JPSM
THE JOINT PROGRAM IN
SURVEY METHODOLOGY

Institute for Employment
Research

The Research Institute of the
Federal Employment Agency

IAB

FCSM/WSS Workshop on Quality of Blended Data

26. Februar 2018

Summary

Frauke Kreuter

Lessons learned

Combining Data Sources

When assessing quality, we need to focus on

Y

We need to get comfortable with proxies in

Y and X

We need to remember the initial question

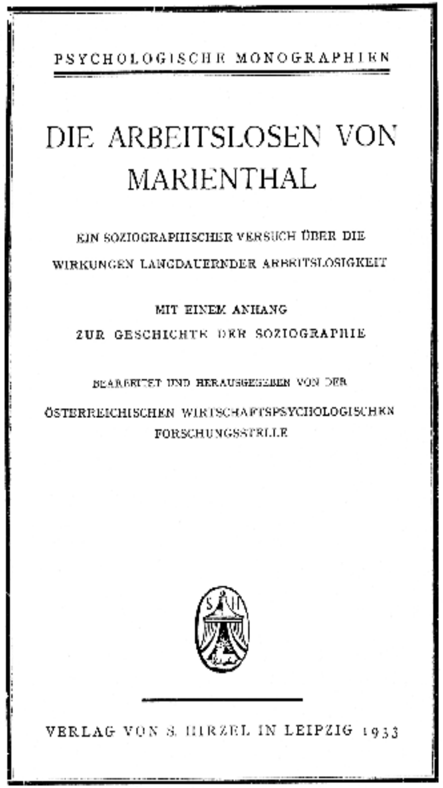
?

We need to change the way we operate

Lessons not yet learned

Combined data collection

Research Question – Effects of Unemployment



PSYCHOLOGISCHE MONOGRAPHIEN

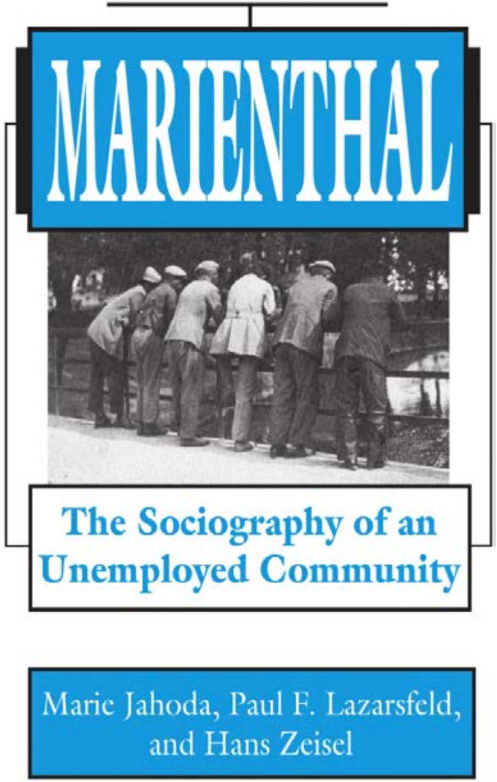
DIE ARBEITSLOSEN VON MARIENTHAL

EIN SOZIOGRAPHISCHER VERSUCH ÜBER DIE WIRKUNGEN LANGDAUERNDER ARBEITSLOSIGKEIT

MIT EINEM ANHANG ZUR GESCHICHTE DER SOZIOGRAPHIE

BEARBEITET UND HERAUSGEGEBEN VON DER ÖSTERREICHISCHEN WIRTSCHAFTSPSYCHOLOGISCHEN FORSCHUNGSSTELLE

VERLAG VON S. HIRZEL IN LEIPZIG 1933



MARIENTHAL

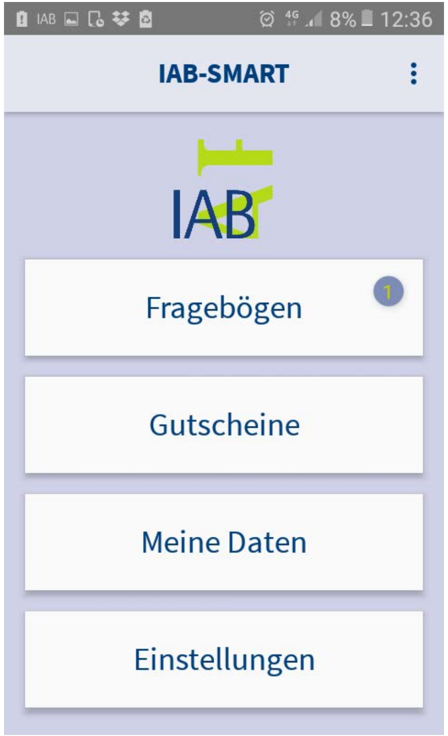
The Sociography of an Unemployed Community

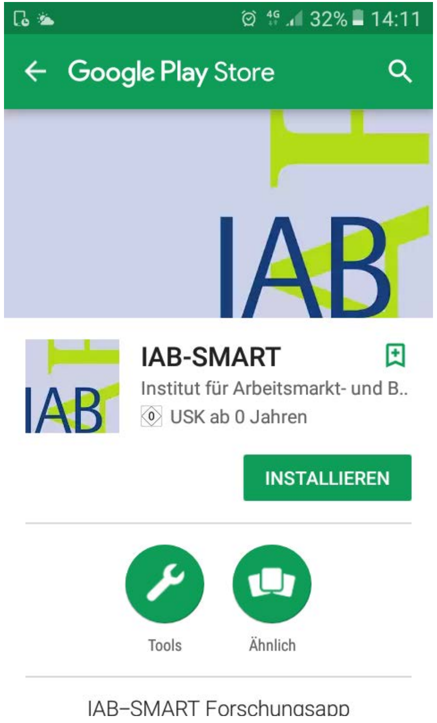
Marie Jahoda, Paul F. Lazarsfeld, and Hans Zeisel

Research Question – Effects of Unemployment

Research-App, that …

- … issues questionnaires
- … collects passive data
- … links to panel survey and administrative data





PASS – Panel (10 years) + Administrative Data

Sample of households with at least one welfare benefit recipient (at reference date)

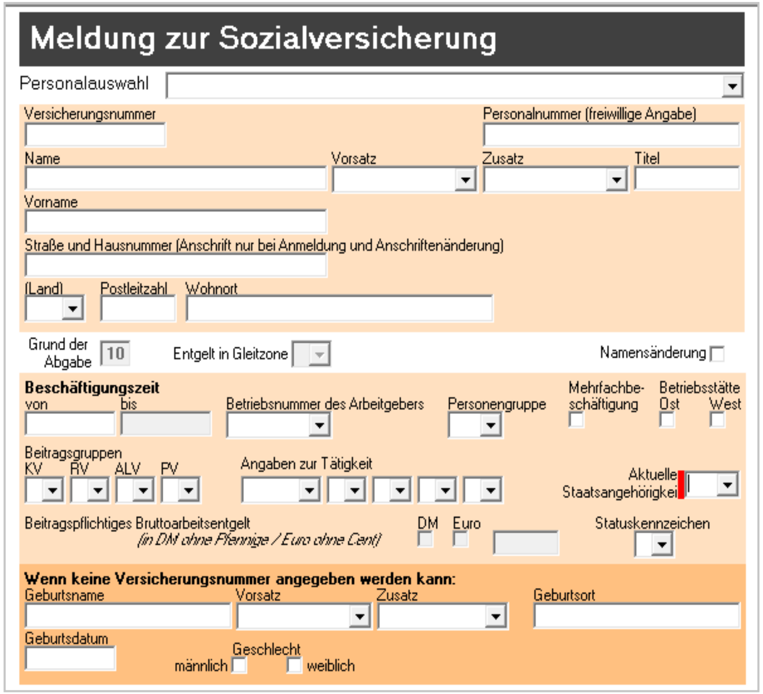
Refreshed annually

Surveyed annually

Random household sample of resident population

Refreshed annually

Surveyed annually



Meldung zur Sozialversicherung

Personalauswahl

Versicherungsnummer

Personalnummer (freiwillige Angabe)

Name

Vorsatz

Zusatz

Titel

Vorname

Straße und Hausnummer (Anschrift nur bei Anmeldung und Anschriftenänderung)

(Land)

Postleitzahl

Wohnort

Grund der Abgabe 10

Entgelt in Gleitzone

Namensänderung □

Beschäftigungszeit

von

bis

Betriebsnummer des Arbeitgebers

Personengruppe

Mehrfachbe-schäftigung □

Betriebsstätte Ost □ West □

Beitragsgruppen KV RV ALV PV

Angaben zur Tätigkeit

Aktuelle Staatsangehörigkei

Beitragspflichtiges Bruttoarbeitsentgelt *(in DM ohne Pfennige / Euro ohne Cent)*

DM □ Euro □

Statuskennzeichen

Wenn keine Versicherungsnummer angegeben werden kann:

Geburtsname

Vorsatz

Zusatz

Geburtsort

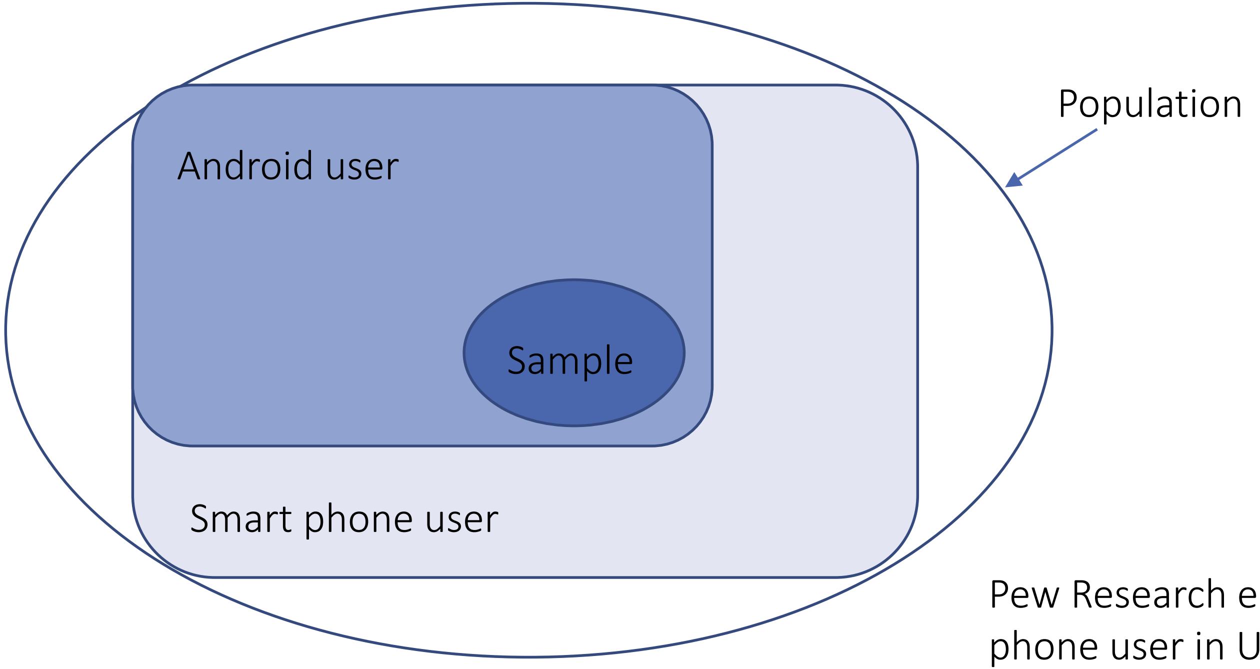
Geburtsdatum

Geschlecht

männlich □ □ weiblich

Trappmann M., Christoph B., Achatz J., Wenzig C. (2009) PASS: a new panel study for labour market research, Int. J. of Manpower , 30, 7, pp.765-770

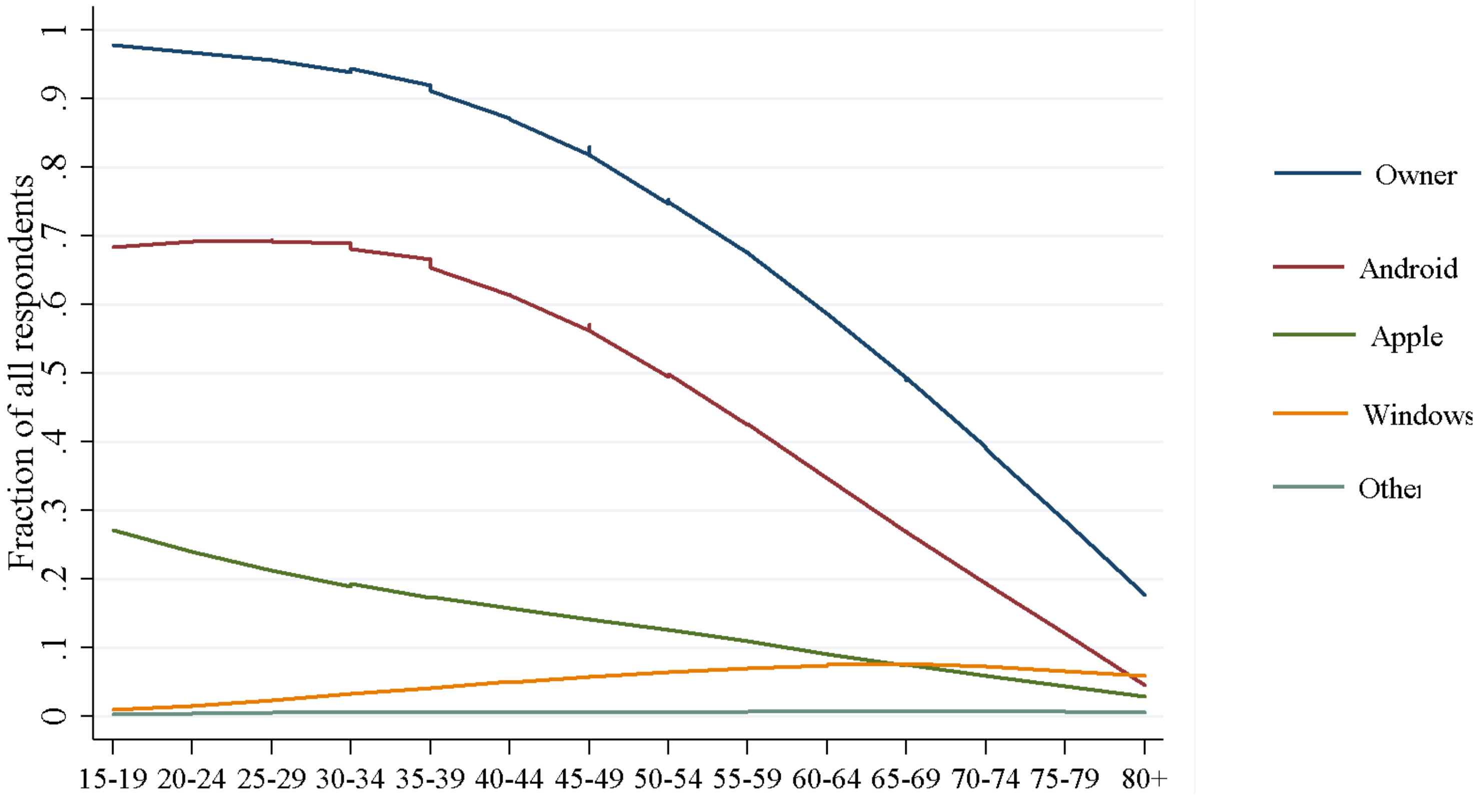
Coverage // Selection // External Validity



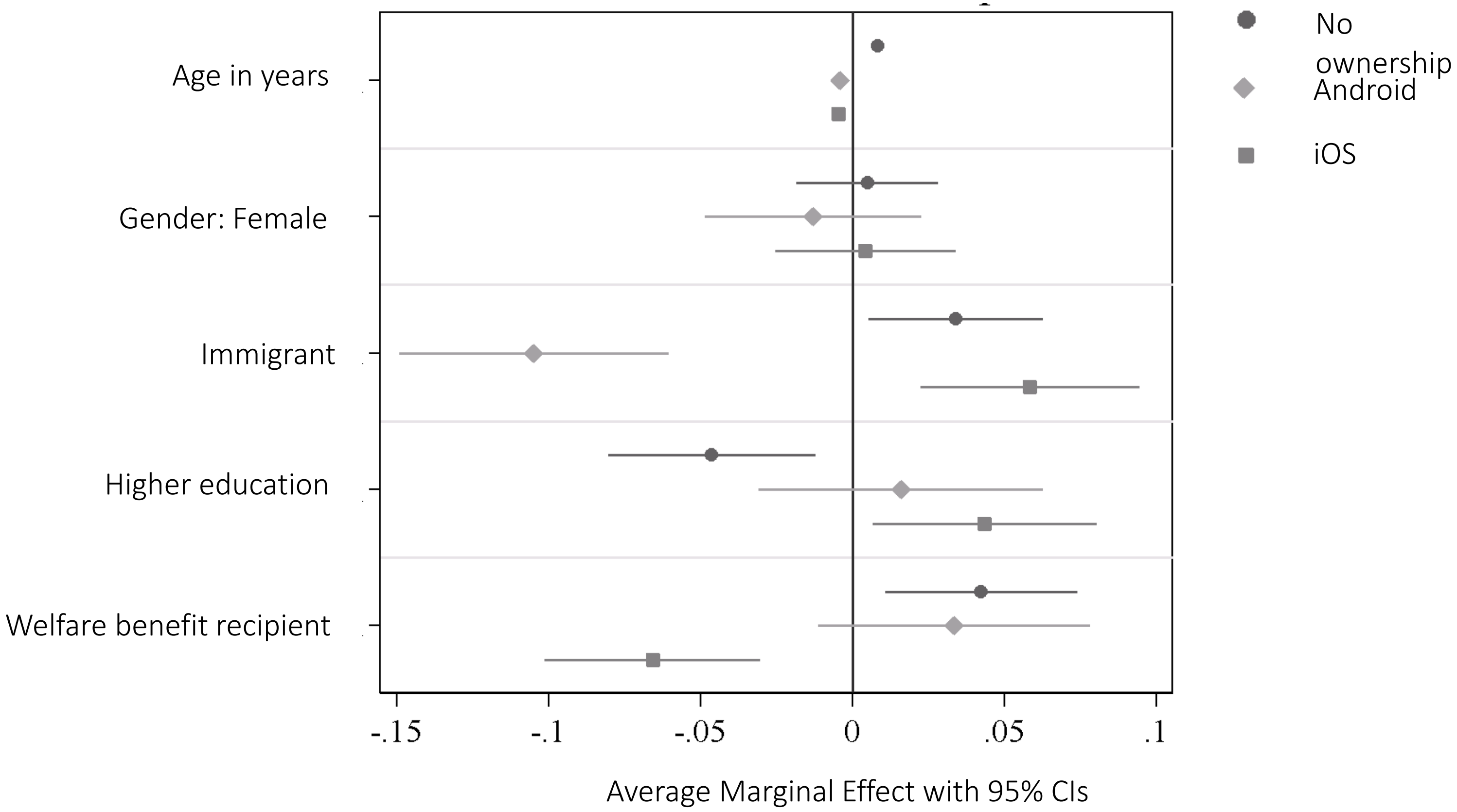
Pew Research estimates: 77% smart phone user in U.S. in 2016

Source: Valliant R, Dever J, Kreuter F (2018): Practical Tools for Designing and Weighting Survey Samples. 2nd Edition. New York: Springer.

Ownership by age groups (unweighted PASS estimates)

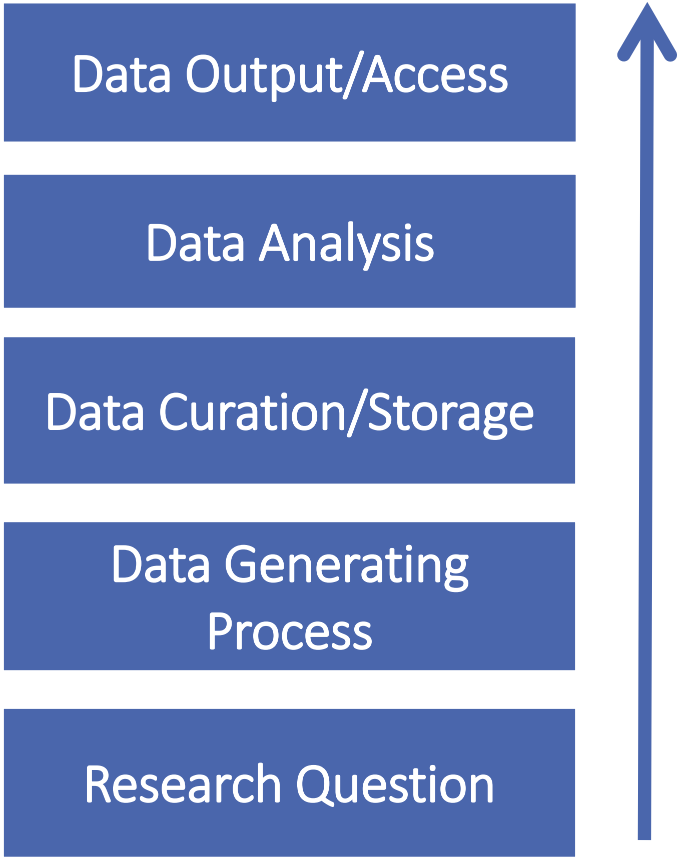


Predicting ownership and device type



Lessons offered

Survey and Data Science



Data Output/Access

Learn how to communicate results and distribute and store your data

Data Analysis

Learn a variety of analysis methods suited for different data types

Data Curation/Storage

Learn how to curate and manage data

Data Generating Process

Understand how to collect data yourself, and how data are generated through administrative and other processes.

Research Question

Learn how to formulate your research goal and which data are best suited to achieve this goal.

Source: Usher in Japec et al 2015

survey-data-science.net

INTERNATIONAL PROGRAM IN SURVEY AND DATA SCIENCE

offered through the University of Mannheim and the Joint Program in Survey Methodology

(Universities of Maryland and Michigan, Westat)

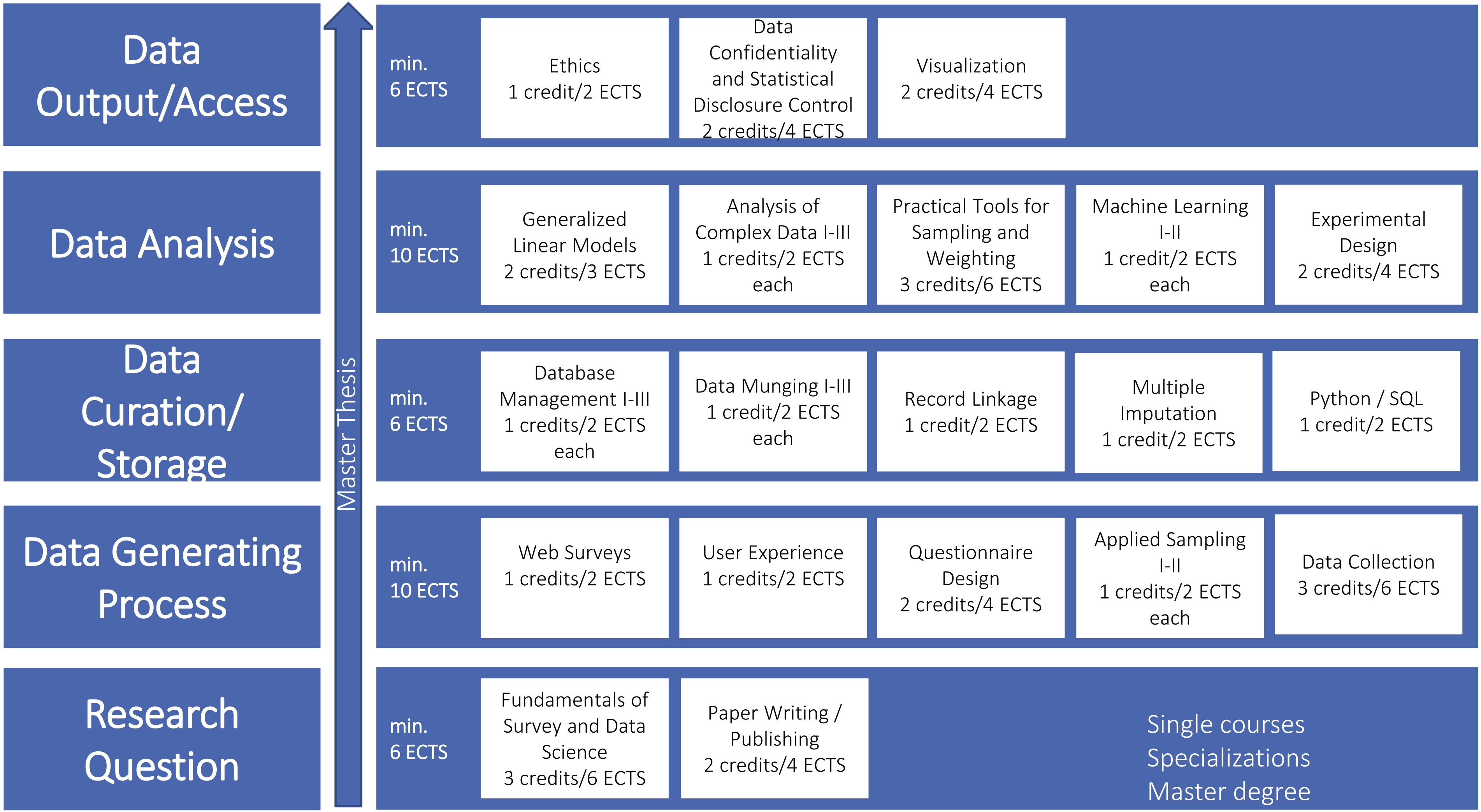
JPSM
THE JOINT PROGRAM IN SURVEY METHODOLOGY

IN OMNIBUS VERITAS
UNIVERSITÄT MANNHEIM

SPONSORED BY THE

Federal Ministry
of Education
and Research

AUFSTIEG DURCH BILDUNG
OFFENE HOCHSCHULEN



Data Output/Access	min. 6 ECTS	Ethics 1 credit/2 ECTS	Data Confidentiality and Statistical Disclosure Control 2 credits/4 ECTS	Visualization 2 credits/4 ECTS		
Data Analysis	min. 10 ECTS	Generalized Linear Models 2 credits/3 ECTS	Analysis of Complex Data I-III 1 credits/2 ECTS each	Practical Tools for Sampling and Weighting 3 credits/6 ECTS	Machine Learning I-II 1 credit/2 ECTS each	Experimental Design 2 credits/4 ECTS
Data Curation/ Storage	min. 6 ECTS	Database Management I-III 1 credits/2 ECTS each	Data Munging I-III 1 credit/2 ECTS each	Record Linkage 1 credit/2 ECTS	Multiple Imputation 1 credit/2 ECTS	Python / SQL 1 credit/2 ECTS
Data Generating Process	min. 10 ECTS	Web Surveys 1 credits/2 ECTS	User Experience 1 credits/2 ECTS	Questionnaire Design 2 credits/4 ECTS	Applied Sampling I-II 1 credits/2 ECTS each	Data Collection 3 credits/6 ECTS
Research Question	min. 6 ECTS	Fundamentals of Survey and Data Science 3 credits/6 ECTS	Paper Writing / Publishing 2 credits/4 ECTS			

Master Thesis

Single courses
Specializations
Master degree

Faculty

U. of Maryland / Michigan:

Chris Antoun

Fred Conrad

Steven Heeringa

Partha Lahiri

James Lepkowski

Richard Valliant

University of Mannheim:

Thomas Gautschi

Florian Keusch

Thomas Fetzer

Heiner Stuckenschmidt

Other universities:

Helmut Kuechenhoff (LMU Munich)

Daniel Oberski (Utrecht University)

Trent Buskirk (U. Mass, Boston)

Simon Munzert (HU Berlin)

Government Agencies:

Manfred Antoni (IAB)

Jörg Drechsler (IAB)

Joseph Sakshaug (IAB)

Stefan Bender (Bundesbank)

Jeffrey Gonzalez (BLS)

Carolina Franco (Census)

Private partners:

Mario Callegaro (Google)

Jennifer Romano-Bergstrom (Facebook)

Jill Dever (RTI)

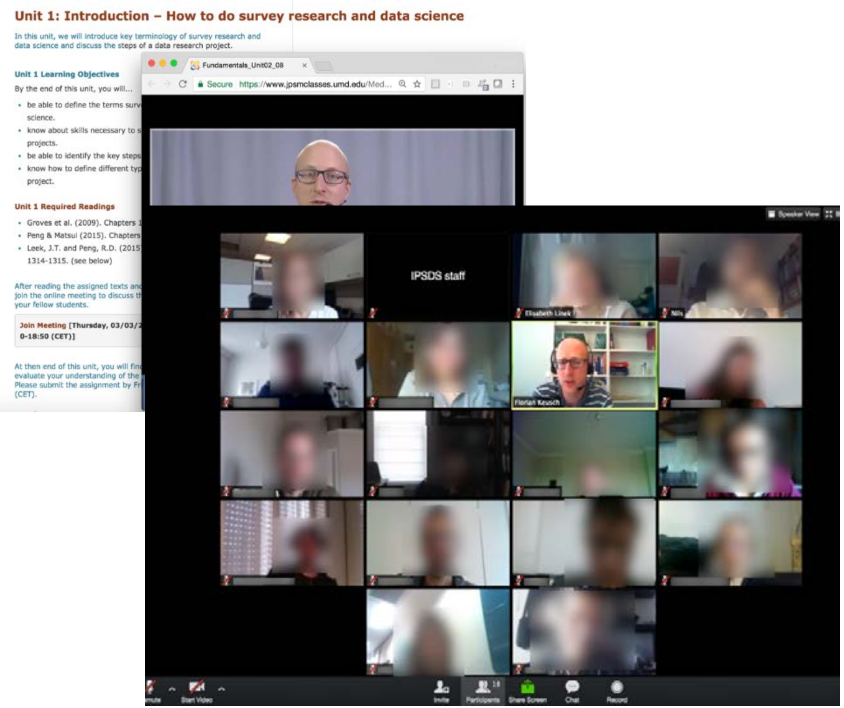
Emily Geisen (RTI)

Raphael Nishimura (Abt)

Roger Tourangeau (Westat)

Onsite (Connect@IPSDS)

Online



Unit 1: Introduction – How to do survey research and data science

In this unit, we will introduce key terminology of survey research and data science and discuss the steps of a data research project.

Unit 1 Learning Objectives

By the end of this unit, you will...

- be able to define the terms surv science.
- know about skills necessary to s projects.
- be able to identify the key steps
- know how to define different typ project.

Unit 1 Required Readings

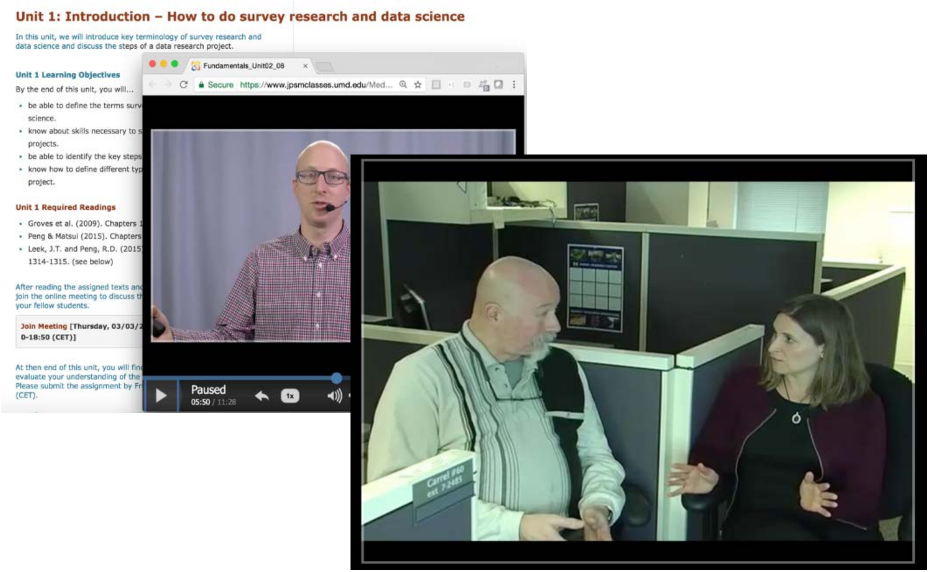
- Groves et al. (2009). Chapters 1
- Peng & Matsui (2015). Chapters
- Leek, J.T. and Peng, R.D. (2015 1314-1315. (see below)

After reading the assigned texts and join the online meeting to discuss th your fellow students.

Join Meeting [Thursday, 03/03/2 0-18:50 (CET)]

At then end of this unit, you will fin evaluate your understanding of the Please submit the assignment by Fr (CET).

Asynchronous



Unit 1: Introduction – How to do survey research and data science

In this unit, we will introduce key terminology of survey research and data science and discuss the steps of a data research project.

Unit 1 Learning Objectives

By the end of this unit, you will...

- be able to define the terms surv
science.
- know about skills necessary to s
projects.
- be able to identify the key steps
- know how to define different typ
project.

Unit 1 Required Readings

- Groves et al. (2009). Chapters 1
- Peng & Matsui (2015). Chapters
- Leek, J.T. and Peng, R.D. (2015
1314-1315. (see below)

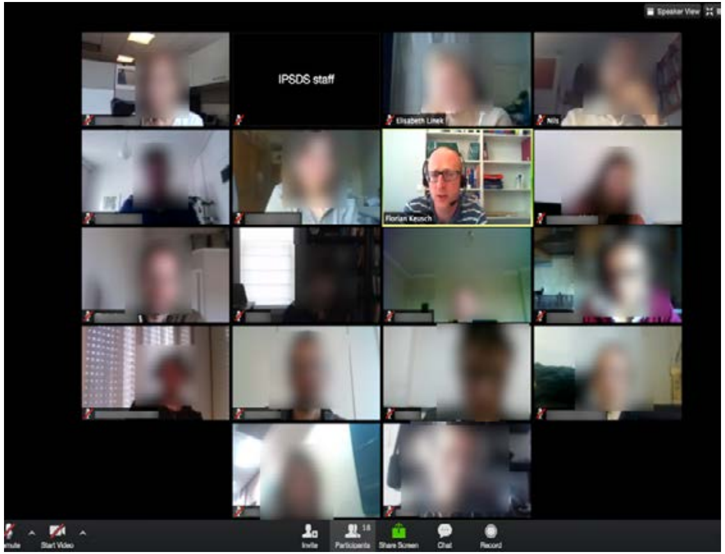
After reading the assigned texts and
join the online meeting to discuss th
your fellow students.

**Join Meeting [Thursday, 03/03/2
0-18:50 (CET)]**

At then end of this unit, you will fin
evaluate your understanding of the
Please submit the assignment by Fr
(CET).

- Pre-recorded lectures
(split into smaller video units)
- (Bi)weekly assignments
- Discussion forums

Synchronous

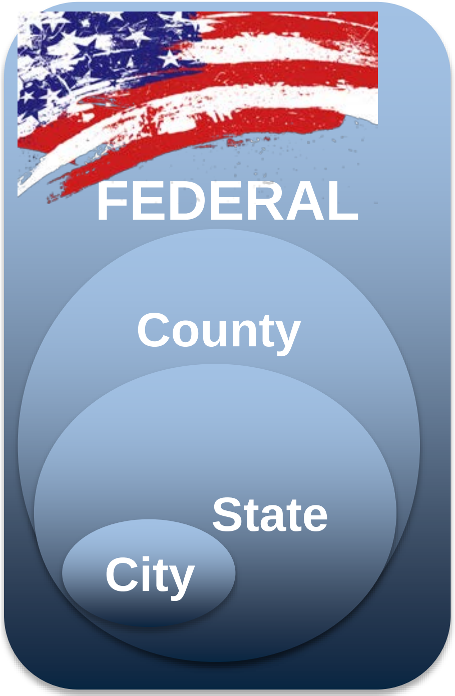


- Small virtual classrooms
- Weekly 50-minute discussions led by the instructor
- Obligatory component

Community is key

Coleridge Initiative

Networks: The first two classes brought together ~40 agencies from city, state, county and federal agencies



HEALTHY CHICAGO
CHICAGO DEPARTMENT OF PUBLIC HEALTH

MECKLENBURG COUNTY NORTH CAROLINA

NEW YORK CITY HOUSING AUTHORITY

CHICAGO POLICE

NEW YORK STATE OF OPPORTUNITY.
Office of Information Technology Services

USDA

Virginia Department of Behavioral Health & Developmental Services

City of Milwaukee

Illinois Department of Corrections

THE CITY OF SOUTH BEND, INDIANA
PEACE
1865

IDES
ILLINOIS DEPARTMENT OF EMPLOYMENT SECURITY

Allegheny County Department of Human Services

UNITED STATES GOVERNMENT ACCOUNTABILITY OFFICE

U.S. DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT

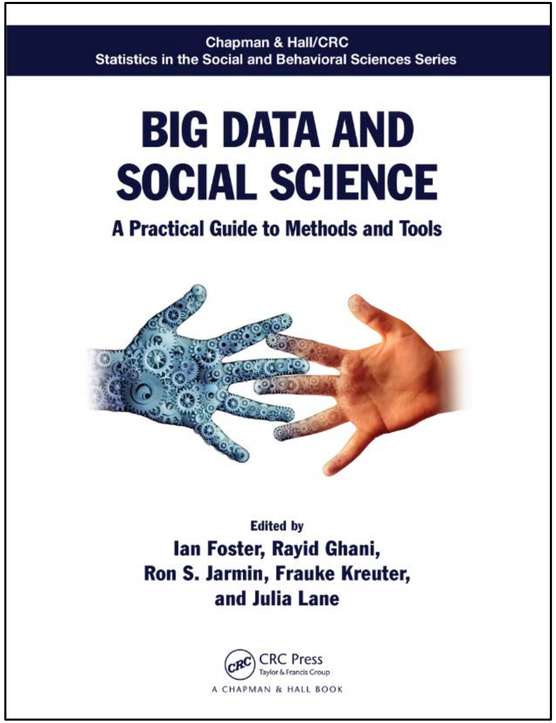
NYC .gov always open

COUNTY of DANE WISCONSIN
1839

NEW YORK STATE | Office of Alcoholism and Substance Abuse Services

CITY OF BOULDER

Professional Training Workshops



Chapman & Hall/CRC
Statistics in the Social and Behavioral Sciences Series

BIG DATA AND SOCIAL SCIENCE

A Practical Guide to Methods and Tools

Edited by
Ian Foster, Rayid Ghani,
Ron S. Jarmin, Frauke Kreuter,
and Julia Lane

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Three Classes

- Different cohorts (ex-offenders, welfare recipients and veterans)
- Joined with housing, transportation and jobs data

Class Format

- Module 1: Foundations – Research Questions, Python, SQL
- Module 2: Data Acquisition – Web Scraping, API, Record Linkage
- Module 3: Data Analysis – Machine Learning, Networks, Text, Spatial
- Module 4: Visualization, Inference, Ethics, Privacy

Additional Information

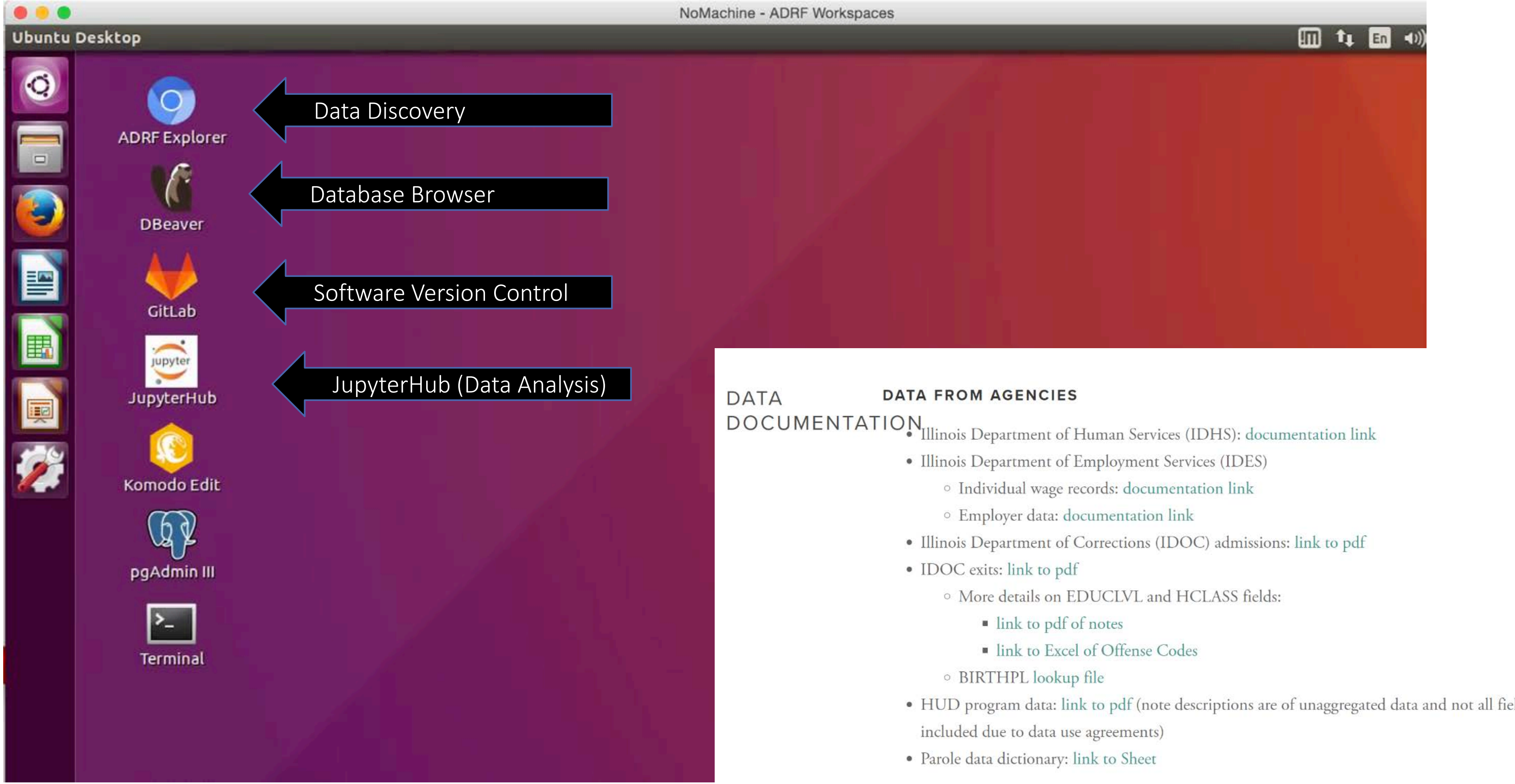
- Final reports are all virtual
- Teaching Assistants and facilitators will be at each site for each module

THE UNIVERSITY OF CHICAGO

NEW YORK UNIVERSITY

UNIVERSITY OF MARYLAND

Collaborative secure environment



Data Discovery

Database Browser

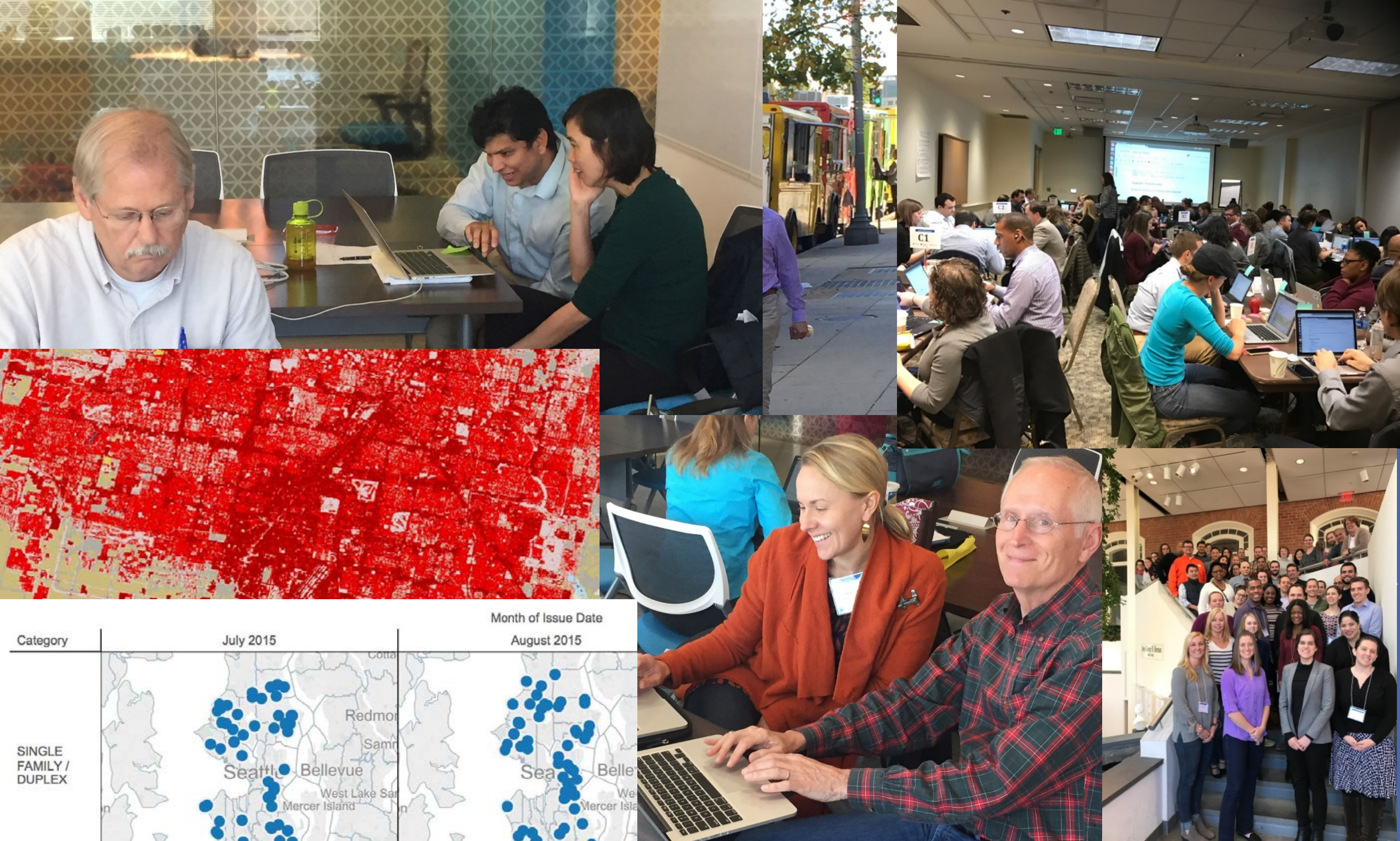
Software Version Control

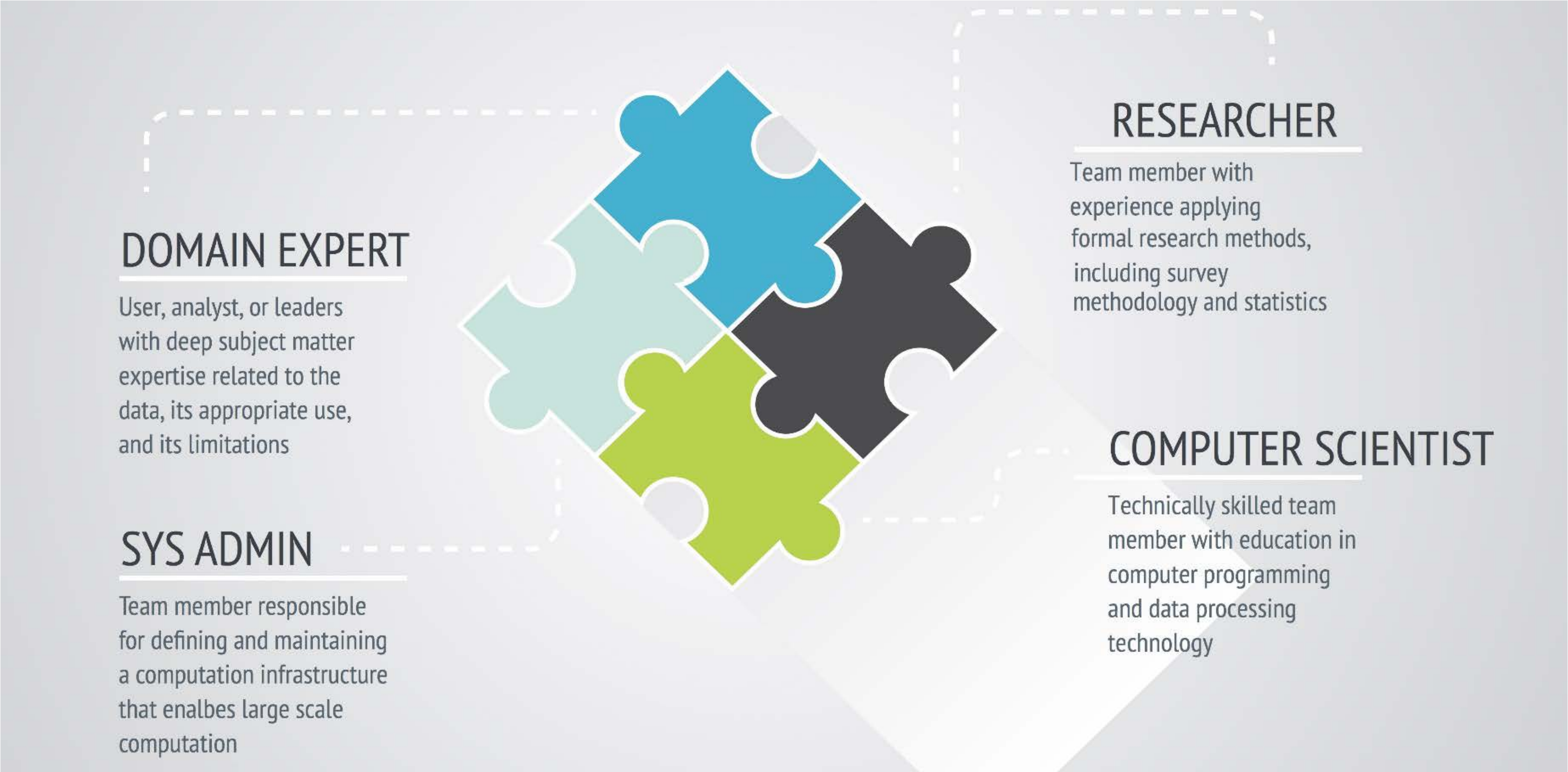
JupyterHub (Data Analysis)

DATA DOCUMENTATION

DATA FROM AGENCIES

- Illinois Department of Human Services (IDHS): documentation link
- Illinois Department of Employment Services (IDES)
  - Individual wage records: documentation link
  - Employer data: documentation link
- Illinois Department of Corrections (IDOC) admissions: link to pdf
- IDOC exits: link to pdf
  - More details on EDUCLVL and HCLASS fields:
    - link to pdf of notes
    - link to Excel of Offense Codes
  - BIRTHPL lookup file
- HUD program data: link to pdf (note descriptions are of unaggregated data and not all fie included due to data use agreements)
- Parole data dictionary: link to Sheet


Month of Issue Date
Category
July 2015
August 2015
SINGLE FAMILY / DUPLEX
Cotta
Redmon
Samm
Seattle
Bellevue
West Lake Sa
Mercer Island
Sea
Belle
We
Mercer Isla
C1
C2



DOMAIN EXPERT

User, analyst, or leaders with deep subject matter expertise related to the data, its appropriate use, and its limitations

SYS ADMIN

Team member responsible for defining and maintaining a computation infrastructure that enalbes large scale computation

RESEARCHER

Team member with experience applying formal research methods, including survey methodology and statistics

COMPUTER SCIENTIST

Technically skilled team member with education in computer programming and data processing technology

Source: Abe Usher

Shift in mindset! Dare to experiment (now)!

Thank you!

Frauke Kreuter (fkreuter@umd.edu)

survey-data-science.net
coleridgeinitiative.org