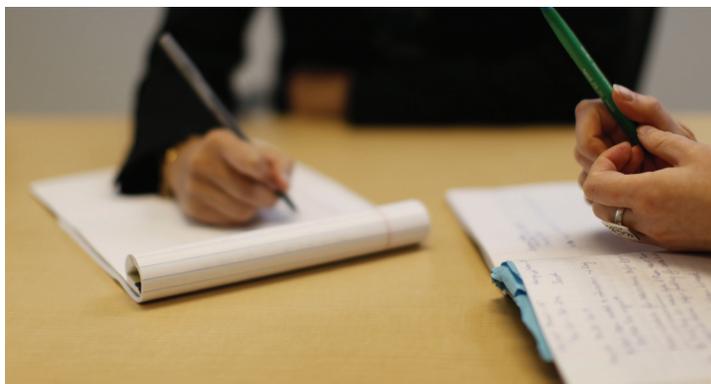


# NORC Data Enclave

The Key to Unlocking Data Securely



NORC data enclave

## SENIOR STAFF

### Chet Bowie

Senior Vice President and  
Director of Economics,  
Labor, and Population Studies  
Department

Interviewing methods, questionnaire  
design, survey automation, combining  
administrative data with survey data  
for policy-relevant research

### Timothy M. Mulcahy

Senior Research Scientist  
and Program Director,  
NORC Data Enclave

Data warehousing, data security,  
privacy and confidentiality protection,  
microdata dissemination

Founded in 1941, NORC at the University of Chicago is an independent social science research organization dedicated to serving the public interest.

## CHICAGO

Downtown Chicago  
55 East Monroe Street  
Chicago, IL 60603  
(312) 759-4000

University of Chicago  
1155 East 60th Street  
Chicago, IL 60637  
(773) 256-6000

## DC & MARYLAND

4350 East-West Highway  
Suite 800  
Bethesda, MD 20814  
(301) 634-9300

## Data Enclave

### CONTACT

Timothy M. Mulcahy  
Mulcahy-tim@norc.org  
(301) 634-9352

[www.norc.org/dataenclave](http://www.norc.org/dataenclave)

Launched in 2006, the NORC Data Enclave provides a confidential, protected environment within which authorized researchers can access sensitive microdata remotely. While public use data can be disseminated in a variety of ways, there is a more limited range of options for disseminating sensitive microdata that have not been fully de-identified for public use. Some data producers have sufficient economies of scale to develop advanced in-house solutions that serve the needs of external researchers, but most lack the resources to archive, curate, and disseminate the datasets they collect. The NORC Data Enclave provides our partner organizations a secure platform where they can both host and build a research community around their data.

The NORC Data Enclave currently serves more than 250 researchers and hosts confidential data for several Federal agencies and foundations.

### Major Capabilities

Our team values NORC's core mission of conducting research of the highest quality that serves the public interest. The NORC Data Enclave offers the full cycle of data services, ranging from study design and concept to data archiving and access. In addition to being experts in metadata documentation and data dissemination, we provide a comprehensive set of analytic services. Members of our team come from a wide range of backgrounds and are thoroughly trained in research methods, statistical analysis, and can leverage the full array of analytic capabilities of NORC.

In addition, the NORC Data Enclave engages the research community in developing a knowledge infrastructure around each dataset through its virtual collaboration, which enables geographically dispersed researchers to share information, replicate results, and provide feedback to the data producer.

### Areas of Expertise:

Privacy and confidentiality protection	Data warehousing
Disclosure analysis	Data security
Statistical analysis	Secure remote access technologies
De-identification & micro-aggregation	Metadata management
Visual and Spatial Analysis	Virtual collaboration
Data diagnostics and harmonization	Microdata dissemination



## Innovative Secure Access Platform

The NORC Data Enclave provides an ideal system for accessing confidential microdata. Employing security technologies used by the military and banking sectors, the Data Enclave is based around a Citrix infrastructure that renders users' computers as "remote terminals," providing them access to data, statistical software, and collaborative tools hosted on a central, secure NORC server.

Only authorized users will gain access to microdata through a customized virtual environment where information inflow and outflow are strictly controlled and monitored. The system is highly flexible as it allows the data custodian to restrict access to particular users and/or particular locations. It is secure because all data manipulations occur within NORC servers that follow strict security protocols.

## Data Security & Certification

The NORC Data Enclave is in its fifth year operating under a NIST-approved System Certification and Accreditation (C&A) package, including an approved IT Security Plan, Data Protection Plan, and a System Certification Test Plan, as outlined in DOC IT Security Program Policy, Section 6.5.2. The NORC Data Enclave's IT Security Plan is fully compliant with the Federal Information Security Management Act (FISMA), provisions of mandatory Federal Information Processing Standards (FIPS), and meets all of NIST's IT, data, system, and physical security requirements.

Per the Federal Information Security Management Act and provisions of mandatory Federal Information Processing Standards (FIPS) 199 and NIST Security Plan 800-60, NORC's Data Enclave system impact levels have been determined as follows:

<b>NIST Sensitivity Element</b>	Confidentiality
<b>NIST Impact Rating</b>	Moderate
<b>Short Description of Basis for Rating</b>	The Data Enclave contains information of a proprietary nature. If this proprietary information were disclosed, it could result in serious loss of customer trust for the Federal Statistical Agency.

### Our Sponsors

U.S. Department of Agriculture (USDA)

- Economic Research Service (ERS)
- National Agricultural Statistics Service (NASS)

National Science Foundation (NSF)

National Institute of Standards and Technology (NIST)

Center for Medicare and Medicaid Services (CMS)

Annie E. Casey Foundation (AECF)

Ewing Marion Kauffman Foundation

National Bureau of Economic Research (NBER)

Financial Crisis Inquiry Commission (FCIC)

Oregon State University

## NORC DATA ENCLAVE NEWSLETTER

DECEMBER 2010

### INSIDE THIS ISSUE:

**2ND ANNUAL CONFERENCE ON MICRODATA ACCESS** 1

**ENCLAVE TEAM SUPPORTS THE FINANCIAL CRISIS INQUIRY COMMISSION** 2

**CENTERS FOR MEDICARE & MEDICAID SERVICES DEPOSIT DATA IN ENCLAVE** 2

**2ND COOPERATIVE AGREEMENT WITH THE USDA** 2

**THE NATIONAL SCIENCE FOUNDATION JOINS THE ENCLAVE** 3

**NORC TEAM IMPLEMENTS BUSINESS INNOVATION SURVEY** 3

**NEW GRANT FROM THE KAUFFMAN FOUNDATION TO BUILD AND SUSTAIN KFS ENCLAVE COMMUNITY** 4

**ENCLAVE CELEBRATES 4TH YEAR ANNIVERSARY WITH FOUNDING SPONSOR NIST** 4

**AECF ADDS NEW RESEARCHERS AND FUNCTIONALITY TO THE ENCLAVE** 4

**NEW WEBSITE COMING AND NEW MEMBERS** 5

**DATA ENCLAVE OFFERS NEW SERVICES AND CAPABILITIES** 5

**TO LEARN MORE ABOUT NEW ANALYTIC SERVICES PROVIDED THROUGH THE ENCLAVE, GO TO PAGE 5**

### 2ND ANNUAL CONFERENCE ON MICRODATA ACCESS: FEBRUARY 10, 2011



Ballroom at the National Press Club

On February 10, 2011, NORC at the University of Chicago and the National Institute of Standards and Technology will host the 2nd Annual Conference on Microdata Access at the National Press Club in Washington D.C. Senior leadership from Federal statistical agencies, foundations, and other stakeholders in the research community will share their insights on best practices and strategies for data sharing. Discussants also will focus on their motivation to do so as well as concerns about extending microdata access. Other sponsors include the U.S. Department of Agriculture, National Science Foundation, Kauffman Foundation, and the Annie E. Casey Foundation.

As part of the current Administration's transparency initiatives, the Office of Management and Budget (OMB) issued a memorandum on November 3, 2010 that encourages responsible data sharing. This guidance memo, which lays out the prospects and challenges of providing microdata access in the 21<sup>st</sup> century, will be discussed during the first of two panel discussions during the conference. The first panel will include leaders from various statistical agencies having a frank conversation about the effects of these changes on their respective agencies. The second panel session will focus more on the goals, objectives, and implications of extending confidential microdata access to wider audiences (e.g., Ph.D. students), various access modalities, and how emerging technologies facilitate secure access to sensitive microdata. The conference will conclude with a discussion on how secure remote access to microdata can influence responsible data sharing for the overarching social science research community.

#### NEW OMB DIRECTIVE ON DATA SHARING: M-11-02 SHARING DATA WHILE PROTECTING PRIVACY

A recent memorandum from the Office of Management and Budget (OMB) further reinforces the central theme of the 2nd Annual Conference on Microdata Access. OMB directive M-11-02, Sharing Data While Protecting Privacy, encourages Federal agencies to identify "new approaches to share high-value data for purposes of supporting important Administration initiatives, informing public policy decisions, and improving program implementation." The directive recognizes the benefits of responsible data sharing for informing public policy decisions, and encourages Federal agencies to pursue initiatives to share high-value data.



Click [here](#) to download the memorandum.

For more information and to register for the event, please contact Data Enclave Program Director, Tim Mulcahy [Mulcahy-tim@norc.org](mailto:Mulcahy-tim@norc.org).

## DATA ENCLAVE SUPPORTS FINANCIAL CRISIS INQUIRY COMMISSION

The Data Enclave team recently assisted the [Financial Crisis Inquiry Commission](#) (FCIC) in administering a market risk survey of hedge funds, as well as aggregating and de-identifying the microdata in order to protect the confidentiality of survey responses. NORC also prepared the final dataset, thus highlighting our capacity to conduct data diagnostics, data imputation, data quality control, and disclosure analysis.

As part of the Fraud Enforcement and Recovery Act of 2009, the congressionally appointed FCIC was established to examine the causes, domestic and global, of the current financial and economic crisis in the United States and determine the levels of interconnectedness and effect of contagion in the financial services industry.

The 10-member bipartisan Commission was entrusted to conduct a comprehensive examination of issues related to the financial crisis and report its findings to Congress, the President, and the public. Data Enclave staff collaborated closely with the Commission in 2010 to support this important effort.



For more information, please contact [mulcahy-tim@norc.org](mailto:mulcahy-tim@norc.org)

## CMS DEPOSITS DATA IN ENCLAVE

NORC's Public Health Department recently entered into a contract with the [Center for Medicare and Medicaid Services](#) (CMS), in partnership with IMPAQ International, to examine the potential for releasing analytically useful de-identified public use files from Medicare and Medicaid claims data. These data include individual beneficiary-specific information, inpatient and out-patient records, part D drugs data, and records from other treatment institutions. All confidential data are now housed in NORC's secure Data Enclave.

The NORC team, comprised of Avi Singh, Fritz Scheuren, Beth Hair, Amanda Yu, Al Crego, Peter Kwok, and Josh Borton, is responsible for creating enhanced public use files that link patients' information to their institutional records. NORC is also responsible for conducting disclosure analysis of the public use files to make certain that no individual can be re-identified from the resulting dataset. The Data Enclave team provides technical support to the project in terms of housing, managing, preparing, and analyzing the data files. The data include millions of records in each 5% summary files, thus requiring significant processing power and computational speed. To meet this need, NORC has invested significant resources to enhance the capacity of the current infrastructure.



For more information, please contact [kuk-kenneth@norc.org](mailto:kuk-kenneth@norc.org)

## USDA AWARDS 2ND COOPERATIVE AGREEMENT TO THE DATA ENCLAVE

The NORC Data Enclave recently entered into a 2nd cooperative agreement with the [United States Department of Agriculture](#). The agreement titled *ARMS Dissemination using the NORC Data Enclave* aims to build on NORC's existing work with USDA, i.e., providing secure desktop access to the ARMS data for academics and senior analysts working on approved projects. The project also will (1) examine the feasibility of extending microdata access, in a tightly controlled pilot test, to Ph.D. students working under Principal Investigators with formally USDA-approved MOUs and Project Agreements; (2) evaluate collaboration incentives, and (3) determine if enclave scholarships encourage wider research participation.

USDA recently had a call for proposals as part of its *Agricultural Resource Management Survey Scholarship Program*, which provides one-year free access (value of \$5,200) to the Data Enclave for scholars from accredited U.S. institutions of higher education or non-profit organizations. We received a number of excellent proposals from this call and will be making final decisions on awardees this December.

In 2010, the USDA enclave research group added 14 new users accessing the ARMS data through the Enclave. Additionally, all USDA researchers now have access to the 2009 Phase III ARMS data via customized secure thin client machines.

For more information, please contact [sapkota-shaswat@norc.org](mailto:sapkota-shaswat@norc.org)



**THE NATIONAL SCIENCE FOUNDATION JOINS THE DATA ENCLAVE**



**NORC ANTICIPATES ADDING TWO NEW DATASETS TO THE ENCLAVE: THE SURVEY OF DOCTORATE RECIPIENTS AND THE SURVEY OF EARNED DOCTORATES**

**Directorate for Social, Behavioral & Economic Sciences**

**Division of Science Resources Statistics**

The National Science Foundation awarded a contract to NORC to create a Secure Data Access Facility (SDAF). The objective of this statement of work is “to house the Survey of Earned Doctorates (SED) and Survey of Doctorate Recipients (SDR) data and metadata and provide secure remote access to these data at the SDAF as well as technical support to approved researchers [as defined by the National Science Foundation (NSF)] as well as Division of Science Resources Statistics (SRS) staff at NSF.”

The Survey of Doctorate Recipients (SDR) gathers information from individuals who have obtained a doctoral degree in a science, engineering or health field (SEH). The SDR is conducted every two years and is a longitudinal survey that follows recipients of research doctorates from U.S. institutions until age 76. This group is of special interest to many decision makers, because it represents some of the most highly educated individuals in the U.S. workforce. The SDR results are used by employers in education, industry, and government to understand and predict trends in employment in SEH fields. The results are also used to evaluate the effectiveness of equal opportunity efforts.



The Survey of Earned Doctorates (SED) began in 1957–58 to collect data continuously on the number and characteristics of individuals receiving research doctoral degrees from all accredited U.S. institutions. The results of this annual survey are used to assess characteristics and trends in doctorate education and degrees. This information is vital for educational and labor force researchers in the Federal government and academia. The SED is sponsored by the following six Federal agencies: the National Science Foundation, the National Institutes of Health, the U.S. Department of Education, the U.S. Department of Agriculture, the National Endowment for the Humanities, and the National Aeronautics and Space Administration.

**DISCOVERY IN RESEARCH PORTFOLIO (DIRP)**

**Directorate for Computer & Information Science & Engineering (CISE)/Directorate for Social, Behavioral & Economic Sciences (SBE)**

**Advisory Subcommittee on Discovery in a Research Portfolio: Tools for Structuring, Analyzing, Visualizing and Interacting with Proposal and Award Portfolios**



The directorates of CISE and SBE created a subcommittee to leverage the NSF community’s interest in technological tools for structuring, analyzing, and visualizing an NSF program portfolio. This includes proposals, award abstracts, and research and education outcomes generated with support from an NSF program. This subcommittee was comprised of individuals with research expertise in machine learning, data mining, information visualization, human-centered computing, science policy, and visual analytics, with leadership provided by members of the CISE and SBE Advisory Committees. The subcommittee was tasked to identify effective tools and approaches for deriving knowledge from NSF program proposals and award portfolios.

The NORC Data Enclave played a pivotal role in the work of the subcommittee by providing research teams access to an environment that allowed for sophisticated scripting, development and database analysis while also conforming to the strict security guidelines necessary to house NSF proposal and award information. Of particular note, the work of this project provided enclave staff opportunity to expand its innovative service offerings to include more complex computational support and technical assistance.

For more information, please contact [fernandes-huessy-johannes@norc.org](mailto:fernandes-huessy-johannes@norc.org) or [abdul-razzak-nour@norc.org](mailto:abdul-razzak-nour@norc.org).

**NORC TEAM CONDUCTS BUSINESS INNOVATION SURVEY**

In collaboration with Principal Investigators John Walsh from Georgia Tech and Ashish Arora and Wes Cohen from Duke University and the National Bureau of Economic Research (NBER), NORC collected survey data from more than 6,500 U.S. businesses via computer-assisted telephone interviews (CATI). The study was sponsored by the National Science Foundation (NSF) and the Kauffman Foundation as part of the Science of Science and Innovation Policy program, and builds on earlier work conducted by the PIs (the Carnegie Mellon Survey on Industrial R&D). The NORC team designed, pre-tested, and conducted the survey, with topics covering areas such as new technology and processes, licensing, collaborative innovation, and firm characteristics.

For more information, please contact [mulcahy-tim@norc.org](mailto:mulcahy-tim@norc.org)



## KAUFFMAN FOUNDATION AWARDS NORC GRANT TO BUILD AND SUSTAIN THE KFS ENCLAVE RESEARCH COMMUNITY

The [Kauffman Foundation](#) and NORC have embarked on a new partnership that builds on the foundation's prior work with the Enclave. With this new 3-year grant, *Sustaining Virtual Research Communities*, the Kauffman Firm Survey (KFS) Enclave research community will expand by 20 researchers per year. The new grant also will involve receiving, archiving, indexing, and curating new waves of KFS business microdata.

KFS5, the fourth follow-up survey, was released in April 2010. The microdata are housed in the enclave along with complete metadata documentation. In addition, three new cohorts of researchers joined the KFS Enclave community this summer and fall, providing additional opportunities for collaboration and innovation.

For more information, please contact [mahmud-ayesha@norc.org](mailto:mahmud-ayesha@norc.org).



## NORC DATA ENCLAVE CELEBRATES 4TH YEAR ANNIVERSARY WITH FOUNDING SPONSOR NIST

As we celebrate the 4<sup>th</sup> anniversary of our partnership with NIST, we are currently supporting 19 researchers who are working on management and economics related projects. This year, we continue to bring in new researchers to the vibrant community.

For example, we recently announced two new waves of research scholarships that include access to confidential business microdata and a travel stipend worth \$1,500. The next call for proposals will be announced in January 2011. In Wave 1, we made five awards. Scholarship awardees include researchers from the University of Maryland, University of Florida, and the University of North Carolina—all focusing on innovation processes and how R&D affects firm outcomes.

It is also noteworthy that these 5 new researchers have been invited to participate in a pilot test that examines the viability of providing secure microdata access through a custom configured thin-client machine. These machines are configured with strict security features that greatly reduce the risk associated with providing access to confidential microdata. Researchers will provide the NORC team with feedback regarding the utility and functionality of the machines.

Also of note, NIST is collaborating with NORC in planning, developing, and implementing the 2<sup>nd</sup> Annual Conference on Microdata Access to be held at the National Press Club. Additional information about the conference may be found on the first page of this issue.

For more information, please contact [kuk-kenneth@norc.org](mailto:kuk-kenneth@norc.org)

## ANNIE E. CASEY FOUNDATION ADDS NEW RESEARCHERS AND FUNCTIONALITY TO THE ENCLAVE

We are pleased to report on the continuing growth of the Making Connections research group in the Data Enclave. The Making Connections group includes researchers from Harvard University, Princeton University, Brown University, the University of Wisconsin at Madison, Stanford, and the University of Chicago, among others. It is also worth noting that the enclave team has made substantial progress in providing harmonized datasets that allow researchers to more easily analyze differences within and across Making Connections sites over the course of multiple waves of data collection.

We are also pleased to report that we will be adding ArcGIS to the Enclave environment for all users. The addition reflects the work that Making Connections researchers in particular have put toward the use of supplementary and geographic data. This is a major addition to the software packages currently available in the Enclave, and we hope that researchers in all Enclave groups will find this a useful analytic tool.

For more information, please contact [hepburn-peter@norc.org](mailto:hepburn-peter@norc.org)



WHAT'S NEW IN THE DATA ENCLAVE?

NEW WEBSITE

The enclave team is in the process of redesigning the website, which we expect to complete early next year. The new website will make it easier for users to retrieve relevant information and also will include a [data cataloging tool](#) developed by the International Household Survey Network. The new tool will make it possible for users to browse through the datasets (devoid of frequencies), examine the variables, compare variables across datasets, and apply for access to the datasets. The tool is DDI compliant and is on-par with XML-based international standards designed for data and metadata exchange.

ENCLAVE WELCOMES NEW TEAM MEMBERS

The enclave team is excited to welcome three new members – Kenneth Kuk, Shaswat Sapkota and Nour Abdul-Razzak. Born and raised in Hong Kong, Kenneth received a B.S. in Policy Analysis and Management at Cornell University in 2010. His research interests involve human migration patterns, human capital investments, antitrust laws, corporate market transactions, and economics of information. Shaswat is originally from Nepal, and just received his Masters in Public Policy from the College of William and Mary in Williamsburg, VA. He is interested in areas of social policy and international development. Nour Abdul-Razzak graduated from the Massachusetts Institute of Technology in 2009 with Bachelor of Science degrees in both Civil Engineering and Political Science. She hopes to pursue a PhD in Economics with research interests in political economy, public economics, and development.

ENCLAVE OFFERS NEW SERVICES

As the enclave team has expanded, so has our ability to offer the full cycle of data services, ranging from study design and concept to data archiving and access. The enclave team will be able to provide a comprehensive set of analytic services, in addition to our main current services of data archiving and dissemination. The enclave team has strong technical and analytical backgrounds in social science research, and has ready access to the full range of analytic capabilities of NORC as a whole. Below are just a few examples of what services we now offer to both data producers and researchers:

- Statistical Analysis
- Disclosure Analysis
- Data Cleaning and Coding
- De-identification
- Micro-aggregation
- Visual Analytic Tools
- Spatial Analysis
- Diagnostics
- Harmonization
- Imputation

ENCLAVE'S INFRASTRUCTURE BUILD-OUT (NORC DATA ENCLAVE 3.0)

As the Data Enclave continues to grow both in size and sophistication, NORC continues to invest in improving the Data Enclave's infrastructure. In early 2011, NORC will add 20 servers to the farms currently running the Data Enclave. This investment will not only increase the possible user base to over 500 researchers, but also keep the Data Enclave on the cutting edge of virtual computing service delivery. In addition to increasing the resources available to Data Enclave users, these upgrades also will allow the Data Enclave to offer a broader spectrum of security options.

TEAM MEMBERS:

Ayesha Mahmud, Johannes Fernandes-Huessy, Kenneth Kuk, Nour Abdul-Razzak, Peter Hepburn, Shaswat Sapkota, Tim Mulcahy

RECENT TEAM PRESENTATIONS/ACCOMPLISHMENTS



- Johannes Fernandes-Huessy (see photo) was awarded a first place blue ribbon prize for a presentation on illicit drug markets (Mulcahy, Fernandes-Huessy, Brownstein, 2010) at the fall 2010 APPAM research conference in Boston.
- In 2010, Tim Mulcahy was invited to present at: the NSF's Future Investments in Large-Scale Survey Data Dissemination Workshop; USDA's National Agricultural Statistics Service National Leadership Conference; Federal Committee on Statistical Methodology's (FCSM) Statistical Policy Seminar and FCSM's Subcommittee on Confidentiality, the ASA Joint Statistical Meetings, the Association for Public Policy Analysis and Management; and the International Association for Social Science Information Services and Technology. In February 2011, Mulcahy has been invited to be a keynote speaker at Eurostat's conference on New Techniques and technologies for Statistics in Brussels.

CONTACT: TIM MULCAHY

NORC at the University of Chicago  
 4350 East-West Highway, Suite 800  
 Bethesda, MD 20814  
 Phone: 301-634-9352  
 Fax: 301-634-9301  
 E-mail: [dataenclavemanager@norc.org](mailto:dataenclavemanager@norc.org)