researching information relevant to researcher interests

- **Data availability; available software**
- **Project budgets and timeline**
- **Contact information**

Guidance on Proposal Development Process

- **Navigating the Census Project Approval Process**

Seed Grant Program for Proposal Development

Assistance and Training for using the Secure Lab

Scheduling On-Site Project and Workgroup Space

The Internal (TXRDC) and External (Agency) Review

Informal Development Review at the TXRDC

Researchers consult with TXRDC staff to learn about the review process and the relevant criteria to receive preliminary feedback on proposals.

Official Proposal Review by Census or NCHS

Typically this takes about 90 days and may involve revisions.

Other Agency Review as Appropriate

- If non-census data are involved, additional federal agencies will be involved as appropriate.**
- Relevant for IRS, Social Security, and other administrative data**
- In some cases this may involve additional fees and protocols**

A Special Category of RDC Project – Internal Projects

RDCs create the opportunity for direct collaborations with researchers on “internal projects” underway at federal statistical agencies.

These are common and can be very important.

They are hard to describe; each one is unique.

- **Collaboration is by invitation often emerging out of professional relationships with agency researchers**
- **Proposal and review process may be highly streamlined**
- **Publication and analysis options vary widely**

The common factor is that the RDC makes it easier for local researchers to collaborate with researchers in federal statistical agencies.

Coordinating with Proposals to Funding Agencies

Proposals to external funding agencies can be built around active or proposed RDC projects

- **Crucial to understand how this can work so proposals can be planned and described realistically**
- **TXRDC Director and Administrator write support letters**

External funding agencies reviews focus primarily on the project's "scientific merit" (and fit with program goals)

RDC agency reviews focus primarily on the project's "viability" and fit with agency needs (benefits to the agency)

Tips for a Good RDC Experience

Met with Mark and Bethany ASAP!

Have Realistic Expectations, Plan Ahead, and Plan Carefully

- **Start the process early – do not underestimate the time line**
- **Consult early and closely with Bethany and Mark to determine project feasibility**
- **Ask about seed grants to support proposal development**

Think Strategically About RDC Projects

When appropriate, write proposals geared to multiple papers over an extended time period (3-5 years)

Learn About the Research Environment

Attend RDC workshops; Read on-line documentation and advice; Consult with RDC researchers; Examine CES research annuals and Working Papers

Useful Sources of RDC Information

Information about Data, Data Sets, and New Developments

Center for Economic Studies (CES, US Census)

[Census Restricted Data Sets & Information](#)

National Center for Health Statistics (NCHS)

[NCHS Restricted Data Sets & Information](#)

Agency for Healthcare Research and Quality (AHRQ)

[AHRQ Restricted Data Sets & Information](#)

Developing Projects

TXRDC Director – Mark Fossett

(m-fossett@tamu.edu)

RDC Administrator – Bethany DeSalvo

(bethany.desalvo@census.gov)

Overview and Conclusions

The Texas RDC is an Important Resource and Opportunity

- **Available data will support a variety of research agendas**
- **The data in RDCs will be increasingly crucial to being at the forefront of research in many fields**
- **The resource can help attract and retain top researchers**

The Location and the Consortium Model are Major Benefits

- **Currently Texas researchers must travel long distances to use an RDC and incur substantial time and travel costs**
- **Consortium membership provides cost advantages; “Pay as you go” RDC access typically costs \$20K annually per project**

Benefits More than Just Senior Researchers

- **Assistant and associate-level researchers can advance careers; Graduate students, recent graduates, and postdocs can benefit**

TXRDC Contact Information

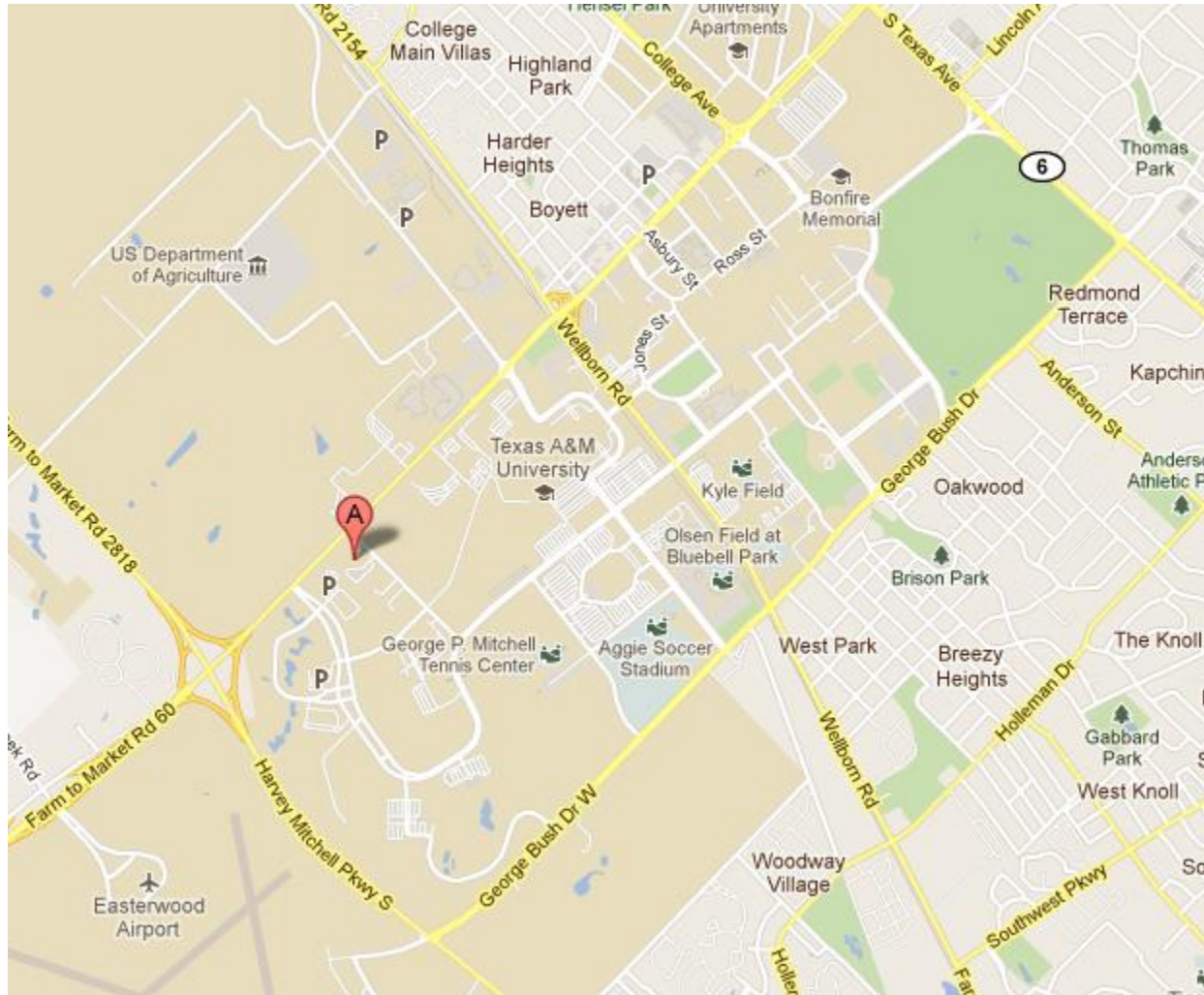
Mark Fossett, Director, Texas A&M University
m-fossett@tamu.edu

Bethany DeSalvo, RDC Administrator, U.S. Census Bureau
bethany.desalvo@census.gov

Texas Census Research Data Center
TAMU 2403
101 Donald L. Houston Building
200 Discovery Drive
College Station, Texas 77843-2406
TXRDC@tamu.edu
979-845-5618

TXRDC website: <http://txrdc.tamu.edu>

Texas RDC Location and Directions



Texas RDC – D. L. Houston Bldg, 200 Discovery Dr.



University of Texas – Austin Branch

Coming to a “40 Acres” near you

Location and hours of operation to be announced

Watch the TXRDC and PRC websites for updates