



Dr. Lars Pelke

Universität Heidelberg, Institut für Politische Wissenschaft

Praxiskurs Datenanalyse und Replikation

Sitzung 2 Was ist replizierbare sozialwissenschaftliche
Datenanalyse?

Leitfragen und Lernziele



- Was ist gute quantitative Forschung?
- Wie empirische Analysen anderer Forschender nachvollziehen und replizieren?
- Was verstehen wir unter Replikation (Stichworte: Reproducibility vs. Replication; duplication study; replication study)
- Wie planen wir unsere Replikationsstudie? Wie arbeiten wir selber reproduzierbar?

Was ist gute quantitative empirische Forschung?



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Gruppenarbeit (ca. 15min)

Tragen Sie gemeinsam in Ihrer Kleingruppe zusammen, was für Sie „gute“ quantitative empirische Forschung ist? Woran erkennen Sie solche Forschung. Halten Sie Ihre Ergebnisse auf dem Flipchart fest

Was ist gute quantitative empirische Forschung?



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Plenumsbesprechung

Eine:r von Ihnen stellt Ihre Ergebnisse mit Hilfe des Flipcharts vor. Sie sind alle dazu aufgerufen Rückfragen zu stellen und zu diskutieren.

Wie empirische Analysen anderer Forschender nachvollziehen und replizieren?



Erster Schritt:

- Ergebnisse (Tabellen und Abbildungen) aus dem Originalartikel extrahieren.
- Methodisches Vorgehen detailliert nachvollziehen mit Blick in die Anhänge

Zweiter Schritt:

- Replikationsfiles der Originalstudie herunterladen und alle dazugehörigen Datensätze herunterladen
- Erster Blick in die Skripte. Verstehe ich was die Autor:innen wie gemacht und gespeichert haben. Gibt es eine klare Zuordnung der Ergebnisse zu den Abbildungen und Tabellen im Paper?

Wie empirische Analysen anderer Forscher nachvollziehen und replizieren?



Kurzer Exkurs: Wenn die Daten/Skripte nicht vorliegen

- Bevor Sie Autor:innen anschreiben, überprüfen Sie die Website der Autor:innen, Dataverse Network, ICPSR Archive und die Journalwebsite
- Wenn die Replikationsmaterialien nicht verfügbar sind, verfassen Sie eine kurze und prägnante Email. Beachten Sie dabei, dass Sie freundlich und auf Autor:innen zugehen und erwarten Sie keine kurzfristige Antwort

Beispiel-Email

Dear Dr/Prof ____,

I am a graduate student at Heidelberg University, working on the relationship between autocracy and social service provision. I read your article with XY (2009) and our article with Z (2020) with great interest. I find especially the empirical analysis in our article with XY (2009) very illuminating.

I am writing to you today because I would like to inquire if you would provide your data sets of the article for replication? This would be immensely helpful for me, not just for learning purposes, but also to check my own data reliability, and to decide which variables to include or drop.

I would be very grateful for your help. Thank you very much in advance!

Best, A

Wie empirische Analysen anderer Forscher nachvollziehen und replizieren?

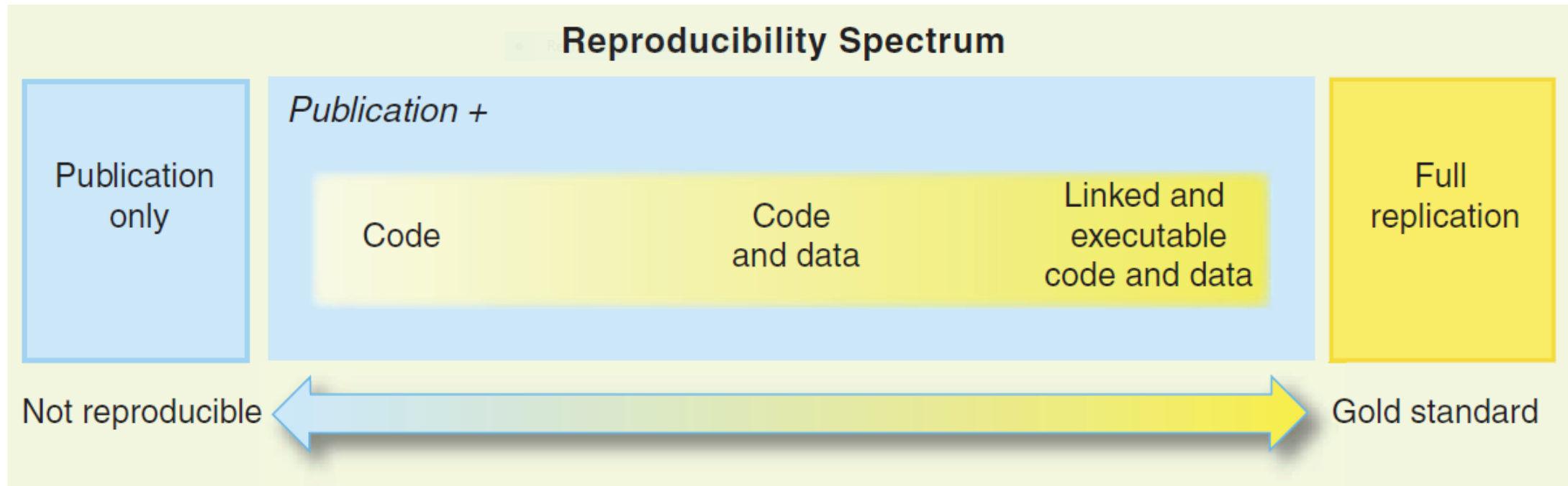


Fig. 1. The spectrum of reproducibility.

Quelle: Peng 2011: Reproducible Research in Computational Science. Science: 334. December 2011.

Wie empirische Analysen anderer Forscher nachvollziehen und replizieren?

Struktur Replikationsmaterialien



tion > Paper > Redistribution and Taxes in Autocracies > replication files

Name	Änderungsdatum	Typ	Größe
calculations	17.06.2020 13:06	Dateiordner	
data	17.06.2020 13:16	Dateiordner	
output	17.06.2020 14:07	Dateiordner	
.Rhistory	30.06.2020 08:00	RHISTORY-Datei	15 KB
README.txt	17.06.2020 14:12	Textdokument	3 KB
Table1.doc	18.06.2020 15:11	Microsoft Word 9...	10 KB

tion > Paper > Redistribution and Taxes in Autocracies > replication files > data

Name	Änderungsdatum	Typ	Größe
cow	17.06.2020 13:06	Dateiordner	
gwf	17.06.2020 13:07	Dateiordner	
Solt 2019 SWIID 8.0	17.06.2020 13:07	Dateiordner	
Solt 2019 SWIID 8.2	17.06.2020 13:07	Dateiordner	
sp_data	17.06.2020 13:07	Dateiordner	
SPAW Rasmussen 2016	17.06.2020 13:07	Dateiordner	
vdem_9	17.06.2020 13:07	Dateiordner	
world_bank_data	17.06.2020 13:07	Dateiordner	
vdem_merged.rds	17.06.2020 13:16	RDS-Datei	27.125 KB
vdem_merged_spaw.rds	17.06.2020 13:18	RDS-Datei	27.141 KB
vdem_social.rds	17.06.2020 13:54	RDS-Datei	27.323 KB

tion > Paper > Redistribution and Taxes in Autocracies > replication files > calculations

Name	Änderungsdatum	Typ	Größe
.Rhistory	29.04.2021 09:13	RHISTORY-Datei	0 KB
01_data wrangling.R	17.06.2020 13:33	R-Datei	8 KB
02_main_models.R	19.06.2020 09:01	R-Datei	47 KB
03_robustness.R	18.06.2020 15:55	R-Datei	141 KB
04_Figures.R	18.06.2020 15:36	R-Datei	14 KB

tion > Paper > Redistribution and Taxes in Autocracies > replication files > output > margins

Name	Änderungsdatum	Typ	Größe
Figure1_main.pdf	19.06.2020 08:53	Adobe Acrobat D...	7 KB
Figure1_main.png	19.06.2020 08:54	PNG-Datei	417 KB
Figure1_main_400dpi.png	19.06.2020 08:54	PNG-Datei	89 KB
Figure1_main_600dpi.png	19.06.2020 08:54	PNG-Datei	155 KB
Figure1_SA.pdf	19.06.2020 08:57	Adobe Acrobat D...	7 KB
Figure1_SA.png	19.06.2020 08:57	PNG-Datei	452 KB
Figure2_main.pdf	19.06.2020 08:58	Adobe Acrobat D...	6 KB
Figure2_main.png	19.06.2020 08:58	PNG-Datei	699 KB
Figure2_main_400dpi.png	19.06.2020 08:59	PNG-Datei	166 KB
Figure2_main_600dpi.png	19.06.2020 08:59	PNG-Datei	309 KB
Figure2_SA.pdf	18.06.2020 15:17	Adobe Acrobat D...	7 KB
Figure2_SA.png	18.06.2020 15:17	PNG-Datei	720 KB
Figure3_main.pdf	19.06.2020 09:02	Adobe Acrobat D...	11 KB
Figure3_main.png	19.06.2020 09:02	PNG-Datei	1,079 KB
Figure3_main_400dpi.png	19.06.2020 09:02	PNG-Datei	245 KB
Figure3_main_600dpi.png	19.06.2020 09:02	PNG-Datei	409 KB
Figure3_SA.pdf	18.06.2020 15:33	Adobe Acrobat D...	10 KB
Figure3_SA.png	18.06.2020 15:33	PNG-Datei	1,059 KB
S1_Figure_1.pdf	17.06.2020 13:38	Adobe Acrobat D...	7 KB
S1_Figure_2.pdf	17.06.2020 13:38	Adobe Acrobat D...	6 KB
S2_Figure_1.pdf	18.06.2020 15:55	Adobe Acrobat D...	11 KB
S2_Figure_2.pdf	18.06.2020 15:55	Adobe Acrobat D...	9 KB
S3_Figure_1.pdf	17.06.2020 13:56	Adobe Acrobat D...	7 KB
S3_Figure_1-1.png	22.04.2021 10:42	PNG-Datei	120 KB
S3_Figure_2.pdf	17.06.2020 13:56	Adobe Acrobat D...	6 KB

Wie empirische Analysen anderer Forscher nachvollziehen und replizieren?



Struktur Skripte (Beispiele)

Kriterien: klar
kommentiert, Outline
Struktur

```
Source on Save
#### "Inclusionary regimes, party institutionalization and redistribution under authoritarianism" ####
# authors: "Lars Pelke"
# date: 2020-06-16
# written under "R version 3.6.1 (2019-07-05)"

#### Preliminaries ####

R.version$version.string

# clear workspace
rm(list=ls())

#libraries
library(tidyverse)
library(ggpubr)
library(readstata13)
library(texreg)
library(dotwhisker)
library(broom.mixed)
library(countrycode)
library(imputeTS)
library(readstata13)

# set working directory

# please use the working directory, where you stored the zip-file.

#### Load Data ####

## V-DEM
vdem <- read_csv("data/vdem_9/V-Dem-CY-Full+Others-v9.csv")
```

Wie empirische Analysen anderer Forscher nachvollziehen und replizieren?



Struktur Skripte (Beispiele)

Kriterien: klar
kommentiert,
Outline Struktur

```
# set working directory
# please use the working directory, where you stored the zip-file.

#####
#####

#### Outline ####

## 1. Small sample based on solt 2019 original data
## 2. REWB with democratic and autocratic countries
## 3. Income Taxation Data
## 4. Social Spending Data
## 5. Social Protection Data
## 6. Public Goods Provision V-Dem Data
## 7. OLS Models relative redistribution

#####
#####

#### APPENDIX B ####
#### 1. Small sample based on solt 2019 original data ####

## Load Data ##

## V-DEM
vdem <- read_csv("data/vdem_9/V-Dem-CY-Full+Others-v9.csv")

vdem$cown <- countrycode(vdem$country_name, "country.name", "cown", warn = TRUE)
vdem$cown[vdem$country_name == "Hong Kong"] <- "715"
vdem$cown[vdem$country_name == "Republic of Vietnam"] <- "817"
vdem$cown <- as.integer(vdem$cown)
```

Was verstehen wir unter Replikation?

Terminologie



- **Reproducibility**

- = *Reproduzierbarkeit von Ergebnissen*
- kann durch Autor:innen verbessert werden, wenn diese folgende Materialien bereitstellen: **Datenzugang**, Details zur **Datenerhebung** und –**verarbeitung**, Details zur **Analyse** der zusammengetragenen Daten, die zu den Ergebnissen geführt haben (siehe auch Lupia and Elman 2014, APSA Guidelines)

- **Duplication**

- = *Duplikation der Originalresultate* einer Studie verifiziert die Ergebnisse indem mit den gleichen Daten und den gleichen Methoden die Originalresultate versucht werden zu duplizieren.
- wir starten mit der Duplikation, um darauf aufbauend Replikationsstudien durchzuführen
- Synonyme: direkte Replikation

Was verstehen wir unter Replikation?

Terminologie



- **Replication**

- Ziel von Replikationsstudien geht über die bloße Duplikation der Originalresultate hinaus.
- Replikation testet typischerweise die Robustheit der Originalbefunde mit Hilfe neuer Daten, neuer Variablen, und/oder anderer statistischer Spezifikationen/Methoden
- Zeithorizont: Datensatzerstellung und Datenbereinigung sind zeitaufwendig, daher zeitaufwendiger als bloße Duplikation
- Synonym: Konzeptionelle Replikation

Was verstehen wir unter Replikation?

Duplication Checklist



1. Herunterladen der Replikationsdateien von der Journalwebsite / Repository (bspw. Harvard Dataverse) und Originalskripte wie beschrieben ausführen. Im Idealfall sollten die Softwareskripte einfach ausführbar sein und keine Fehlermeldungen produzieren.
 2. Alle empirische Modelle durchlaufen lassen, genau so wie diese im Artikel spezifiziert worden sind. Tabellen und Abbildungen reproduzieren.
 3. Vergleichen der Originalresultate mit den Duplikationsergebnissen (händisch, oder besser: mit Hilfe des Einlesens der Originalergebnisse in die Software)
Im Fall von Abweichungen: Diskutieren von potentiellen Fehlern in den Daten, Fehlern in der Datentransformation und Modellspezifikationen. Fehler in den Softwareskripten durch die Originalautoren oder einen selbst?
-

Was verstehen wir unter Replikation?

Replication Checklist



1. Starten wie bei der Duplikation von Ergebnissen mit dem Herunterladen aller Daten und Skripte. Die Originalskripte und Daten spezifizieren genau, was die Autor:innen empirisch Schritt für Schritt ausgeführt haben. Guter Ausgangspunkt zum Aufspüren von Coding Errors, fehlerhaften Skripte, oder fehlerhafter Variablen (bspw. Country IDs als numerische Variablen in ein Model einbeziehen)
2. Der zweite Schritt umfasst typischerweise das Sammeln neuer Daten (andere Datensätze, Ergänzung von weiteren Beobachtungspunkten). Andere Datensätze können häufig zu einer verbesserten Messung eines Konzepts führen oder weitere Variablen enthalten die potentielle confounder sind.
Repository
3. Analyse der Originalmodellen mit neuen Variablen, neuen Daten. Wie verändern diese Modifikationen die Ergebnisse im Vergleich zur Originalstudie?
4. Veränderung der Modellspezifikationen um die Daten oder die Theorie besser zu modellieren (bspw. zusätzliche time lags, Umgang mit missings; Umgang mit Kausalitätsproblemen etc.)
5. Diskussion der Implikationen der neuen Daten, neu eingeführten Variablen/Interaktionseffekte oder neuer Modellspezifikationen, die ggf. die Veränderungen in den Resultaten hervorrufen.
6. Gold Standard Replikationsstudie: Transparente und nachvollziehbare Dokumentation alle Analyseschritte; Veröffentlichung aller Daten und Softwareskripte auf einem dauerhaften zugänglich Repository. Erwäge ferner die Prä-Registrierung der Analyse!

Wie planen wir unsere Replikationsstudie? Wie arbeiten wir selber reproduzierbar?



5 Minute Paper

Schreiben Sie mit Hilfe der bisher diskutierten Punkten zu folgender Frage ein kurzes 5 Minute Paper (kann auch Stichpunkte umfassen):

Wie planen wir unsere Replikationsstudie?

Wie planen wir unsere Replikationsstudie? Wie arbeiten wir selber reproduzierbar?

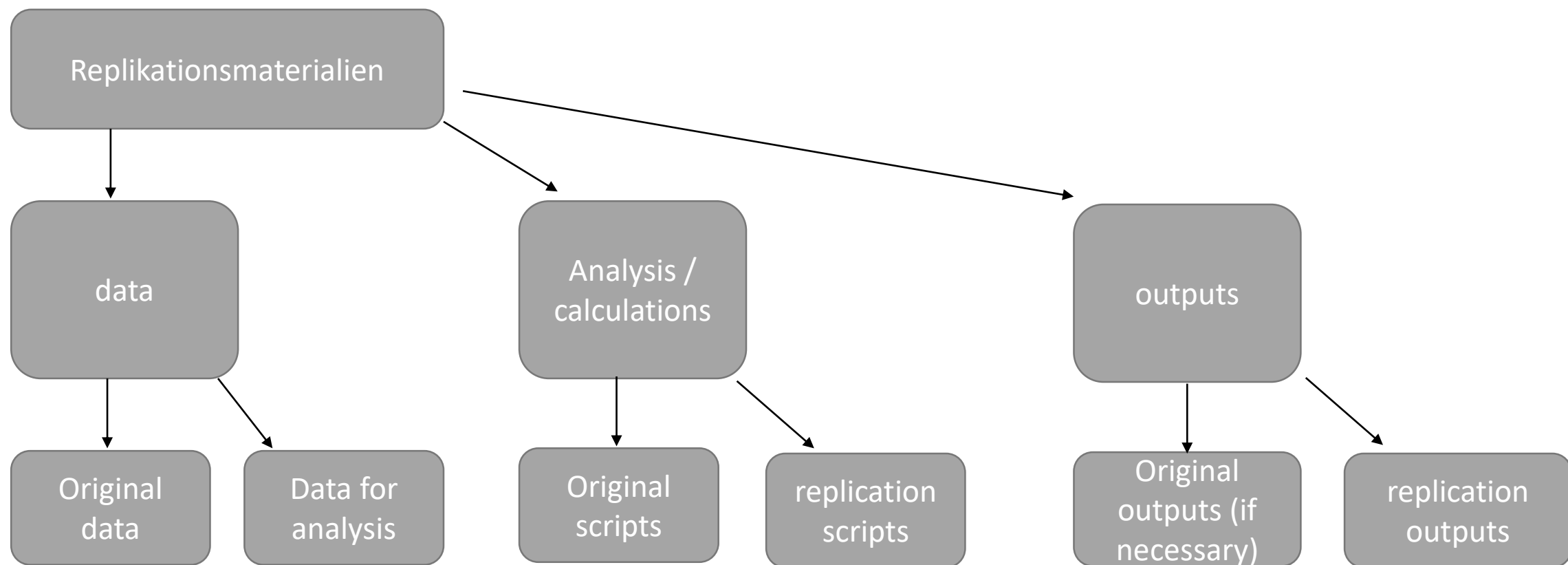


Diskussion 5 Minute Paper

Tauschen Sie sich zunächst mit der:m Sitznachbar:in aus.

Anschließend diskutieren wir das im Plenum

Vorgeschlagene Ordnerstruktur



Wie planen wir unsere Replikationsstudie?
Wie arbeiten wir selber reproduzierbar?



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Wie arbeiten wir selber reproduzierbar?

Suchen Sie zur nächsten Sitzung Replikationsmaterialien eines veröffentlichten Artikel heraus, und schauen Sie in die Skripte und Ordner.
Was fällt Ihnen auf?