Teaching Survey and Data Science Outside the Regular Classroom

Frauke Kreuter

University of Maryland, U.of Mannheim

Atlanta 3/26/18

Project coordinators and funding

Laura & John
Arnold Foundation

U.S. National
Science Foundation

SPONSORED BY THE

Federal Ministry
of Education
and Research

AUFSTIEG DURCH BILDUNG
OFFENE HOCHSCHULEN

The project on which this report is based was funded by the Federal Ministry of Education and Research under the number [16OH22064]. Responsibility for the contents of this publication lies with the author.

AAPOR

AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH

AAPOR Report on Big Data

AAPOR Big Data Task Force

February 12, 2015

Prepared for AAPOR Council by the Task Force, with Task Force members including:

Lilli Japec, Co-Chair, Statistics Sweden
Frauke Kreuter, Co-Chair, JPSM at the U. of Maryland, U. of Mannheim & IAB
Marcus Berg, Stockholm University
Paul Biemer, RTI International
Paul Decker, Mathematica Policy Research
Cliff Lampe, School of Information at the University of Michigan
Julia Lane, American Institutes for Research
Cathy O'Neil, Johnson Research Labs
Abe Usher, HumanGeo Group

Acknowledgement: We are grateful for comments, feedback and editorial help from Eran Ben-Porath, Jason McMillan, and the AAPOR council members.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

CONSENSUS STUDY REPORT

FEDERAL STATISTICS, MULTIPLE DATA SOURCES, AND PRIVACY PROTECTION

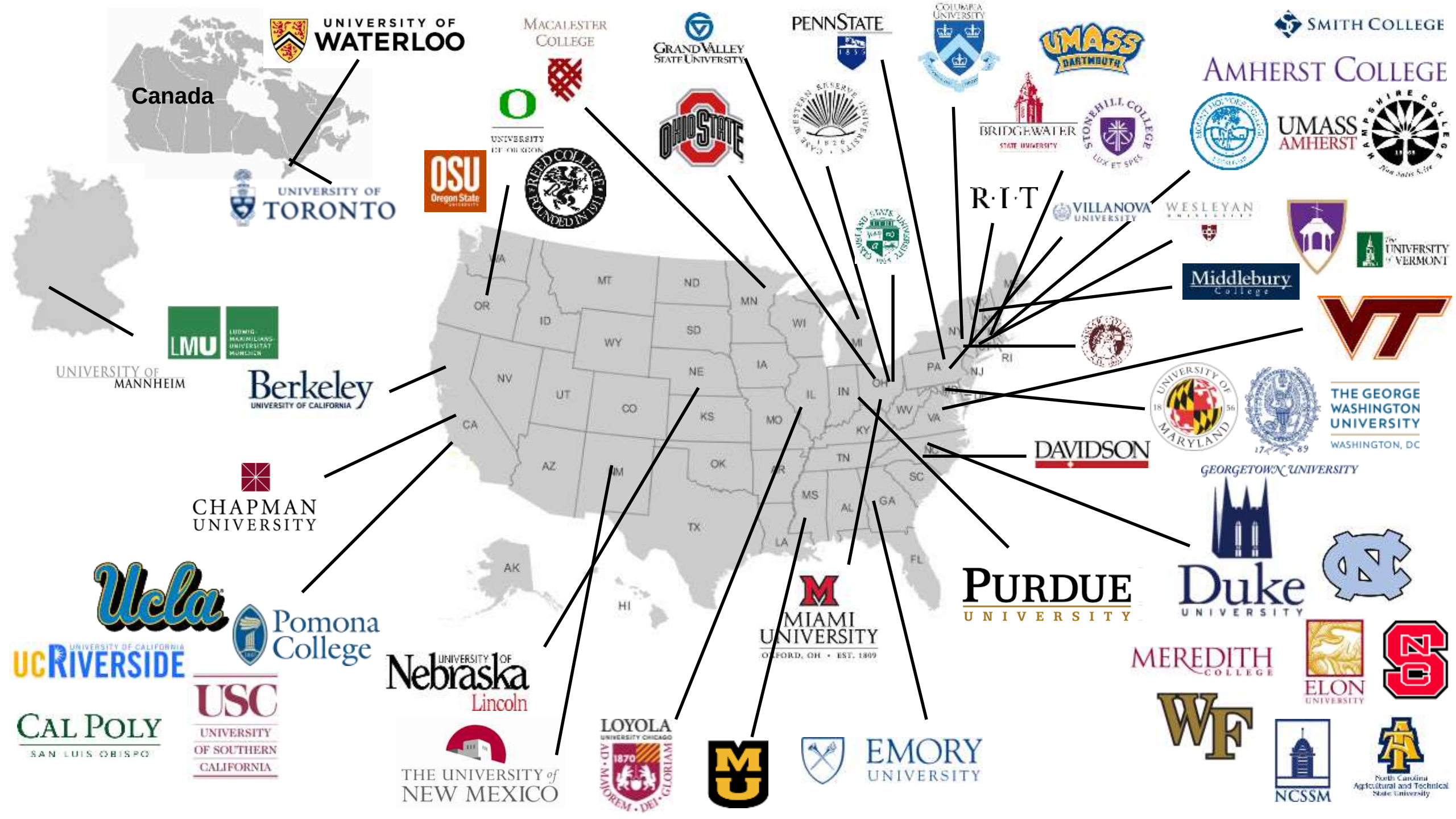
Next Steps

THE PROMISE OF EVIDENCE-BASED POLICYMAKING

Report of the Commission on Evidence-Based Policymaking

1st Example – ASA DataFest


WA
OR
CA
ID
NV
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
OK
TX
MN
IA
MO
AR
LA
WI
IL
MS
MI
IN
KY
TN
AL
OH
WV
GA
FL
SC
NC
VA
PA
NY
MD
DE
NJ
CT
RI
MA
VT
NH
ME
AK
HI


Canada
UNIVERSITY OF WATERLOO
UNIVERSITY OF TORONTO
MACALESTER COLLEGE
GRAND VALLEY STATE UNIVERSITY
PENNSTATE
COLUMBIA UNIVERSITY
UMASS DARTMOUTH
SMITH COLLEGE
AMHERST COLLEGE
UNIVERSITY OF OREGON
OHIO STATE
CASE WESTERN RESERVE UNIVERSITY
BRIDGEWATER STATE UNIVERSITY
STONEHILL COLLEGE
UMASS AMHERST
HAMPSHIRE COLLEGE
OSU Oregon State
REED COLLEGE
R·I·T
VILLANOVA UNIVERSITY
WESLEYAN
CASE WESTERN RESERVE UNIVERSITY
Middlebury College
THE UNIVERSITY OF VERMONT
LMU LUDWIG MAXIMILIANS UNIVERSITÄT MÜNCHEN
UNIVERSITY OF MANNHEIM
Berkeley UNIVERSITY OF CALIFORNIA
WA
OR
CA
ID
NV
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
OK
TX
MN
IA
MO
AR
LA
WI
IL
MS
MI
IN
KY
TN
AL
OH
WV
GA
FL
SC
NC
VA
PA
NY
NJ
RI
MD
DE
CHAPMAN UNIVERSITY
DAVIDSON
UNIVERSITY OF MARYLAND
THE GEORGE WASHINGTON UNIVERSITY WASHINGTON, DC
GEORGETOWN UNIVERSITY
AK
HI
Ucla
Pomona College
UC RIVERSIDE UNIVERSITY OF CALIFORNIA
CAL POLY SAN LUIS OBISPO
USC UNIVERSITY OF SOUTHERN CALIFORNIA
University of Nebraska Lincoln
THE UNIVERSITY of NEW MEXICO
LOYOLA UNIVERSITY CHICAGO
MIAMI UNIVERSITY OXFORD, OH · EST. 1809
PURDUE UNIVERSITY
Duke UNIVERSITY
MEREDITH COLLEGE
ELON UNIVERSITY
EMORY UNIVERSITY
NCSSM
North Carolina Agricultural and Technical State University

ASA DataFest™

- Teams of 3-5 undergraduates
- Friday evening – Sunday afternoon
- One (unknown) data set
- Three winning categories:
  Best insights
  Best visualization
  Best use of outside data

- Best educational experience ever !

2nd Example – Coleridge Initiative

COLERIDGE INITIATIVE

Building the capacity needed to accelerate the effective use of new data.

"Data, data everywhere, we have to stop and think"

with apologies to the Rime of the Ancient Mariner

UPCOMING TRAINING PROGRAMS

Spring 2018
Kansas City, MO

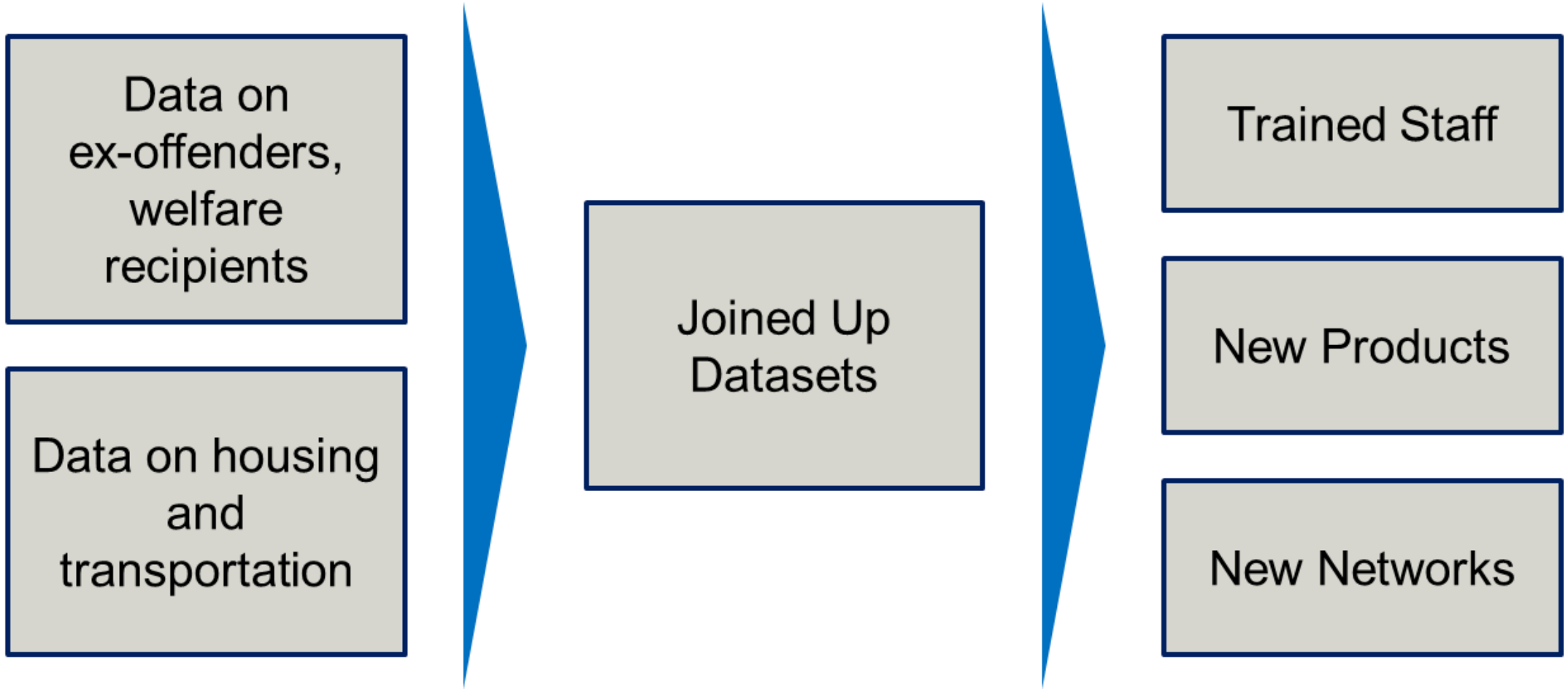
Application Closed

Summer 2018

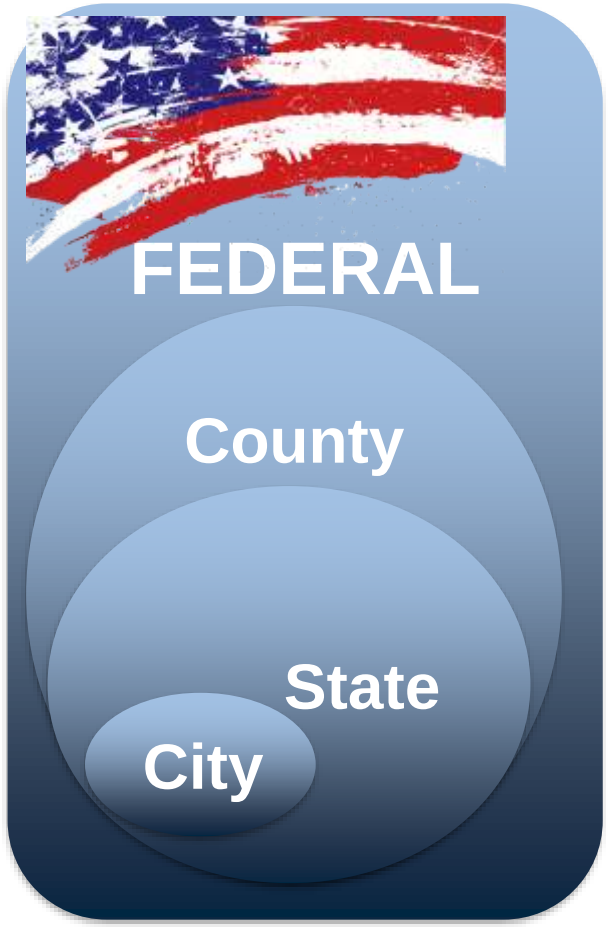
Apply by May 11, 2018

Program directors Rayid Ghani, Frauke Kreuter, and Julia Lane are also co-editors of "Big Data and Social Science: A Practical Guide to Methods and Tools.", the text book for the class.

Approach: hands-on with real microdata



The first 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

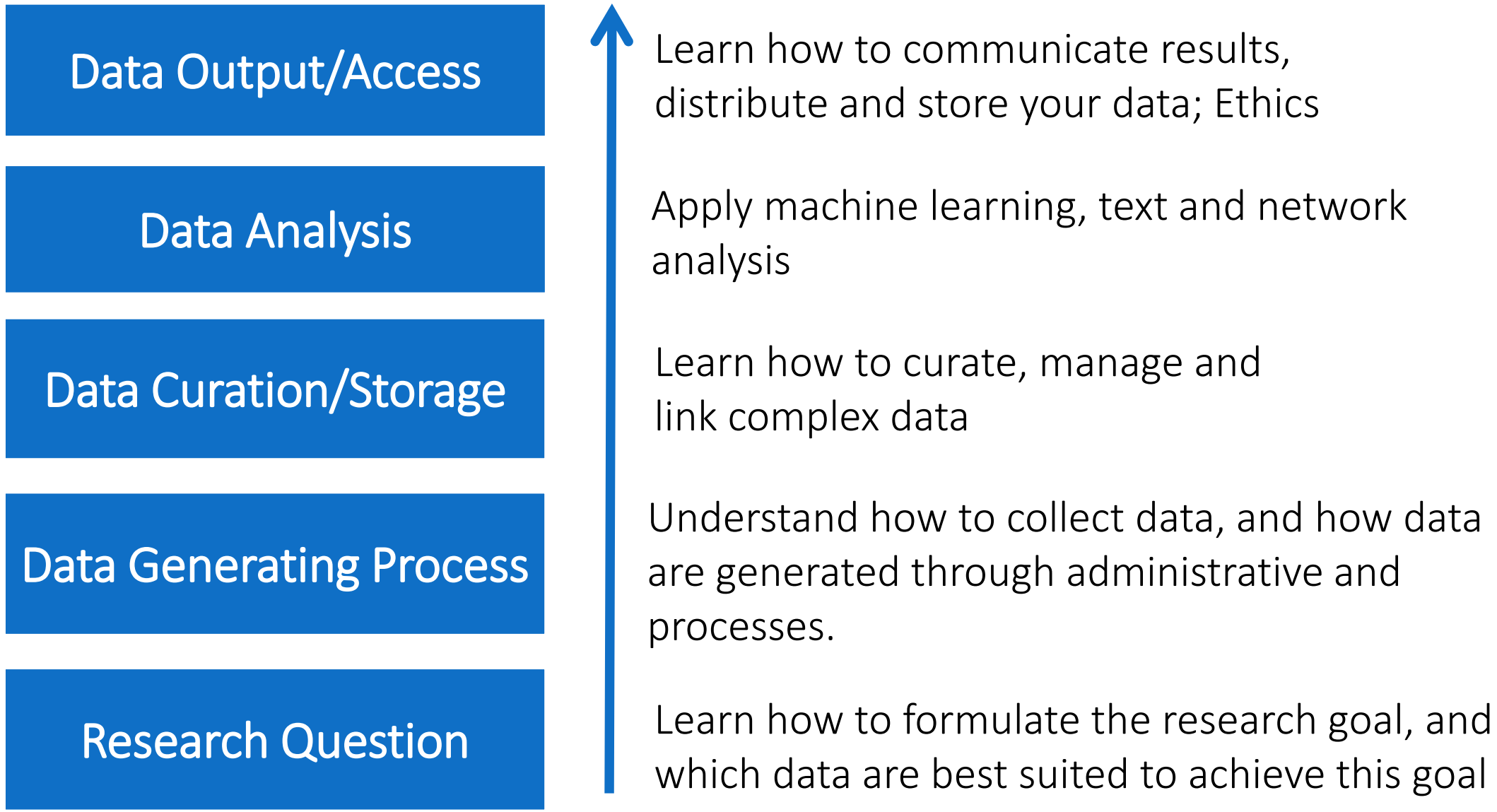
COUNTY of DANE WISCONSIN
1839

NYC.gov
always open

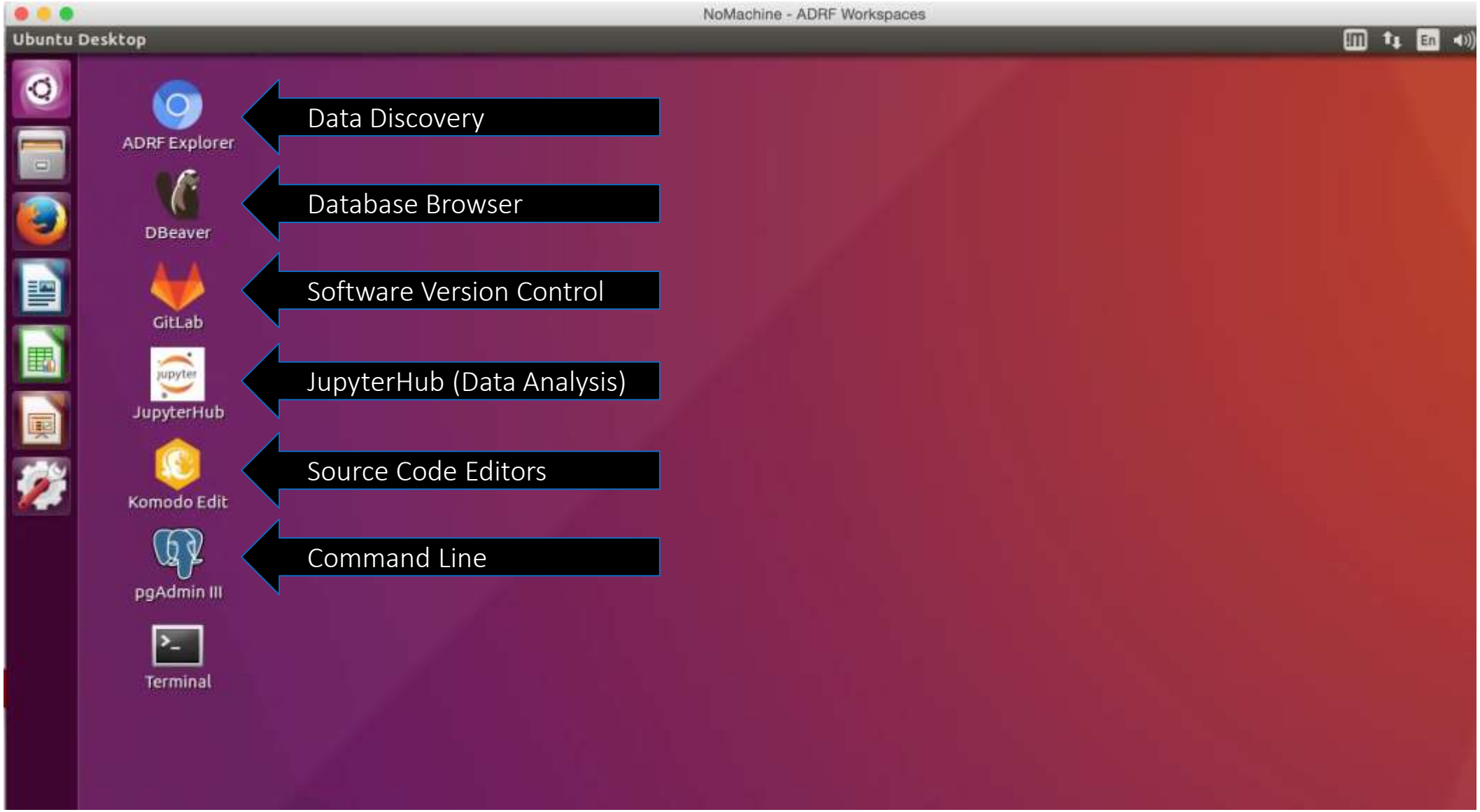
NEW YORK STATE | Office of Alcoholism and Substance Abuse Services

CITY OF BOULDER

Modules

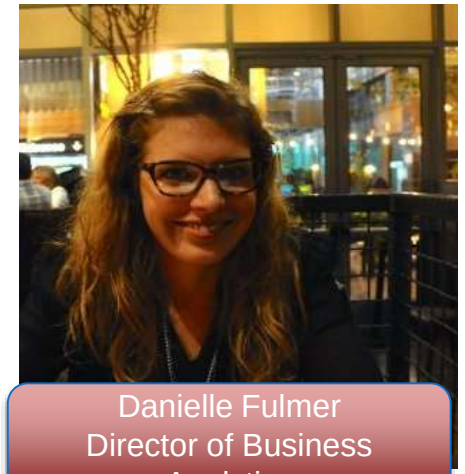

Data Output/Access
Learn how to communicate results, distribute and store your data; Ethics
Data Analysis
Apply machine learning, text and network analysis
Data Curation/Storage
Learn how to curate, manage and link complex data
Data Generating Process
Understand how to collect data, and how data are generated through administrative and processes.
Research Question
Learn how to formulate the research goal, and which data are best suited to achieve this goal

Collaborative and Secure Environment



What our participants say about the program

"Love the Jupyter notebooks!! ... I love how the code snippets and explanations are set up in the Jupyter notebooks. The format of going through it individually and discussing questions/challenges in your group, with the experts available when needed, worked really well for my learning style."



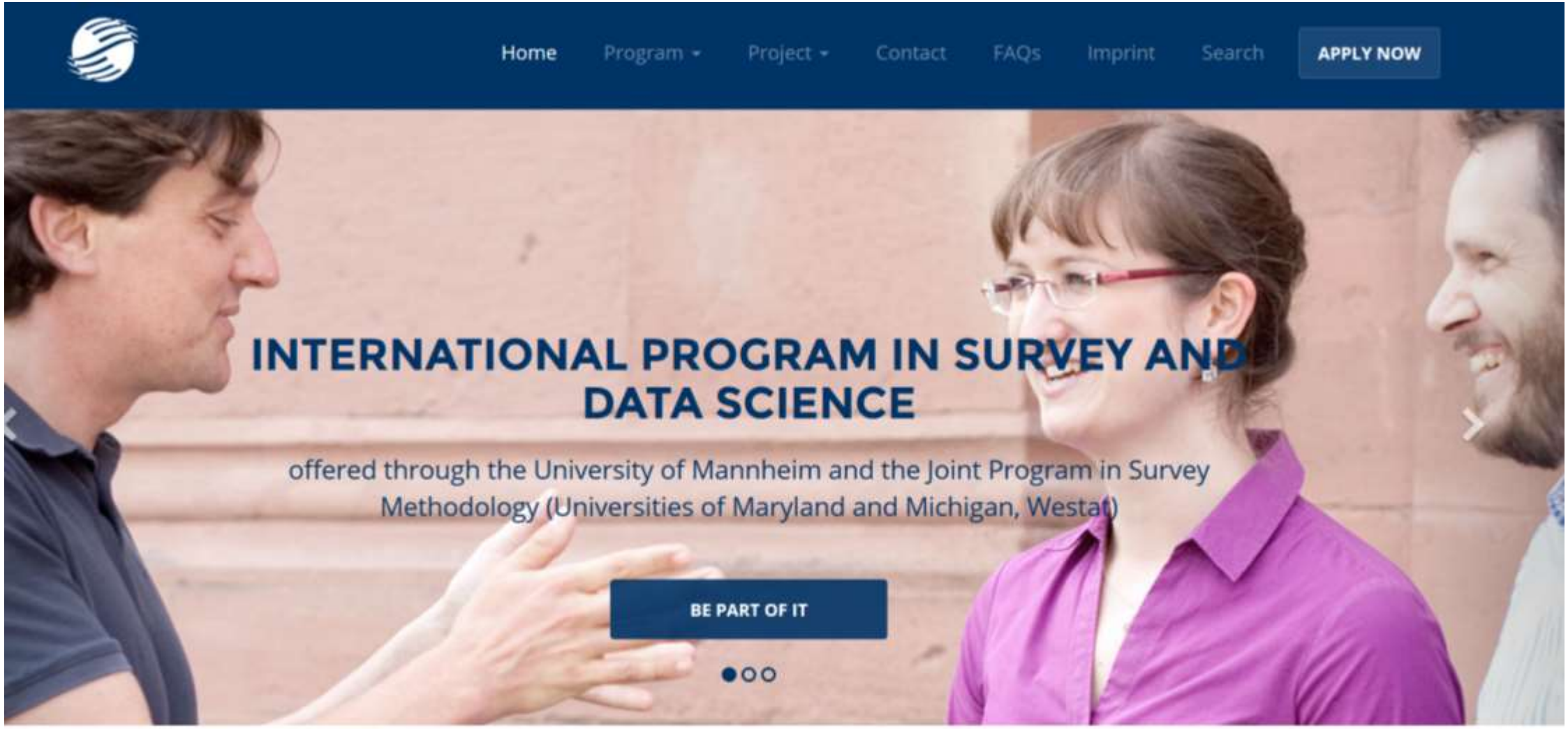
Danielle Fulmer
Director of Business Analytics

I could see our agency benefiting potentially from something like this in that, as the system builds out and collects additional resources/datasets that impact criminal justice system practices, this may be an option for a place for us to look for the results of studies using evidence based practices.

Katy Fitzgerald
Management Analyst

3rd Example – International Program in Survey and Data Science

survey-data-science.net



We are pleased to announce the launch of the International Program in Survey and Data Science (IPSDS). Fundamental changes in the nature of data, their availability, the way in which they are collected, integrated, and disseminated are a big challenge for all those working with designed data from surveys as well as organic data. IPSDS was developed in response to the increasing demand from researchers and practitioners for the appropriate methods and right tools to face these changes. We offer a multidisciplinary curriculum, world-class faculty, and a web-based learning environment that allows you to take courses from anywhere in the world.

Problem we tried to solve – In brief

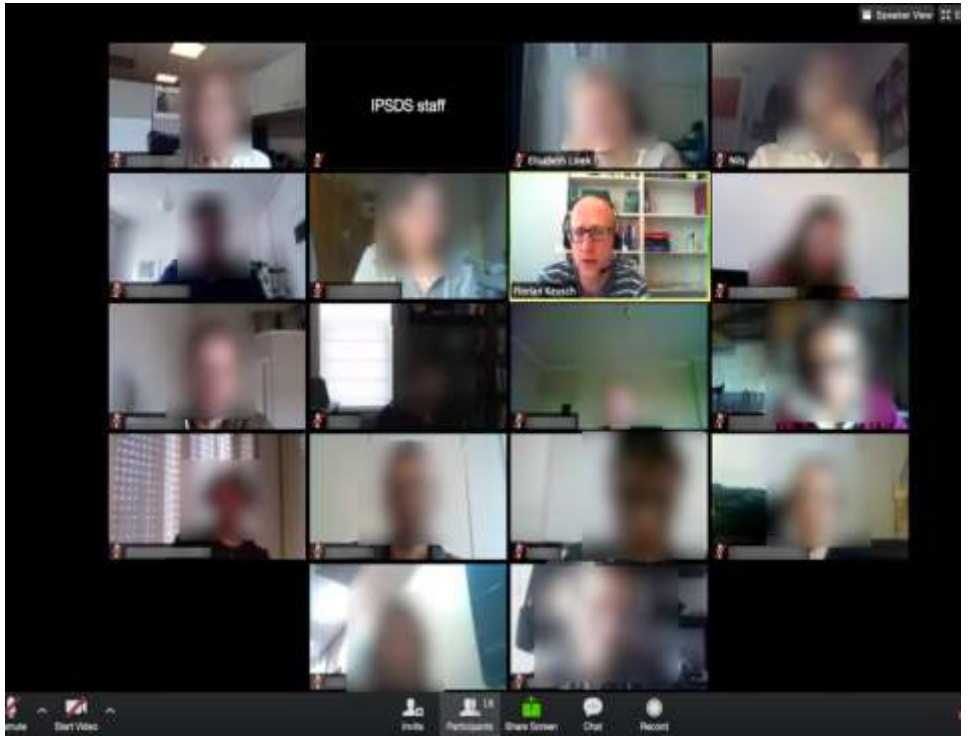
- Allow for multidisciplinary curriculum
- Modularized – adapt to prior skills and work needs
- Relevant methods and tools
- Mix of faculty from academia and industry

Key elements:

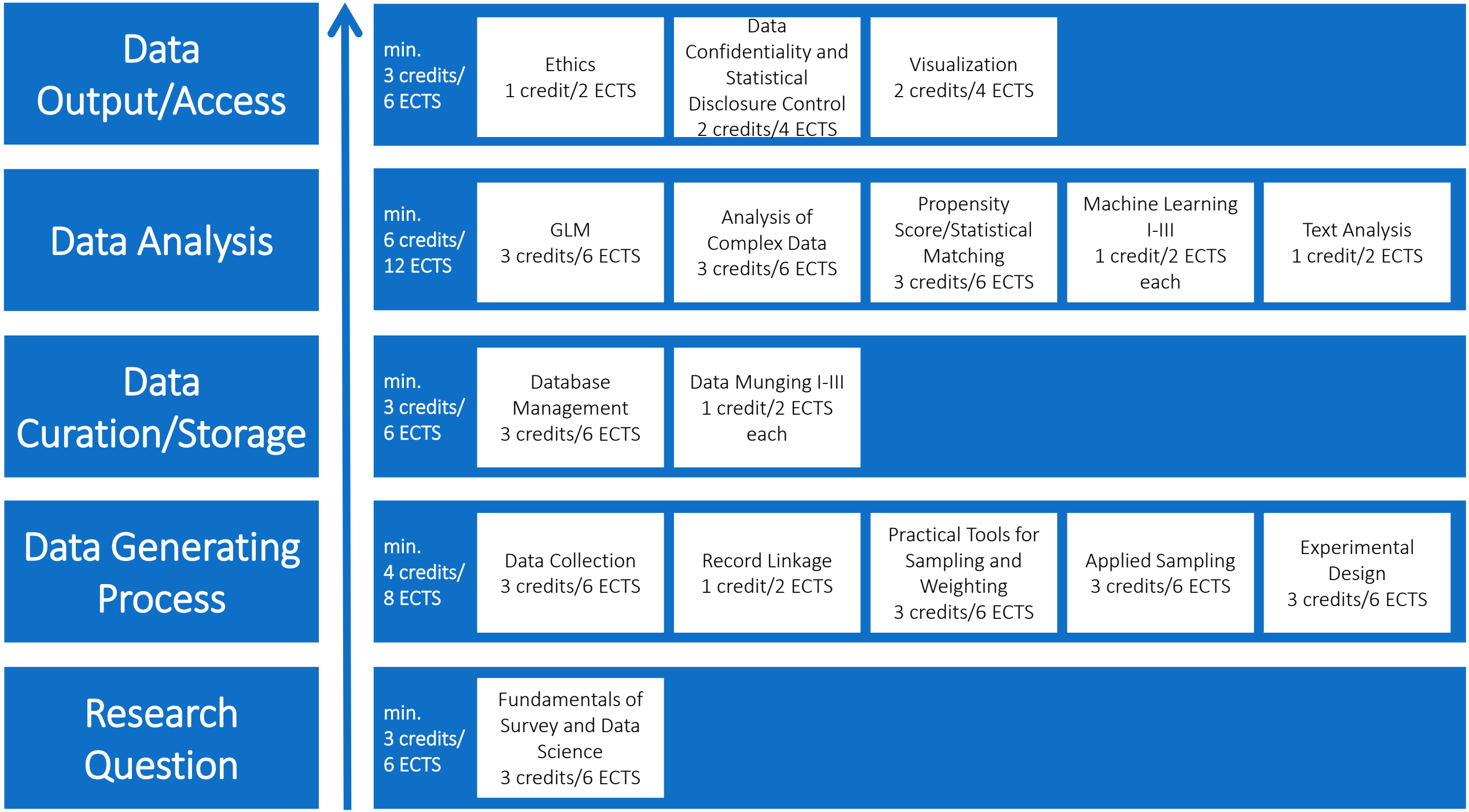
- Flexible web-based learning environment
- Live (video) interaction with faculty and students
- Face-to-face networking meetings

Mix of asynchronous and synchronous formats

- Pre-recorded lectures (small video units)
- (Bi)weekly assignments
- Discussion forums



- Small virtual classrooms
- Weekly 50-minute discussions with instructor
- Obligatory component


Data Output/Access
min. 3 credits/ 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. 6 credits/ 12 ECTS
GLM 3 credits/6 ECTS
Analysis of Complex Data 3 credits/6 ECTS
Propensity Score/Statistical Matching 3 credits/6 ECTS
Machine Learning I-III 1 credit/2 ECTS each
Text Analysis 1 credit/2 ECTS
Data Curation/Storage
min. 3 credits/ 6 ECTS
Database Management 3 credits/6 ECTS
Data Munging I-III 1 credit/2 ECTS each
Data Generating Process
min. 4 credits/ 8 ECTS
Data Collection 3 credits/6 ECTS
Record Linkage 1 credit/2 ECTS
Practical Tools for Sampling and Weighting 3 credits/6 ECTS
Applied Sampling 3 credits/6 ECTS
Experimental Design 3 credits/6 ECTS
Research Question
min. 3 credits/ 6 ECTS
Fundamentals of Survey and Data Science 3 credits/6 ECTS

Lessons Learned

- Learning with application at hand is key

- Teams can quickly overcome shortcomings
- Modular approach much appreciated by working professionals
- Privacy and confidentiality very important

- Hardest to learn and hardest to teach: Asking the right question!

https://ww2.amstat.org/education/datafest/

http://coleridgeinitiative.org

http://survey-data-science.net/

fkreuter@umd.edu

Contact me if you want to host locally or become a partner