Research resources @ Kellogg

Scholarly research is an essential component of the work of faculty and doctoral students at the Kellogg School of Management. Empirical research, including econometric analysis, numerical simulations and conducting experiments and surveys are integral parts of contemporary research endeavors. The Kellogg School facilitates these tasks in a variety of ways. Over the years, more than two thirds of the doctoral dissertations produced by graduate students at the school have included some empirical component.

This document provides an overview of the research services and resources available at Kellogg; it does not fully reflect the scope of existing information technology resources at the school. Please visit the following site for more complete descriptions and documentation on the resources mentioned in this document:

www.kellogg.northwestern.edu/researchcomputing/

1. The research support group

The goal of the research support group is to facilitate the use of the research resources and facilities available at Kellogg. This broad task includes not only the logistical aspect of maintaining resources, but also supporting their use, educating users, and troubleshooting in statistical computing.

The group provides consultation services on statistical and mathematical software, assistance with the data sources to which Kellogg subscribes, as well as guidance and support in Institutional Review Board (IRB) submissions. In addition, the group can provide user-training sessions for groups of doctoral students. One-on-one sessions may be arranged for faculty.

The research support team also helps faculty in the preparation of research-oriented resources for use in MBA classes, such as handouts on how to access a specific resource or how to use specific routines of supported statistical and mathematical software. They can also help the faculty member by training teaching assistants.

Currently, the research support team includes four senior research associates and a behavioral research manager. It is led by Assistant Dean Patricia Ledesma and works under the guidance of Senior Associate Dean: Faculty & Research David Austen-Smith.

2. Sources of data

Researchers at Kellogg use data from a wide range of providers; in some cases, the data are gathered by the researcher via surveys or experiments. Through Kellogg, faculty and doctoral

students gain access to data repositories and commercial services with economic and corporate (financial and accounting) information. There are two main institutional subscriptions of interest to Kellogg members: Kellogg's subscription to Wharton Research Data Services (WRDS) and Northwestern Library's membership in the Inter-University Consortium for Political and Social Research (ICPSR).

Wharton Research Data Services (WRDS)

Data from the Center for Research in Security Prices (CRSP) and Standard & Poor's *Compustat* are among the most widely used at Kellogg. Both datasets, in addition to data from the Federal Deposit Insurance Corporation (FDIC), the NYSE's *Trade and Quote database* (TAQ), Compustat's *ExecuComp*, and Thomson Financial's *CDA/Spectrum Institutional (13f) and Mutual Fund Holdings, OptionMetrics' IvyDB*, etc can be accessed through Wharton Research Data Services, *WRDS,* to which Kellogg subscribes. Doctoral students and faculty are eligible for login ids for this service. Data can be accessed either through a web interface, or via SAS and FORTRAN programs through secure shell.

Bloomberg, Datastream, SDC Platinum and other business specific data sources

Kellogg also subscribes to other specialized data services such as *Bloomberg*, *FactSet*, *Datastream*, *SDC Platinum*, *Thomson ONE Analytics*, *Hoover's Online*, *Zacks History Files*, etc. In addition, Kellogg has acquired data from sources such as the Loan Price Corporation and the Futures Industry Institute.

Through the Northwestern Library, researchers also have access to *DRI/Basic Economics*, the IMF's *International Financial Statistics*, the World Bank's *World Development Indicators* and *Global Development Finance*, and the ILO's *Key Indicators of the Labour Market*, among others. The NU Library is a federal repository of data and publications. Thus, data produced by US government agencies (such as the Bureau of the Census or the Bureau of Labor Statistics) is available through the library.

Chicago Research Data Center (CRDC)

Through the CRDC, researchers with approved projects can access a wide variety of Census economic micro data that are not normally available for public use. The CRDC is the result of collaboration between the Census Bureau and a consortium consisting of Argonne National Laboratory, the Federal Reserve Bank of Chicago, Northwestern University, the University of Chicago and the University of Illinois at Chicago. The center is housed at the Federal Reserve Bank of Chicago.

Inter-University Consortium for Political and Social Research (ICPSR)

The Northwestern Library, through the Social Science Data Center, is the University's representative to ICPSR, which is the largest data repository in the world. Topics range from census enumerations to surveys about organizational behavior, to values and norms in different nations. In addition to being a rich source of data, the study codebooks provided by ICPSR can be a reference source for researchers who intend to carry out their own survey. Another benefit of membership in ICPSR is reduced tuition in ICPSR's Summer Program in Quantitative Methods,

which includes courses in topics such as hierarchical linear models, scaling and dimensional analysis, social network analysis, etc.[1]

Many of the ICPSR datasets are installed in the Social Sciences computing facility, described below. Thus, work with large-scale datasets such as the US Census, the Panel Study of Income Dynamics, or the National Longitudinal Survey of Youth, is greatly facilitated.

Roper Center for Public Opinion Research

NU Library's Social Science Data Center also manages the institutional membership to the Roper Center, an archive of polls by leading survey research organizations in the US and abroad, including from institutions such the National Science Foundation, National Opinion Research Center, Gallup Organization, etc.

Other sources of data

Many international and government agencies allow free access to their data (or part of it) through the web. Good examples of this are the Bank for International Settlements, the World Bank, the Securities and Exchange Commission (SEC), and other government agencies. Northwestern's Library also manages the institutional subscription to the Roper Center for Public Opinion Research.

Many faculty and doctoral students have obtained data from private sources other than the ones subscribed to by Kellogg and the University Library. In the past, researchers have used studies carried out by Information Research Inc. (IRI), A.C. Nielsen, Goldman Sachs, etc. Generally, access to these types of data is arranged through individual departments and research centers within Kellogg.

The table at the end of this document is a partial list of datasets available to faculty and students and is intended to give a rough idea of data availability by topic. In the table, a number of datasets available through ICPSR and government agencies have been included as a reference of the type of data available through those sources.

Bibliographic and full text resources

In addition to "numerical" databases and services, both Kellogg and the Northwestern Libraries subscribe to various full-text databases. Finally, journals are increasingly available in electronic formats. Kellogg faculty and doctoral students may sign up to receive e-mail journals from the Social Science Research Network (SSRN).

Among the most widely used bibliographic databases are EconLit, Sociological Abstracts, the Social Science Citation Index (Web of Science) and the Current Index to Statistics.

---

[1] Since Northwestern is a member of the Committee on Institutional Cooperation (CIC), faculty and doctoral students may take ICPSR courses paying the tuition through Northwestern. Membership in CIC also facilitates admission to summer courses offered by the University of Michigan's Survey Research Center.

Full text databases include Academic Search Elite, Academic Universe (Lexis/Nexis), ABI/Inform, JSTOR, Investext Plus, ISI Emerging Markets, and the Business & Industry Database, among others.

3. Facilities for behavioral and experimental research

Behavioral studies and experiments have been part and parcel of research in organizational behavior and marketing for many decades. Experiments are of increasing importance in economics, political economy and finance. Kellogg has well equipped state of the art, behavioral research facilities and support.

Subject pools

Kellogg actively maintains three subject pools, managed with SONA systems, a web based subject pool administration platform. The first two subject pools have approximately 4,000 and 3,700 participants, mostly Northwestern University undergraduate students. The second subject pool, dubbed "eLab," consists of over 6,800 local and non local participants, recruited nationwide.

The SONA system allows researchers to create studies and recruit participants from the subject pool. The setup includes a basic pre-test questionnaire that allows researchers to screen out potential participants or target specific subsets of the pool.

Laboratories and break-out rooms

Kellogg currently has two behavioral laboratories for computer aided experiments. In addition, there are six 'break-out' rooms for experimental studies on small group negotiations and interactions.

Scheduling of the laboratories is done via a web application, Meeting Room Manager.

The Kellogg School has been a pioneer in the study of negotiations and dispute resolution. To facilitate this type of research, the break-out rooms are discreetly equipped with video cameras mounted near the ceiling and wired for video and audio recording. The cameras are powered remotely from the control room and the video is fed into six DVD recorders on a recording rack. The control room also has a video editing station.

Software for behavioral research and experiments and web surveys

MediaLab, DirectRT and z-Tree are licensed by the school and installed in all the workstations in the behavioral labs. There are also two platforms for web surveys, Qualtrics and ViewsFlash.

fMRI studies

In addition to the school's behavioral labs, students interested in neuroscience experiments (e.g., neurofinance or neuroeconomics) may gain access to functional magnetic resonance imaging (fMRI) facilities through the Center for Advanced MRI at Northwestern University.

4. Computing facilities

The backbone of the school's computing architecture is a Windows Server network. Most doctoral students are assigned to shared offices in which they have access to desktop computers or can bring their own computers and connect to the network. Doctoral students in good academic standing have the option of requesting the purchase of a laptop during their first three years in the program – this option is managed by the PhD Program Office.

Computers and scholarly research are inextricably linked – the last decade has seen an explosion in the availability of data and in the processing capacity of technology. Beyond a desktop or laptop, faculty and doctoral students at Kellogg have access to a variety of computing platforms that can cater to a very broad range of research problems and research styles.

"Skew," the school's dedicated research Linux server

The school maintains a dedicated research server with a rich suite of software and compilers. The current server, "skew5," runs Red Hat Linux Enterprise on a system with 24 cores and 256GB of RAM. Among the software packages installed in the system are Matlab, Mathematica, SAS, Stata/MP2, R, Gauss, S Plus, RATS, TSP, Maple, etc.

"kap01," the school's research Windows application server and Data Warehouse

Later this academic year, a new research server is going into production, the Kellogg Data Warehouse and an associated Windows-based application server ("kap01"). This is a high performance system for research problems that require massive amounts of data and are computationally intensive. The Data Warehouse offers SQL services and high performance data storage (110TB), while the Application server offers access to Stata/MP32, Matlab and SAS, with 80 cores and 256GB of RAM.

Social Science Computing Cluster (SSCC)

The SSCC is a Linux cluster customized for social science research. It offers access to over 260 cores with 4GB of RAM per core. Like skew, the SSCC include a wide variety of statistical and econometric software. The SSCC is a collaborative venture between Northwestern's Office

of the President, Northwestern's Academic Technologies, the Kellogg School and the Weinberg College of Arts & Sciences.

Quest Cluster

Quest is a university-wide large scale high performance computing system, with over 7,000 cores and 100 TB of attached storage. In addition to standard compilers and debuggers (C, C++, and FORTRAN), Quest also includes Mathematica, Matlab, R, KNITRO and Open MPI, among other applications and libraries.

Sheridan Cluster, a Windows virtual machine farm

The Sheridan Cluster is a Microsoft Windows based virtual machine "farm": researchers log into a Windows server remotely using a client application and access a dedicated virtual machine with two processors, running Windows 7 x64, 6GB of RAM and access to specialty software such as ArcGIS, Atltas.ti, Nvivo, MPlus, WinBUGS, Ucinet, etc, in addition to R, SAS and Stata/SE. Launched during the fall of 2009, the Sheridan Cluster is a joint venture between the Kellogg School and Weinberg. It is currently open only to behavioral researchers.

"vault," research data storage services

Vault is a university-wide secure storage system for research. Users can access the service through the web and desktop clients, and are able to share data with internal and external collaborators for specified periods of time.

5. Software

There is a wide variety of software for research computing. In this category, we include statistical and mathematical software, as well as some programming languages, and utilities.

A list of the statistical and mathematical software applications is shown in the table below. In addition to these applications, you may use:

- *Terminal emulation packages*: To access Linux and UNIX servers, you need a terminal emulation package. The recommended packages are SSH Secure Shell and StarNet's *X-Win32*, both available under a Northwestern site license. *X-Win32*, in particular, is an X Windows emulation package.

- *Programming languages and compiler*s: Skew and other UNIX servers have Fortran (f77 and f90), C, C++, perl, etc.

- *Text editors*: In UNIX systems, there is a variety of text editors used to write programs. Among the most popular choices are *pico*, *emacs*, *nedit*, and *vi*. All of these are available on skew.

- *Utilities*: *DBMS/Copy* is a utility program, available on skew, which translates datasets between different formats.

- *Word processing*: *Scientific WorkPlace*, a specialized package that allows easy creation of technical documents using LaTeX, is installed at the Jacobs Center computer lab. Some departments make this program available in public access computers for their

doctoral students. Another popular choice of LaTeX users is MiKTeX, a shareware implementation of TeX for Windows. teTeX is installed in skew.

STATISTICAL AND MATHEMATICAL SOFTWARE AVAILABLE AT KELLOGG

	Where?								Type of application				
	skew	*kap01*	*Kellogg PC lab*	*Sheridan cluster*	*SSCC*	*Quest*	*Behavioral labs*	*Site license **	*Statistics, general purpose*	*Statistics, special applications*	*Matrix based numerical computation*	*Math, symbolic problem solving*	*Behavioral, experimental research*
Atlas-ti				✓						✓			
DirectRT							✓	✓					✓
Gauss	✓				✓						✓		
HLM				✓	✓					✓			
Limdep					✓					✓			
Lisrel, Prelis					✓					✓			
Maple	✓							✓				✓	
Mathematica	✓					✓		✓				✓	
Matlab	✓	✓	✓		✓	✓		✓			✓		
MediaLab							✓	✓					✓
Nvivo				✓						✓			
Ox					✓						✓		
R	✓				✓	✓			✓				
RATS	✓									✓			
SAS	✓	✓	✓		✓	✓		✓	✓		✓		
S-plus	✓				✓				✓				
SPSS			✓				✓	✓	✓				
Stata	✓	✓	✓	✓	✓				✓				
TSP	✓										✓		
z-Tree							✓	✓					✓

* Personal copies available through a discount plan (GradPlan).

Datasets by General Topic
(This list is <u>not</u> comprehensive)

Dataset	Equity Prices / Returns / Distributions	Accounting data	Macroeconomic data	Fixed Income	Derivatives	Executive compensation	Financial Transactions	Mutual Funds	Demographics / Consumer level	Industry data	Where to access
Annual Survey of Manufactures										✓	Web (Bureau of the Census)
Bloomberg	✓	✓	✓	✓							Special terminals in Jacobs PC Labs, Finance
CDA Spectrum	✓	✓									WRDS
Census data									✓		Web (Bureau of the Census), NU Social Science Data Services
Census of Manufactures										✓	Web (Bureau of the Census)
CME tick data					✓						skew
Compustat		✓									WRDS
comScore									✓		WRDS
CRSP	✓							✓			WRDS
Current Population Survey									✓	✓	Web (Bureau of the Census), NU Social Science Data Services
Datastream	✓	✓	✓	✓							Windows 2010, Kellogg server
DRI Basic Economics			✓							✓	Windows / NU Library web site
Execucomp						✓					WRDS
Economática	✓	✓									Windows 2010, Kellogg server
Fixed Investment Securities Database				✓							skew
Future Industry Institute				✓							skew
International Financial Statistics			✓								NU Library
IBES	✓										WRDS
ISSM Transactions File Databases	✓										skew, WRDS
Lancer Analytics Insider	✓	✓									skew, WRDS
Panel Study of Income Dynamics									✓		Web, PSID site at University of Michigan
OptionMetrics	✓										WRDS
SDC - Joint Ventures							✓				Windows 2010, Kellogg server
SDC - Mergers							✓				Windows 2010, Kellogg server
SDC - New Issues							✓				Windows 2010, Kellogg server
SDC - Repurchases							✓				Windows 2010, Kellogg server
Survey of Consumer Finances									✓		Web, Board of Governors of the Federal Reserve System
TAQ	✓										WRDS
Zacks History Files	✓										skew

Servers available to Kellogg researchers

Server	"Audience"	Specs	Software	Type of jobs
skew5	Kellogg faculty and doctoral students	Red Hat Linux, 4 x six core Intel Xeon E7540, 3 GHz, 256GB of RAM, 10TB storage	Matlab, Mathematica, Maple, SAS, Stata/MP2, Gauss, Rats, Splus, TSP, R, Limdep, StatTransfer, Fotran, C++ compilers, CPLEX, KNITRO, …	General use, jobs that demand RAM or "boutique" applications
SSCC (Social Science Computing Cluster)	Kellogg and Weinberg faculty and doctoral students	Red Hat Linux, 260 AMD Opteron CPU cores, 2.8GHz, 4GB of RAM per CPU, 14TB of storage	Matlab, SAS, Stata/SE, Gauss, Splus, R, Limdep, StatTransfer, Fotran, C++ compilers, KNITRO, Lisrel, HLM, …	Simulations/optimizations, multiple jobs, parallel computing (users limited to 20 cores), "boutique" applications
Quest	Northwestern faculty and doctoral students	Red Hat Linux, 4032 cores Intel Nehalem E5520 (2.26GHz, 6GB of RAM per core) and 3024 Intel Westmere X5650 (2.66GHz, 6GB of RAM per core) , 100 TB storage	Matlab, Mathematica, R, KNITRO, OpenMPI, Fortran, C++, libraries (ATLAS, gsl, Boost, Open MPI)	Simulations/optimizations, parallel computing, computationally intensive jobs
Sheridan Cluster	Kellogg and Weinberg behavioral faculty and doctoral students	Win7 64 bit / Citrix; up to 15 simultaneous users, clients get virtual terminal with 2 cores (2.13GHz) and 6GB of RAM	Atlas.ti, ArcGIS, HLM, Mplus, Nvivo, R, Stata/SE, SPSS, Ucinet, WinBUGS, StatTransfer	Work that requires special applications, not much RAM; "boutique" applications
Application server + data warehouse (in production by summer 2012)	Kellogg faculty and doctoral students	Application server : Win 7 64 bit, 80 cores Intel Westmere-EX, 256GB of RAM, 80TB storage (SSD w/SATA3) ; clients get virtual terminal. SQL data warehouse : 40 cores, 128GB of RAM	Matlab, SAS, Stata/MP32	Jobs that demand RAM, some types of computationally intensive or data intensive