



IAB-Workshop

Nutzerschulungen: Bedarf und Umsetzung bei Datenanbietern –
Beispiele aus der Praxis

12. April 2018

IPSDS project

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)

BE PART OF IT



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.

survey-data-science.net

Problems we tried to solve – in brief

Key elements:

- Multidisciplinary curriculum
- Modularized – adapt to prior skills and work needs
- Mix of faculty from academia and industry

Project coordinators and funding



SPONSORED BY THE



Federal Ministry
of Education
and Research



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

IPSDS curriculum

Problems we tried to solve – in brief

Key elements:

- Multidisciplinary curriculum
- Modularized – adapt to prior skills and work needs
- Mix of faculty from academia and industry

Data
Output/Access

min.
6 ECTS

Ethics
2 ECTS

Data
Confidentiality and
Statistical
Disclosure Control
4 ECTS

Visualization I-II
2 ECTS each

Data Analysis

min.
10 ECTS

Generalized Linear
Models
4 ECTS

Analysis of
Complex Data I-III
2 ECTS each

Propensity
Score/Statistical
Matching
4 ECTS

Machine Learning
I-III
2 ECTS each

Item Nonresponse
and Imputation
2 ECTS

Data
Curation/
Storage

min.
6 ECTS

Database
Management I-III
2 ECTS each

Data Munging I-III
2 ECTS each

Data Generating
Process

min.
10 ECTS

Data Collection
Courses
2 ECTS
each

Record Linkage
2 ECTS

Practical Tools for
Sampling and
Weighting
6 ECTS

Applied Sampling
I-III
2 ECTS
each

Experimental
Design
4 ECTS

Research
Question

min.
6 ECTS

Fundamentals of
Survey and Data
Science
6 ECTS

Total: 75 ECTS
Master Thesis: 15 ECTS

Master Thesis

IPSDS structure

Problems we tried to solve – in brief

Key elements:

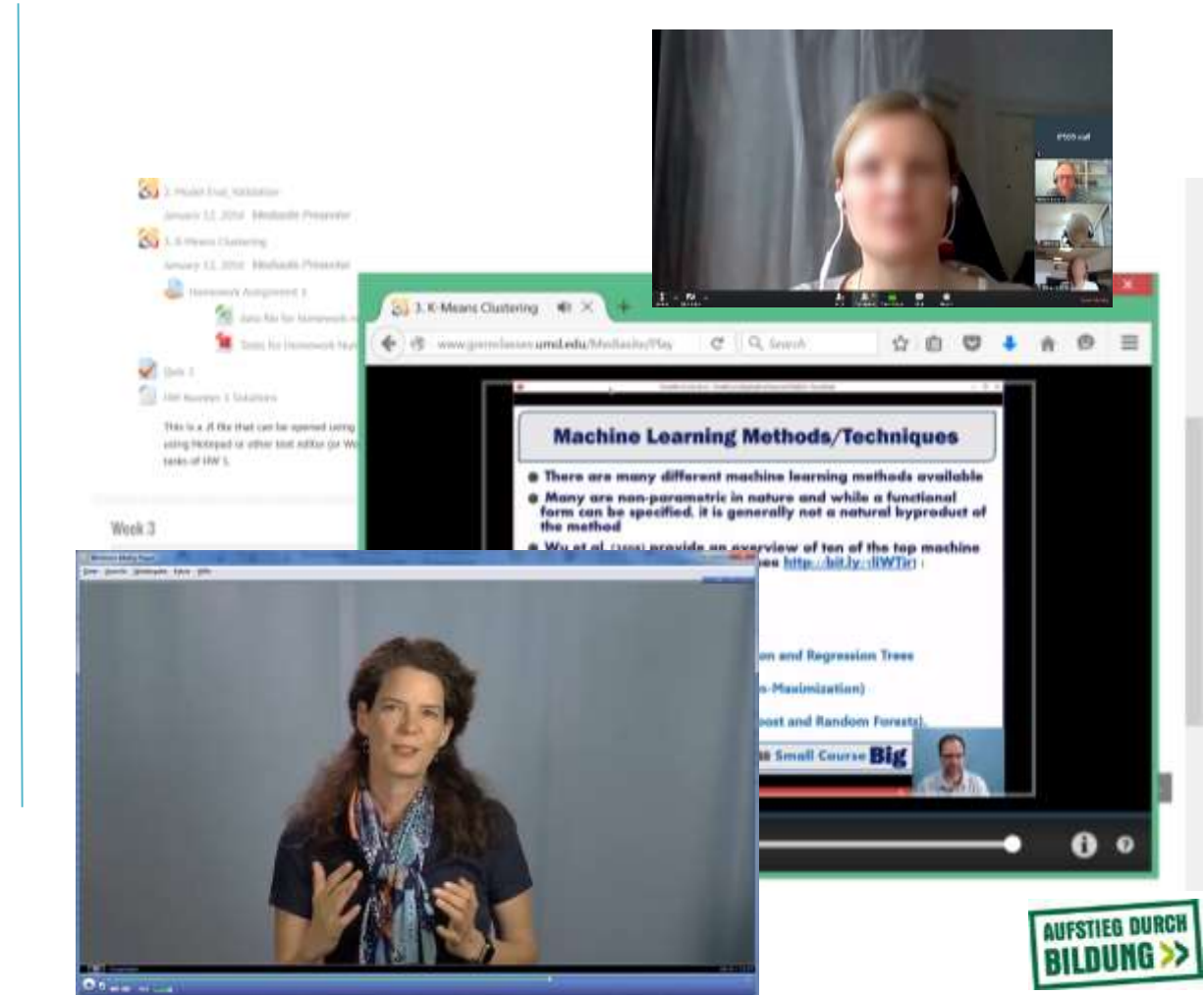
- Flexible web-based learning environment
- Live (video) interaction with faculty and students
- Face-to-face networking meetings

Onsite/Online

Onsite (Connect@IPSDS)



Online



Asynchronous/Synchronous

Asynchronous



- Pre-recorded lectures
- Readings/Other materials
- Assignments/Quizzes
- Discussion forums

Synchronous



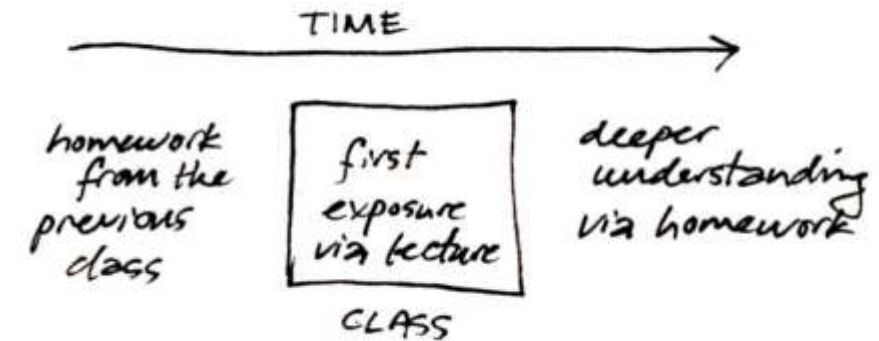
- Small virtual classrooms
- Weekly discussions led by the instructor
- Obligatory component

IPSDS courses

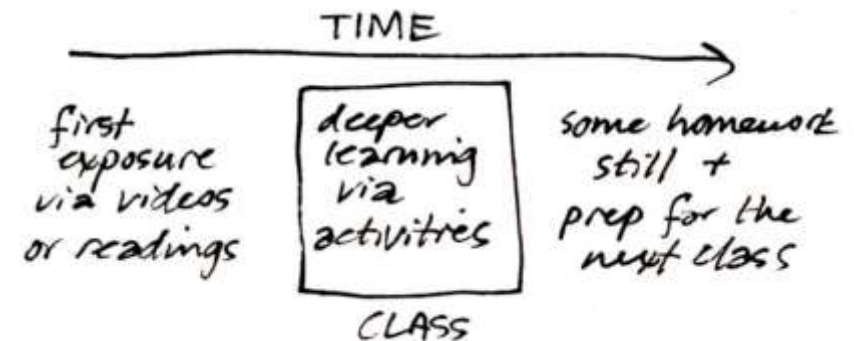
Courses

Advantages of Flipped Teaching

- more opportunities for interactivity in (online) discussions
- more personalized guidance
- more time for feedback
- deeper learning



Traditional classroom



Flipped classroom

Further Design Features

Pacing of the instruction

- Course length: 4 to 12 weeks (2 to 6 ECTS)
- Theme based units per calendar week
- Weekly assessment of topic/progress

Assessment

- Submit questions for online discussions
- Quiz
- Programming assignment
- Presentation
- Etc.

Feedback Mechanisms

- Automated feedback in quizzes
- Detailed from instructor for assignments and in online discussions
- Peer to peer feedback in some classes

Level of Interaction

- Weekly meetings
- Feedback for assignments
- Discussion forum and/or e-mails
- Group assignment

Welcome to SURV 736 !

This short course provides a condensed overview of web technologies and techniques to collect data from the web in an automated way. To this end, students will use the statistical software R. The course introduces fundamental parts of web architecture and data transmission on the web. Furthermore, students will learn how to scrape content from static and dynamic web pages and connect to APIs from popular web services. Finally, practical and ethical issues of web data collection are discussed.

Course and Learning Objectives

By the end of the course, students will...

- have an overview of state-of-the-art research that draws on web-based data collection,
- have a basic knowledge of web technologies,
- be able to assess the feasibility of conducting scraping projects in diverse settings,
- be able to scrape information from static and dynamic websites as well as web APIs using R, and
- be able to tackle current research questions with original data in their own work.

Before you start working on Unit 1, please make sure to thoroughly read the syllabus.

Mandatory weekly online meetings are Wednesdays, 12 PM (EST)/6 PM (CET). Please join the meeting via Zoom (directions below).

To join the weekly online meeting, go to <https://psm.zoom.us/> and select "Join". Enter the meeting ID: 9467 501955. If you are having connectivity issues with Zoom, see if a different browser (Explorer, Chrome, etc.) works better.

Moodle

- Course description/General Information
Topics covered, syllabus, additional resources
- New units auto-display each week.
Each unit includes:
 - Readings (Note reference to book chapter, URLs, PDFs)
 - Slides
 - Lecture videos
 - (Link to external resources)
 - (Additional material)
 - Zoom link for online meeting + date and time
 - Discussion forum for submitting questions/student-instructor interaction
 - Homework
 - Quiz (autograded)
 - Assignment submission (time restrictions)

Note: US Daylight Saving Time starts on March 11, 2018 and clocks are turned forward 1 h our. Daylight saving time begins in the EU on March 25, 2018. Therefore, look carefully at the time s of meetings and deadlines!

Slides for Unit 1

- 01-01 Introduction.pdf
- 01-02 Web-Scraping-Overview.pdf

4 problems to be solved w

- Due Tuesday, March 13
- Please submit your solu
- To generate the HTML f
- RStudio
- Here's a 2-minute YouTube tutorial on how to link RMarkdown files to HTML:

Quiz 1

Due Tuesday, March 13, 1 PM (EST)/6 PM (CET)

Place to post your questions

AUFSTIEG DURCH BILDUNG >>

Moodle

Welcome to SURV 736 !

This short course provides a condensed overview of web technologies and techniques to collect data from the web in an automated way. To this end, students will use the statistical software R. The course introduces fundamental parts of web architecture and data transmission. Static and dynamic web pages and common web data collection are discussed.

Course and Learning Objectives

By the end of the course, students will...

- have an overview of state-of-the-art
- have a basic knowledge of web technol
- be able to assess the feasibility of con
- be able to scrape information from st
- be able to tackle current research que

Before you start working on Unit 1, please read the following information. Mandatory weekly online meetings are Webinars (directions below).

To join the weekly online meeting, go to [Zoom Meeting](#). If you are having connectivity issues, please contact the instructor.

Syllabus- Spring 2018

Useful resources for R and web scraping

list-of-web-scraping-research

Unit 1: Introduction - Web Technologies

Join Zoom meeting on [Wednesday, 03/07/2018, 12:00 p.m.-12:50 p.m. EST/6:00 p.m.-6:50 p.m. CET]

Readings:

From textbooks-

- Munzert et al. (2015): Chapters Preface, 1, 2, 8

Video lectures (Simon Munzert): available online Wednesday, February 28

Note: US Daylight Saving Time starts on March 11, 2018 and clocks are turned forward 1 hour. Daylight saving time begins in the EU on March 25, 2018. Therefore, look carefully at the times of meetings and deadlines!

Slides for Unit 1

- 01-01-Introduction.pdf
- 01-02-Web-Scraping-Overview.pdf



- surv736-01-08-Summary
- surv736-01-case-study-br
- Assignment 1
- 4 problems to be solved w
- Due Tuesday, March 13
- Please submit your solu
- To generate the HTML f
- RStudio
- Here's a 2-minute YouTube tutorial on how to link R Markdown files to HTML!
- Quiz 1
- Due Tuesday, March 13, 1 PM (EST)/6 PM (CET)
- Place to post your questions

- Course description/General Information
Topics covered, syllabus, additional resources
- New units auto-display each week.
Each unit includes:
 - Readings (Note reference to book chapter, URLs, PDFs)
 - Slides
 - Lecture videos
 - (Link to external resources)
 - (Additional material)
 - Zoom link for online meeting + date and time
 - Discussion forum for submitting questions/student-instructor interaction
 - Homework
 - Quiz (autograded)
 - Assignment submission (time restrictions)



Welcome to SURV 736 !

This short course provides a condensed overview of web technologies and techniques to collect data from the web in an automated way. To this end, students will use the statistical software R. The course introduces fundamental parts of web architecture and data transmission. Static and dynamic web pages and common web data collection are discussed.

Course and Learning Objectives

By the end of the course, students will...

- have an overview of state-of-the-art
- have a basic knowledge of web techn
- be able to assess the feasibility of con
- be able to scrape information from st
- be able to tackle current research que

Before you start working on Unit 1, please read the following instructions. Mandatory weekly online meetings are Webinars (directions below).

To join the weekly online meeting, go to [Zoom Meeting](#). If you are having connectivity issues, please contact the IT helpdesk.

- Syllabus- Spring 2018
- Useful resources for R and web scraping
- list-of-web-scraping-research

Unit 1: Introduction – Web Technologies

Join Zoom meeting on [Wednesday, 03/07/2018, 12:00 p.m.-12:50 p.m. EST/6:00 p.m.-6:50 p.m. CET]

Readings:

From textbooks-

- Murzert et al. (2015): Chapters Preface, 1, 2, 8

Video lectures (Simon Murzert): available online Wednesday, February 28

18 and clocks are turned forward 1 h. Therefore, look carefully at the time

surv736-01-08-Summary

surv736-01-case-study-breweries

Assignment 1

4 problems to be solved with R.

- Due Tuesday, March 13, 1 PM (EST)/6 PM (CET)
- Please submit your solution as an HTML file
- To generate the HTML file, use the Rmd file, enter your solutions, and click on "Knit to HTML" in RStudio
- Here's a 2-minute YouTube tutorial on how to knit Rmd files to HTML:



Quiz 1

Due Tuesday, March 13, 1 PM (EST)/6 PM (CET)

Place to post your questions

Moodle

- Course description/General Information
Topics covered, syllabus, additional resources
- New units auto-display each week.
Each unit includes:
 - Readings (Note reference to book chapter, URLs, PDFs)
 - Slides
 - Lecture videos
(Link to external resources)
(Additional material)
 - Zoom link for online meeting + date and time
 - Discussion forum for submitting questions/student-instructor interaction
 - Homework
 - Quiz (autograded)
 - Assignment submission (time restrictions)



Videos

- Lectures, interviews and discussions with experts, demonstrations of specific techniques and software tools
- Lectures are broken into easily-digestible sessions to help students to better focus on the material.
- Students engage with the material at their own pace: e.g., replay parts that cover difficult concepts

The image is a screenshot of a video player. The main content is a survey form titled "Question with visual aids". The form includes the following text:

Question with visual aids

VIII. Neighborhood Networks:

INTERVIEWER SHOW CARD D TO RESPONDENT.

Now I would like to ask you some questions about the people who live in this neighborhood. For the first set of questions, please use the options on CARD D when giving your answers.

INTERVIEWER: WRITE IN THE RATING NUMBER THAT CORRESPONDS TO THE RESPONDENT'S ANSWER.

NONE	1-2	SEVERAL	MANY	ALMOST EVERYONE
1	2	3	4	5

___ 1. First, how many people in your neighborhood would you recognize?

RESIDENT SURVEY OF NEIGHBORHOOD CONDITIONS. <http://povertycenter.corn.edu>

At the bottom right of the video player, there is a small inset video showing a woman with glasses and a yellow scarf.

Virtual Classrooms

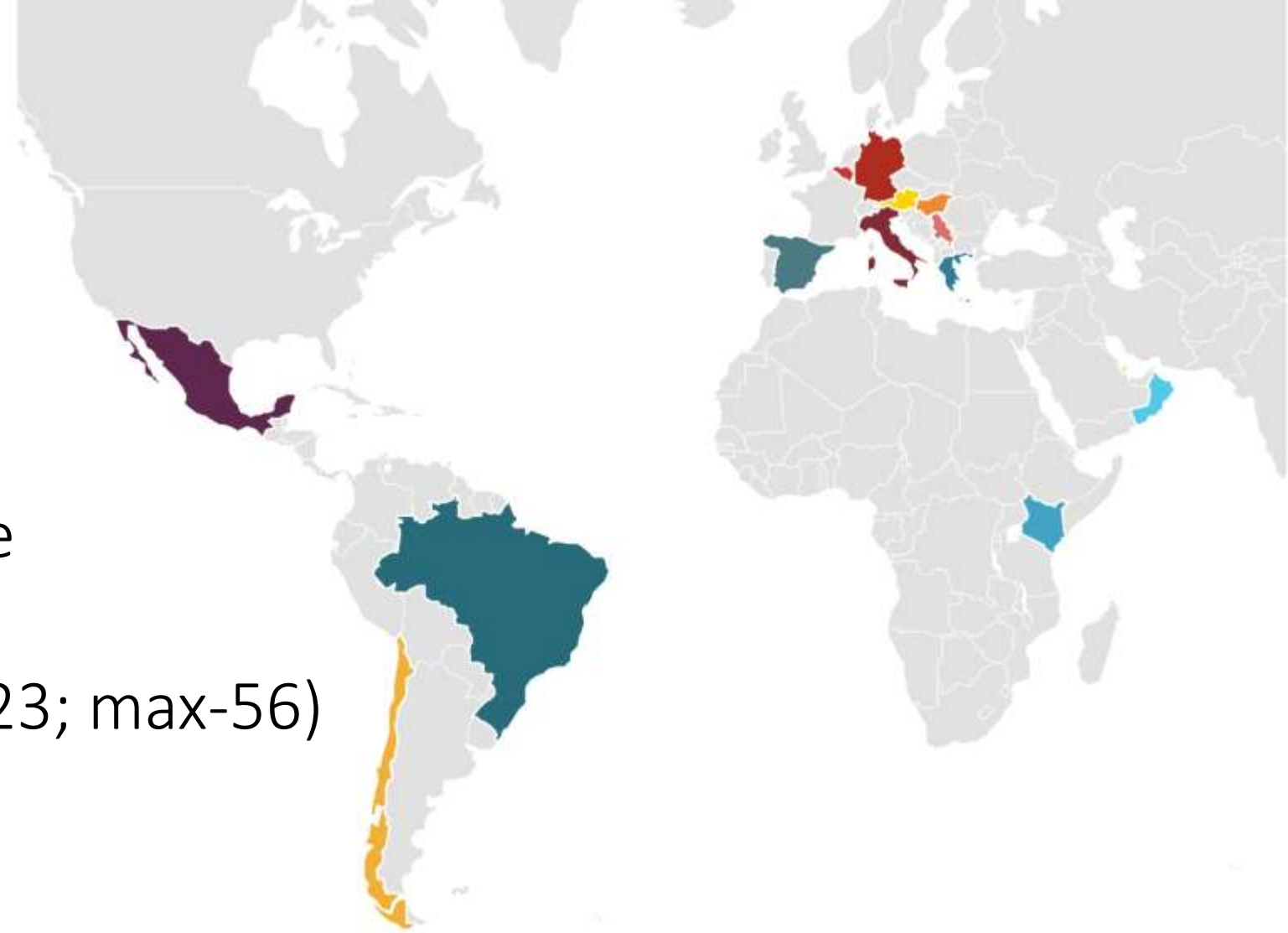
- Weekly mandatory online meetings (50 minutes)
- Discuss students' questions
- Review problems with assignments
- Collaborative problem solving
- Motivate students to persist in the course
- Break out rooms, (private and public) chats, polls ...



IPSDS students

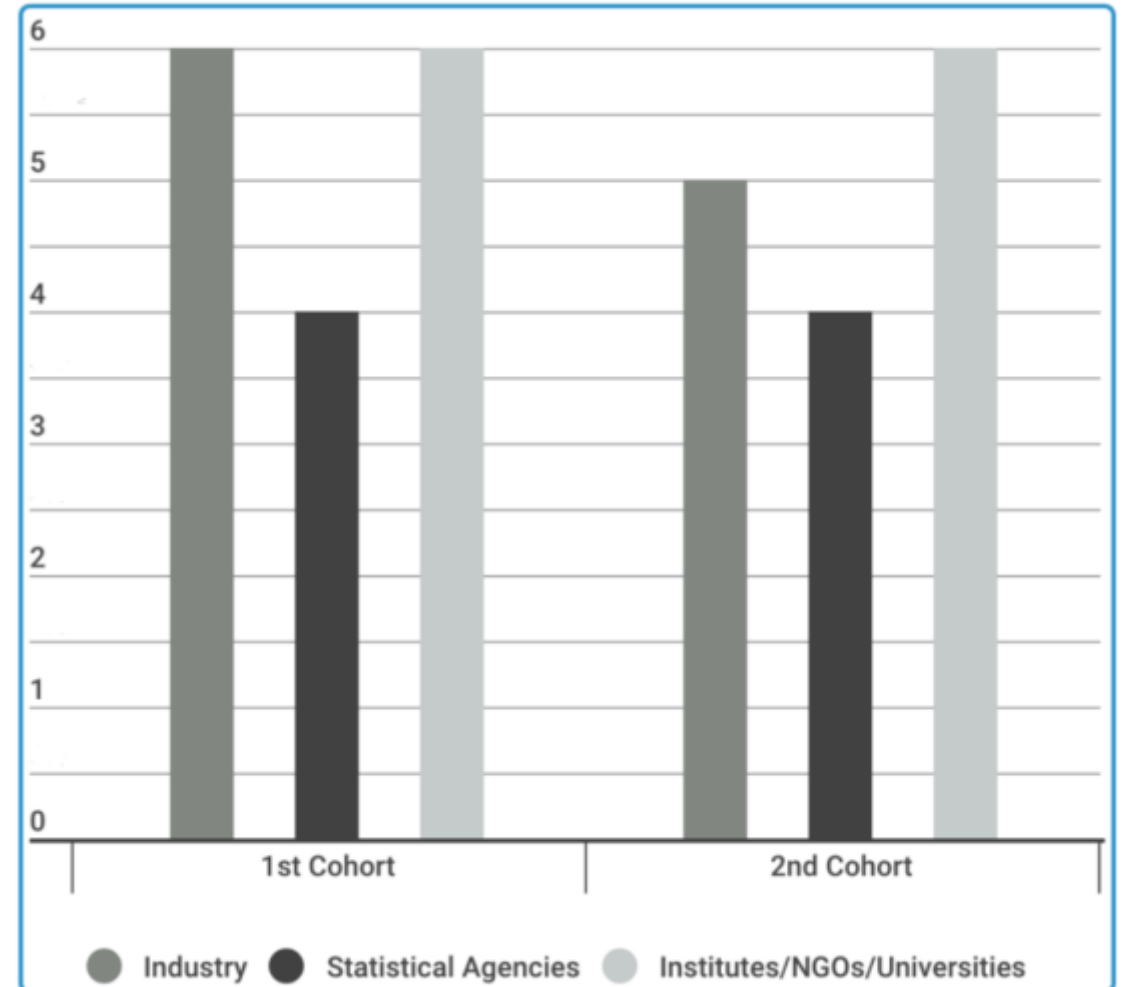
2 Test Cohorts

- 31 Participants
(18 f + 13 m)
- 15 countries of residence
- Age: median=30,5 (min-23; max-56)



Test Cohorts 1 & 2

- 22 (71 %) M.A./M.S.
- 41 hours/week
(Min-10; Max-55)
- 8 students with family duties



IPSDS evaluations & studies

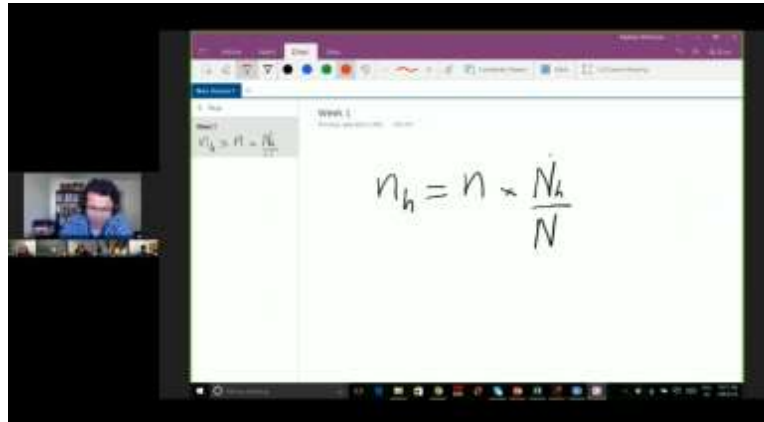
Evaluations

- Start of the program/yearly milestone survey/qualitative interviews of students
- Post-course student survey
- Learning analytics
- Qualitative interviews with instructors
- Pilot studies
- Curriculum and courses review

BMBF-Study: Online Communication

Synchronous Communication

Fundamentals of Survey and Data Sc.



- 11 online discussions (mandatory)
- 16 students

- 12 weeks/6 ECTS
- Post questions
- Answer/comment questions of other students

Asynchronous Communication

Data Collections Methods

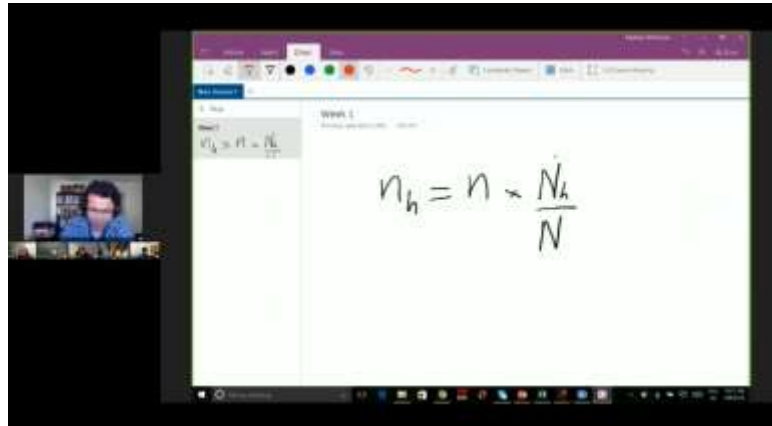


- 1 introductory online meeting
- Discussion forums
- 15 students

BMBF-Study: Online Communication

Synchronous Communication

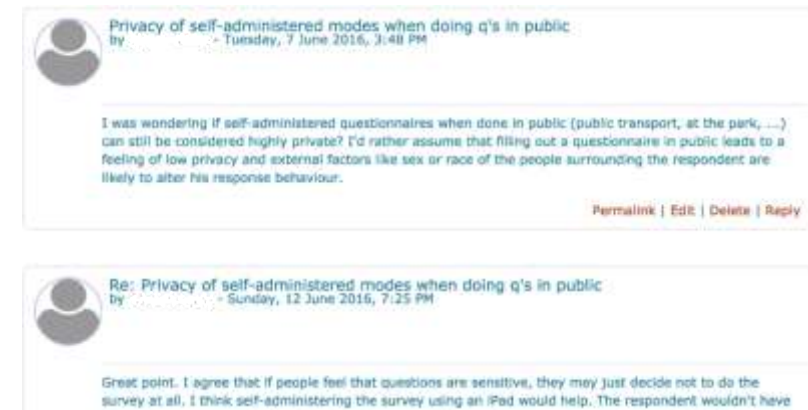
Fundamentals of Survey and Data Sc.



- Stronger sense of community and belonging
- Immediate feedback
- All students successfully finished the course

Asynchronous Communication

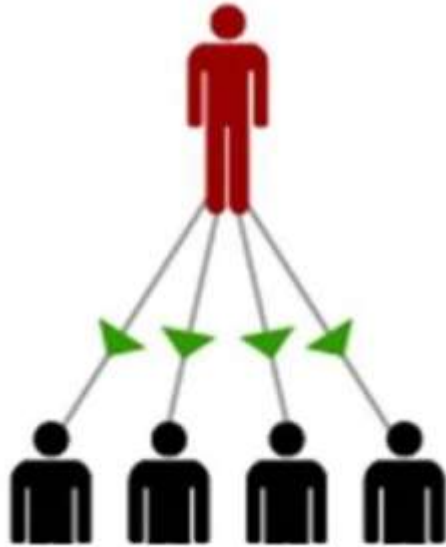
Data Collections Methods



- Greater flexibility
- More time for reflection
- Less workload
- 2 drop outs

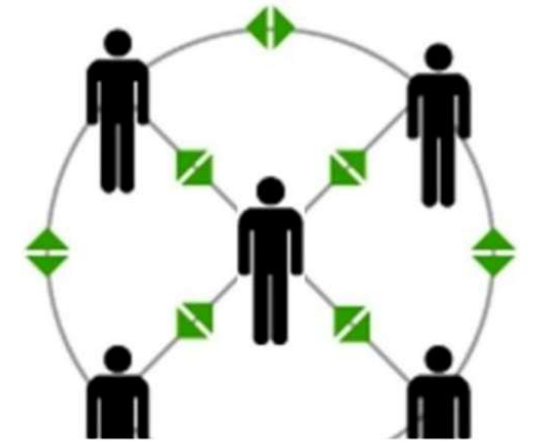
BMBF-Study: Flexibility

Instructor-Paced Format



- Weekly online meetings
- Set dates for all assignments and final exam

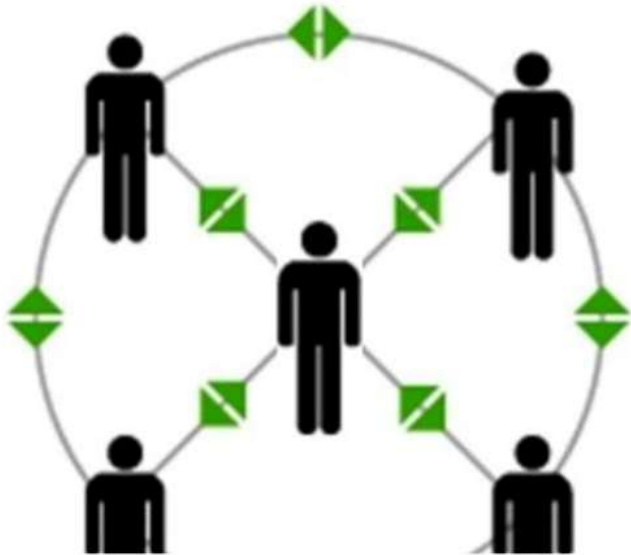
Self-Paced Format



- 4 ECTS/ 8 + 1 weeks
- 8 participants
- 1 introductory online meeting
- Only 1 deadline: all assignments and final exam due in the ninth week

BMBF-Study: Flexibility

Self-Paced Format



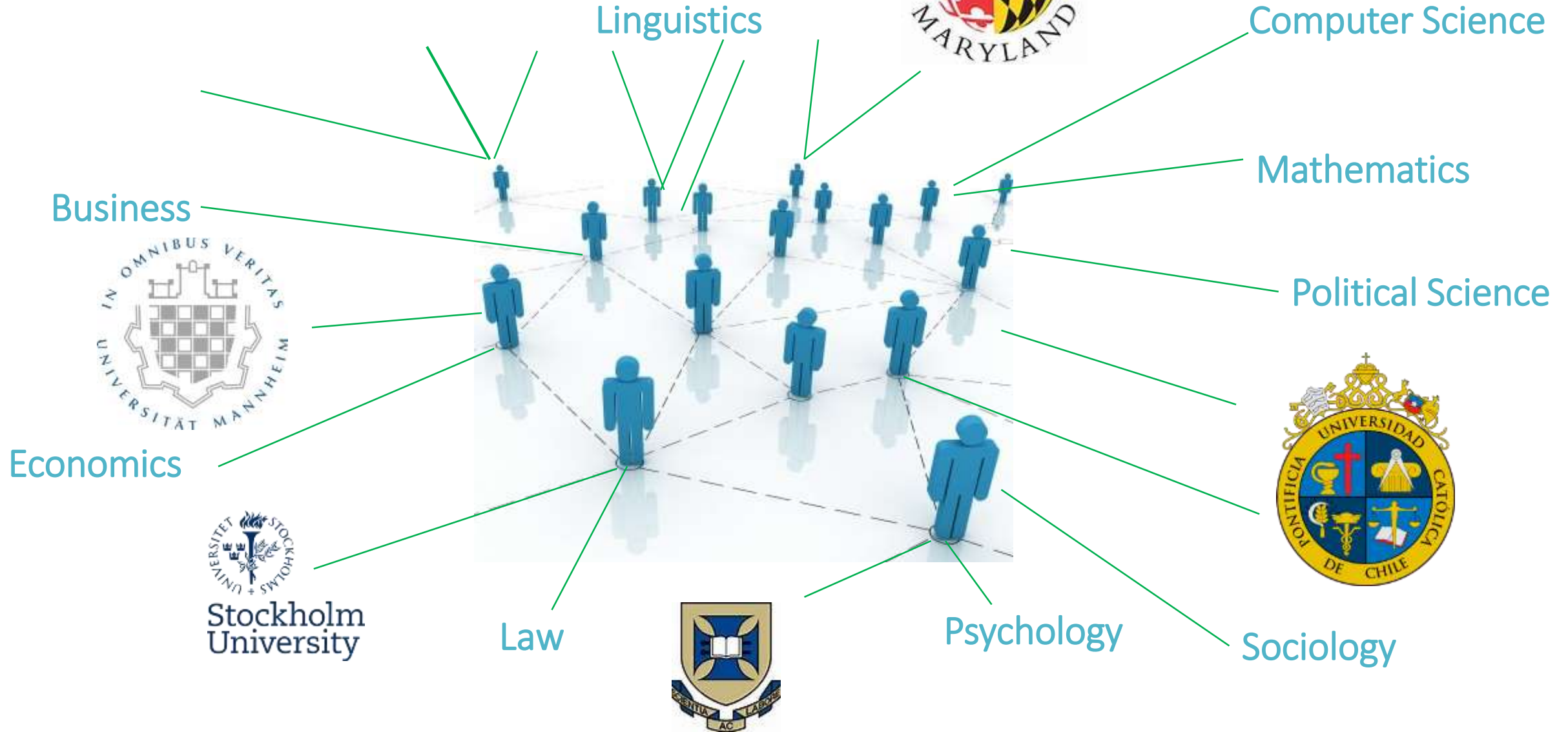
- 4 drop outs
- Flexibility with deadlines for assignments appreciated
- Biggest challenge: time management

IPSDS lessons learning

- Modular approach much appreciated by working professionals
- Feasibility of combining studies with work and family (6 ECTS/term, 3 drop outs in 2 years)
- Biggest challenge: workload management (average workload: 10 hours/week)
- Balancing flexibility and consistency (time management, self-discipline)
- Workplace orientation:
 - Hands-on application and working with data is key
 - Project-oriented courses/working on real-world projects

your QUESTIONS

#collaborations



Partner

Universities

- University of Maryland
- University of Mannheim

- Catholic University of Santiago de Chile
- Australian National University
- Beijing University
- Ashoka University (expressed interest)
- University of Capetown (planned)

Others

- SRO - Michigan
- PEW
- German Record Linkage Center
- GESIS
- Bureau of Labour Statistics
- U.S. Census Bureau
- Statistics Netherlands

Was haben wir gelernt?

Durchschnittlich 6 ECTS/Term

IPSDS ermöglicht Weiterbildung und Familie zu koordinieren:

- *7 Teilnehmer/innen mit Kindern*
- *3 Geburten während der ersten Förderphase*

Kurs-Design berücksichtigt Motivationsförderung:

- *synchrones Lernen*
- *Connect@IPSDS*
- *fest terminierter Kursrhythmus*

Nur 3 Studienabbrecher (nach 1. Jahr)



IPSDS faculty

✓ **Manfred Antoni**

(German Institute for Employment Research)

✓ **Stefan Bender**

(Deutsche Bundesbank)

✓ **Trent Buskirk**

(Vice President at Statistics and Methodology at Marketing Systems Group/University of Massachusetts Boston)

✓ **Mario Callegaro**

(Google UK)

✓ **Frederick Conrad**

(University of Maryland/University of Michigan)

✓ **Jill Dever**

(RTI)

✓ **Jörg Drechsler**

(German Institute for Employment Research/LMU Munich)

✓ **Jeffrey Gonzalez**

(Bureau of Labor Statistics OSMR)

✓ **Steven Heeringa**

(University of Maryland/University of Michigan)

✓ **Matt Jans**

(UCLA Center for Health Policy Research/University of Maryland/University of Connecticut)

✓ **James Lepkowski**

(University of Maryland/University of Michigan)

✓ **Daniel Oberski**

(Utrecht University)

✓ **Louis Rizzo**

(Westat)

✓ **Jennifer Romano-Bergstrom**

(Facebook)

✓ **Joseph Sakshaug**

(University of Michigan/German Institute for Employment Research)

✓ **Richard Valliant**

(University of Maryland/ University of Michigan)

Team

Frauke
Kreuter



Karin
Frößinger



Daria
Korfant

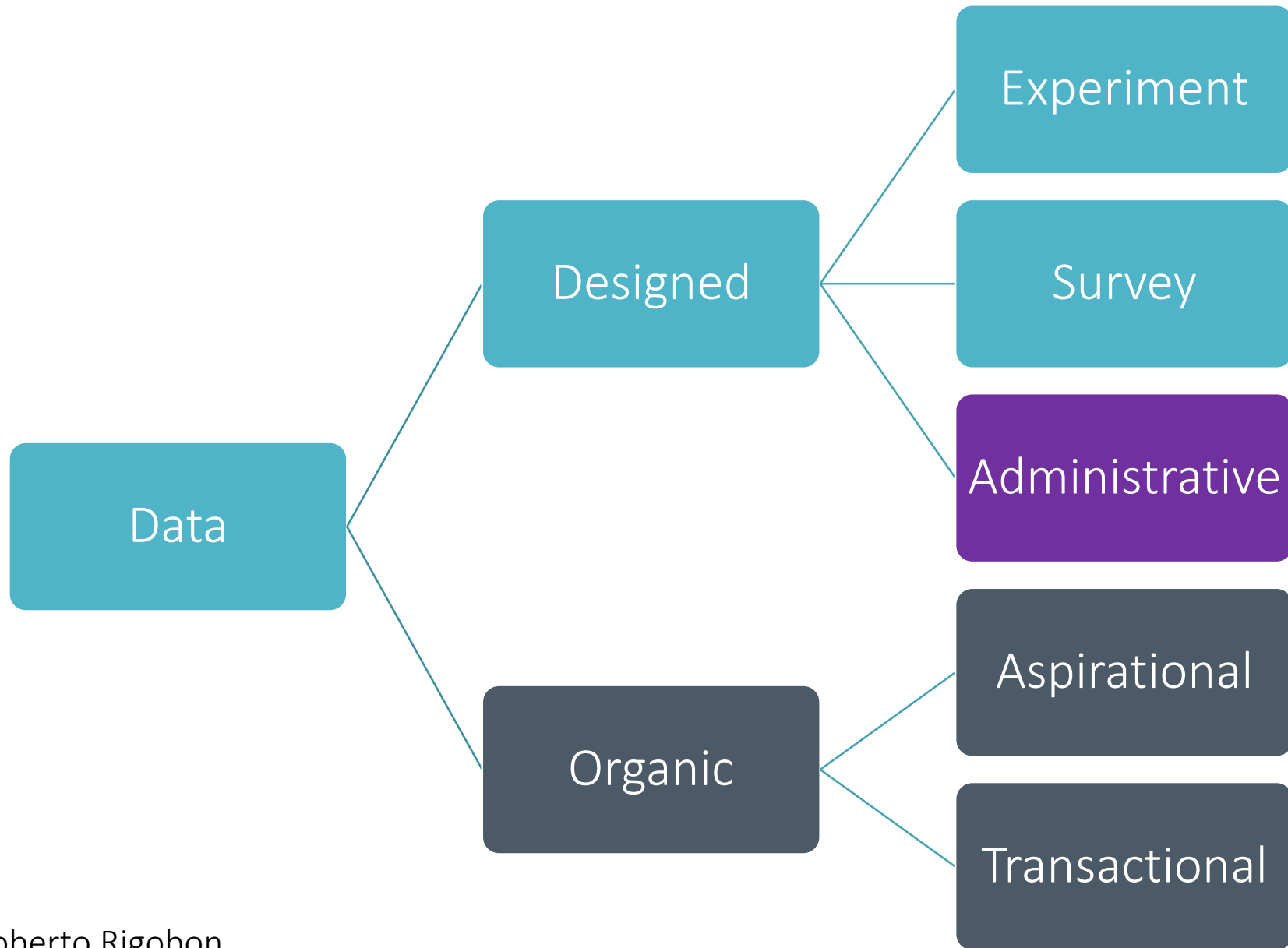


Evgenia
Samoilova



Florian
Keusch





Source: Roberto Rigobon

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 enables 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

Modules

Data Output/Access

Learn how to communicate results, distribute and store your data; Ethics

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 processes.

Research Question

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