

Digitale Datenspuren nutzbar machen

Datenspenden als Methode der Kommunikationswissenschaft

Sitzung **5**: Datenspende-Studien aus Sicht der Forscher:innen

Valerie Hase (Ludwig-Maximilians-Universität München)

 github.com/valeriehase & valerie-hase.com

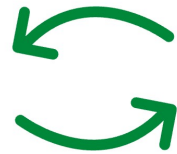
1) Wie laufen Datenspende-Studien aus Sicht der Forscher:innen ab?



Quelle: Foto von Markus Winkler auf Unsplash

*Welche methodischen Entscheidungen müssen
Forscher:innen treffen, die Datenspende-Studien
durchführen wollen? 🤔*

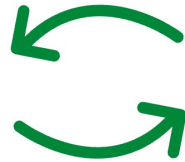
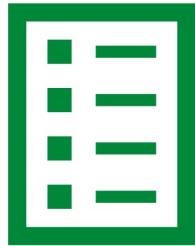
Ablauf - Perspektive der Forschung



- 1** Forschungsdesign
 - Forschungsfragen
 - Technische Umsetzung
 - Umfrage
 - Datenspende
- 2** Bereinigung
 - Klassifikation (Links, Accounts)
 - Matching & Aggregation
 - Test auf Verzerrungen
- 3** Auswertung

Grafik. Ablauf einer Datenspende-Studie - Perspektive der Forschung

Schritt I: Forschungsdesign



1 Forschungsdesign

- Forschungsfragen
- Technische Umsetzung
 - Umfrage
 - Datenspende

Grafik. Ablauf einer Datenspende-Studie - Perspektive der Forschung

Schritt I: Forschungsdesign

Zentrale Entscheidungen:

- Welche theoretischen Fragen wollen & können wir beantworten?
- Welches Datenspende-Tool nutzen wir wie?
- Wie wird die Datenspende in die Umfrage integriert?

Schritt I: Forschungsdesign

Zentrale Entscheidungen:

- **Welche theoretischen Fragen wollen & können wir beantworten?**
- Welches Datenspende-Tool nutzen wir wie?
- Wie wird die Datenspende in die Umfrage integriert?

Schritt I: Forschungsdesign - Forschungsfragen

Bitte überlegt euch eine Forschungsfrage/Hypothese, die ihr auf Basis einer Datenspende-Studie beantworten wollen würdet. 🤔

Schritt I: Forschungsdesign

Zentrale Entscheidungen:

- **Welche theoretischen Fragen wollen & können wir beantworten?**
 - Können wir theoretische Konzepte mit Datenspenden operationalisieren?
 - Können wir relevante Grundgesamtheiten abbilden?

Schritt I: Forschungsdesign

Zentrale Entscheidungen:

- Welche theoretischen Fragen wollen & können wir beantworten?
- **Welches Datenspende-Tool nutzen wir wie?**
- Wie wird die Datenspende in die Umfrage integriert?

Schritt I: Forschungsdesign - Datenspende-Tool

Datenspende-Tools

- PORT (Boeschoten et al., 2022) als Weiterentwicklung von OSD2F (Araujo et al., 2022) (Niederlande)
- Data Donation Module (Pffner et al., 2022) (Schweiz)

Schritt I: Forschungsdesign - Datenspende-Tool

- Teilnehmer:innen laden DDPs hoch
- Daten werden lokal (d.h. auf Rechner der Teilnehmer:innen) gefiltert

Schritt I: Forschungsdesign - Datenspende-Tool

- Teilnehmer:innen laden DDPs hoch
- Daten werden lokal (d.h. auf Rechner der Teilnehmer:innen) gefiltert

Extraktion
spezifischer Ordner
via CSS



Name	Typ
ads_and_businesses	Dateiordner
ads_and_topics	Dateiordner
apps_and_websites	Dateiordner
autofill_information	Dateiordner
avatars_store	Dateiordner
comments	Dateiordner
contacts	Dateiordner
content	Dateiordner
device_information	Dateiordner
digital_wallets	Dateiordner
events	Dateiordner
followers_and_following	Dateiordner
fundraisers	Dateiordner
guides	Dateiordner
information_about_you	Dateiordner
likes	Dateiordner
login_and_account_creation	Dateiordner
loyalty_accounts	Dateiordner
media	Dateiordner
media_settings	Dateiordner
messages	Dateiordner

Grafik. Ablauf einer Datenspende-Studie - Perspektive der Forschung

Schritt I: Forschungsdesign - Datenspende-Tool

- Teilnehmer:innen laden DDPs hoch
- Daten werden lokal (d.h. auf Rechner der Teilnehmer:innen) gefiltert
- Daten werden lokal anonymisiert, z. B. über Whitelists



```
liked_posts - Editor
Datei Bearbeiten Format Ansicht Hilfe
{
  "likes_media_likes": [
    {
      "title": "kiricheeeeeese"
      "media_list_data": [
        ],
      "string_list_data": [
        {
          "href": "https://www.instagram.com/p/kttM6Wy/",
          "timestamp": 1393065075
        }
      ]
    },
    {
      "title": "tagesschau",
      "media_list_data": [
        ],
      "string_list_data": [
        {
          "href": "https://www.instagram.com/p/Cs3LwsrI8/",
          "timestamp": 1392964862
        }
      ]
    }
  ]
}
```

z.B. Ausschluss, da nicht auf Whitelist

Grafik. Ablauf einer Datenspende-Studie - Perspektive der Forschung

Schritt I: Forschungsdesign - Datenspende-Tool

- Teilnehmer:innen laden DDPs hoch
- Daten werden lokal (d.h. auf Rechner der Teilnehmer:innen) gefiltert
- Daten werden lokal anonymisiert, z. B. über Whitelists
- Teilnehmer:innen können Daten löschen

The screenshot shows the 'Datenspende-Tool' interface. At the top, it says 'Einträge in Ihrer Datenspende: 180'. Below this are two buttons: 'Ansehen/Filtern' and 'Spenden'. There are four tabs: 'Ihre Datenspende', 'liked_posts.json', 'liked_comments.json', and 'following.json'. The 'liked_posts.json' tab is active. Below the tabs, it says 'Datei: liked_posts.json', 'Einträge in dieser Datei: 179', and 'Einträge je Seite: 10'. There is a search bar with the text 'Suche ...' and a red button 'Zeilen entfernen'. Below the search bar is a table with two columns: 'string_list_data' and 'title'. The table contains five rows of data, each representing an Instagram post. The first row is:

```
{ "href": "https://www.instagram.com/p/CV54rkojsq7/", "timestamp": 1634861174 }
```

 with the title '<user>'. The second row is:

```
{ "href": "https://www.instagram.com/p/B8YdCejiMdtHY74miApr2HvNAA2WfpPexAJbqY0/", "timestamp": 1581329148 }
```

 with the title '<user>'. The third row is:

```
{ "href": "https://www.instagram.com/p/B6bfmLXpGzPGT0zK8RPjh3vnbHfoNkXUABknE0j/", "timestamp": 1577174728 }
```

 with the title '<user>'. The fourth row is:

```
{ "href": "https://www.instagram.com/p/B45i3iehJSRUWp8V94cSzsdnA0rB7id1q6bpM0/", "timestamp": 1573842550 }
```

 with the title '<user>'. The fifth row is:

```
{ "href": "https://www.instagram.com/p/B4mulaDiNzQj/", "timestamp": 1573226073 }
```

 with the title '<user>'. The sixth row is:

```
{ "href": "https://www.instagram.com/p/B4cev3uo8egj/", "timestamp": 1572874850 }
```

 with the title 'tagesschau'.

Grafik. Ablauf einer Datenspende-Studie - Perspektive der Forschung

Schritt I: Forschungsdesign - Datenspende-Tool

- Teilnehmer:innen laden DDPs hoch
- Daten werden lokal (d.h. auf Rechner der Teilnehmer:innen) gefiltert
- Daten werden lokal anonymisiert, z. B. über Whitelists
- Teilnehmer:innen können Daten löschen
- nach informierter Einwilligung werden Daten an Server der Forscher:innen weitergeleitet

Schritt I: Forschungsdesign - Datenspende-Tool

Beispiel: OSD2F (Araujo et al., 2022, S. 377)

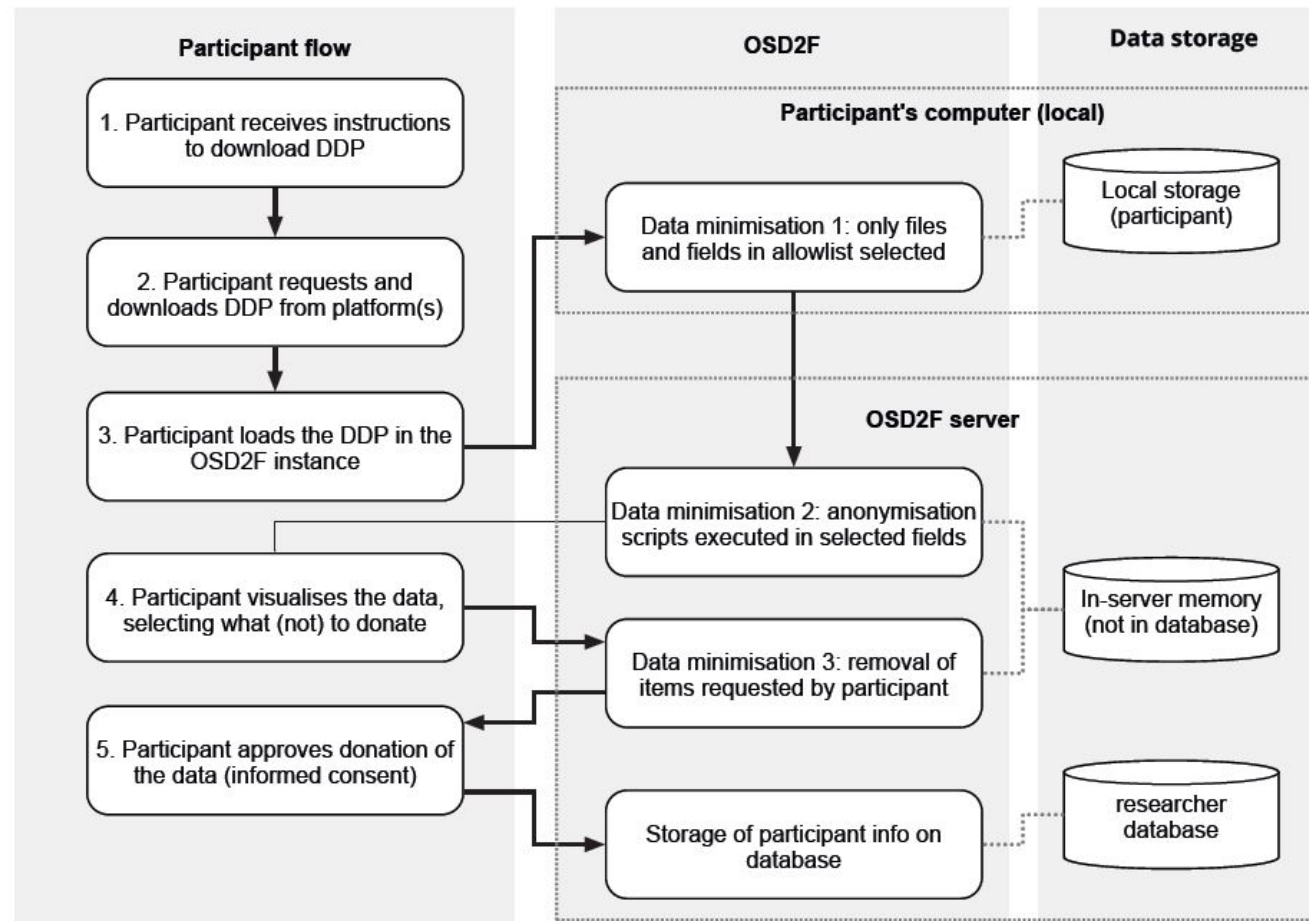


Figure 1. Overview of OSD2F flow

Araujo et al., 2022, S. 377

Schritt I: Forschungsdesign

Zentrale Entscheidungen:

- Welche theoretischen Fragen wollen & können wir beantworten?
- **Welches Datenspende-Tool nutzen wir wie?**
 - Kann das Tool relevante Daten datenschutzkonform extrahieren & anonymisieren?
 - Können wir Skripte schnell an Veränderungen durch Plattformen anpassen?
- Wie wird die Datenspende in die Umfrage integriert?

Schritt I: Forschungsdesign

Zentrale Entscheidungen:

- Welche theoretischen Fragen wollen & können wir beantworten?
- Welches Datenspende-Tool nutzen wir wie?
- **Wie wird die Datenspende in die Umfrage integriert?**

Schritt I: Forschungsdesign - Umfrage

- geläufiges Vorgehen: Weiterleiten an externe Seite
- neu: Integration in Umfragen, z. B. via SoSci Survey (Haim et al., 2023); ermöglicht Messwiederholungen, Incentivierung, Experimentaldesigns, etc.

The screenshot shows the configuration page for a data donation study in SoSci Survey. The interface includes several input fields and a code editor. A red box highlights a warning message about the function's development status.

Frage-text: Keine Nummer anzeigen
Hier können Sie Ihre Twitter-Daten spenden.

Anweisungen:

Warning (red box): Diese Funktion befindet sich noch in der Entwicklung. Möglicherweise werden Funktion und/oder Aussehen ohne Vorwarnung geändert. Bitte kontaktieren Sie Ihren Server-Administrator, bevor Sie diese Funktion in einer Befragung verwenden. Bitte informieren Sie SoSci Survey auch über alle Fehler und Probleme, die Ihnen auffallen.

URL des OSD2F-Servers:

Server-Token:

Weiter-Knopf anzeigen:

Upload (JSON):

```
{
  "blocked_accounts json": {
    "in_key": "relationships_blocked_users",
    "anonymizers": [
      { "insta_anonymize_usernames": "string_list_data.value" }
    ]
  },
  "accepted_fields": [
    "string_list_data.value",
    "string_list_data.timestamp"
  ]
},
  "comments json": {
    "in_key": "comments_v2",
    "anonymizers": [
      { "fb_anonymize_comments": "title" }
    ]
  },
  "accepted_fields": [
    "title",
    "data.comment.timestamp"
  ]
},
  "following json": {
    "in_key": "relationships_following",
    "anonymizers": [
      { "insta_anonymize_usernames": "string_list_data.value" }
    ]
  }
}
```

Content/Upload ID: 6 / 7

Grafik. Ablauf einer Datenspende-Studie - Perspektive der Forschung

Schritt I: Forschungsdesign - Umfrage

Dadurch dass Datenspenden aufwendig sind (und damit Verzerrungen begünstigen), stellen sich weitere Fragen nach...

- Incentivierung
- Platzierung der Anfrage nach Datenspenden
- Formulierung der Bedeutsamkeit von Datenspenden

Schritt I: Forschungsdesign

Zentrale Entscheidungen:

- Welche theoretischen Fragen wollen & können wir beantworten?
- Welches Datenspende-Tool nutzen wir wie?
- **Wie wird die Datenspende in die Umfrage integriert?**
 - Wie können wir die Datenspende technisch so leicht wie möglich für Teilnehmer:innen machen?
 - Welche Design-Strategien können wir nutzen, um Teilnehmer:innen für die Datenspende zu begeistern?

Schritt II: Bereinigung



1 Forschungsdesign

- Forschungsfragen
- Technische Umsetzung
 - Umfrage
 - Datenspende

Grafik. Ablauf einer Datenspende-Studie - Perspektive der Forschung

Schritt II: Bereinigung



1 Forschungsdesign

- Forschungsfragen
- Technische Umsetzung
 - Umfrage
 - Datenspende

2 Bereinigung

- Klassifikation (Links, Accounts)
- Matching & Aggregation
- Test auf Verzerrungen

Grafik. Ablauf einer Datenspende-Studie - Perspektive der Forschung

Schritt II: Bereinigung

Zentrale Entscheidungen:

- Wie klassifizieren wir Inhalte?
- Wie matchen & aggregieren wir Inhalte?
- Wie testen wir für Verzerrungen?

Schritt II: Bereinigung

Zentrale Entscheidungen:

- Wie klassifizieren wir Inhalte?
- Wie matchen & aggregieren wir Inhalte?
- Wie testen wir für Verzerrungen?


👉 Übergreifendes Tutorial hierzu: Hase, V. (2023, Juli). Data Donations—Tutorial Automated Content Analysis. [Link](#)

Schritt II: Bereinigung

Zentrale Entscheidungen:

- **Wie klassifizieren wir Inhalte?**
- Wie matchen & aggregieren wir Inhalte?
- Wie testen wir für Verzerrungen?

Schritt II: Bereinigung - Klassifikation

- Anonymisierung als erster Schritt der Klassifikation basiert zumeist auf Whitelists, d.h. z.B. Listen von Nachrichtenmedien  [Beispiel](#)
- Aber auch anschliessend müssen...
 - Daten nachträglich via APIs gesammelt
 - Daten via automatisierter Inhaltsanalyse klassifiziert werden

Schritt II: Bereinigung - Klassifikation

id	submission_id	filename	n_deleted	insert_timestamp	update_timestamp	entry	
7868	308142	5345	liked_posts.json	0	2022-12-09 10:37:45.458707+00:00	2022-12-09 10:37:45.458714+00:00	{"string_list_data":[{"timestamp":1654035032}], "title": "<user>"}
7869	308143	5345	liked_posts.json	0	2022-12-09 10:37:45.458731+00:00	2022-12-09 10:37:45.458737+00:00	{"string_list_data":[{"timestamp":1654034499}], "title": "<user>"}
7870	308144	5345	liked_posts.json	0	2022-12-09 10:37:45.458754+00:00	2022-12-09 10:37:45.458761+00:00	{"string_list_data":[{"timestamp":1654034341}], "title": "<user>"}
7871	308145	5345	liked_posts.json	0	2022-12-09 10:37:45.458777+00:00	2022-12-09 10:37:45.458784+00:00	{"string_list_data":[{"timestamp":1654020807}], "title": "<user>"}
7872	308146	5345	liked_posts.json	0	2022-12-09 10:37:45.458801+00:00	2022-12-09 10:37:45.458808+00:00	{"string_list_data":[{"timestamp":1654020127}], "title": "<user>"}
7873	308147	5345	liked_posts.json	0	2022-12-09 10:37:45.458824+00:00	2022-12-09 10:37:45.458831+00:00	{"string_list_data":[{"timestamp":1654020057}], "title": "tagesschau"}
7874	308148	5345	liked_posts.json	0	2022-12-09 10:37:45.458847+00:00	2022-12-09 10:37:45.458854+00:00	{"string_list_data":[{"timestamp":1654019851}], "title": "<user>"}
7875	308149	5345	liked_posts.json	0	2022-12-09 10:37:45.458871+00:00	2022-12-09 10:37:45.458878+00:00	{"string_list_data":[{"timestamp":1654019739}], "title": "<user>"}
7876	308150	5345	liked_posts.json	0	2022-12-09 10:37:45.458894+00:00	2022-12-09 10:37:45.458901+00:00	{"string_list_data":[{"timestamp":1654019708}], "title": "<user>"}
7877	308151	5345	liked_posts.json	0	2022-12-09 10:37:45.458918+00:00	2022-12-09 10:37:45.458925+00:00	{"string_list_data":[{"timestamp":1653940335}], "title": "<user>"}
7878	308152	5345	liked_posts.json	0	2022-12-09 10:37:45.458941+00:00	2022-12-09 10:37:45.458948+00:00	{"string_list_data":[{"timestamp":1653938012}], "title": "<user>"}
7879	308153	5345	liked_posts.json	0	2022-12-09 10:37:45.458965+00:00	2022-12-09 10:37:45.458971+00:00	{"string_list_data":[{"timestamp":1653937848}], "title": "<user>"}
7880	308154	5345	liked_posts.json	0	2022-12-09 10:37:45.458988+00:00	2022-12-09 10:37:45.458995+00:00	{"string_list_data":[{"timestamp":1653937307}], "title": "<user>"}
7881	308155	5345	liked_posts.json	0	2022-12-09 10:37:45.459011+00:00	2022-12-09 10:37:45.459018+00:00	{"string_list_data":[{"timestamp":1653808843}], "title": "<user>"}
7882	308156	5345	liked_posts.json	0	2022-12-09 10:37:45.459035+00:00	2022-12-09 10:37:45.459042+00:00	{"string_list_data":[{"timestamp":1653781269}], "title": "<user>"}
7883	308157	5345	liked_posts.json	0	2022-12-09 10:37:45.459058+00:00	2022-12-09 10:37:45.459065+00:00	{"string_list_data":[{"timestamp":1653753711}], "title": "sz"}
7884	308158	5345	liked_posts.json	0	2022-12-09 10:37:45.459082+00:00	2022-12-09 10:37:45.459089+00:00	{"string_list_data":[{"timestamp":1653691455}], "title": "<user>"}
7885	308159	5345	liked_posts.json	0	2022-12-09 10:37:45.459105+00:00	2022-12-09 10:37:45.459112+00:00	{"string_list_data":[{"timestamp":1653674965}], "title": "<user>"}
7886	308160	5345	liked_posts.json	0	2022-12-09 10:37:45.459128+00:00	2022-12-09 10:37:45.459135+00:00	{"string_list_data":[{"timestamp":1653674398}], "title": "<user>"}


Grafik. Beispiel eines Datenspende-Datensatzes

Schritt II: Bereinigung - Klassifikation

	id	submission_id	filename	n_deleted	insert_timestamp	update_timestamp	entry
1	708905	9073	Suchverlauf.json	0	2022-12-17 12:43:07.127782+00:00	2022-12-17 12:43:07.127790+00:00	{"title":"Gesucht nach: kinocheck","titleUrl":"https://www.youtube.com/results?search_query=kinocheck"}
2	1050798	10102	Suchverlauf.json	0	2022-12-20 11:08:43.968028+00:00	2022-12-20 11:08:43.968035+00:00	{"title":"Gesucht nach: anno 1602 denkmal","titleUrl":"https://www.youtube.com/results?search_query=anno+1602+denkmal"}
3	619493	8665	Suchverlauf.json	0	2022-12-16 21:04:58.414825+00:00	2022-12-16 21:04:58.414832+00:00	{"title":"Gesucht nach: ytitti","titleUrl":"https://www.youtube.com/results?search_query=ytitti"}
4	938862	9908	Suchverlauf.json	0	2022-12-19 13:26:30.762649+00:00	2022-12-19 13:26:30.762657+00:00	{"title":"Coop Erbjudande v6 angesehen","titleUrl":"https://www.youtube.com/watch?v=q1goWZD8nQ"}
5	1289477	10178	Suchverlauf.json	0	2022-12-28 15:33:30.872355+00:00	2022-12-28 15:33:30.872362+00:00	{"title":"The spring collection angesehen","titleUrl":"https://www.youtube.com/watch?v=fi49A9iB1hA"}

Grafik. Beispiel eines Datenspende-Datensatzes

Schritt II: Bereinigung - Klassifikation

- Anonymisierung als erster Schritt der Klassifikation basiert zumeist auf Whitelists, d.h. z.B. Listen von Nachrichtenmedien  [Beispiel](#)
- Aber auch anschliessend müssen...
 - Daten nachträglich via APIs gesammelt
 - Daten via automatisierter Inhaltsanalyse klassifiziert werden
- Fraglich ist, ob
 - wir Nachrichten auf Basis ihrer Quelle (z.B. Nachrichtenaccounts) klassifizieren können ([Reiss, 2022](#))
 - binäre Klassifikationen als Nachricht ja/nein überhaupt sinnvoll sind

Schritt II: Bereinigung

Zentrale Entscheidungen:

- **Wie klassifizieren wir Inhalte?**
 - Nutzen wir automatisierte Methoden?
 - Wie definieren wir, theoretisch gesehen, Nachrichten?
- Wie matchen & aggregieren wir Inhalte?
- Wie testen wir für Verzerrungen?

Schritt II: Bereinigung

Zentrale Entscheidungen:

- Wie klassifizieren wir Inhalte?
- **Wie matchen & aggregieren wir Inhalte?**
- Wie testen wir für Verzerrungen?

Schritt II: Bereinigung - Matching & Aggregation

Matching 🤝, d.h. Zusammenführung von Umfragedaten und Datenspuren (Munzert et al., 2023)

- meist via anonymer ID (z.B. "Interviewnummer")
- Daten sollten, soweit möglich, getrennt gespeichert werden

Schritt II: Bereinigung - Matching & Aggregation

Aggregation , d.h. Zusammenfassung von Datenpunkten über Individuen oder Zeit (Munzert et al., 2023)

- zumeist notwendig für Analyse (und um Daten ggf. zu teilen)
- aber: Aggregation kann grossen Einfluss auf Resultate haben

Schritt II: Bereinigung

Zentrale Entscheidungen:

- Wie klassifizieren wir Inhalte?
- **Wie matchen & aggregieren wir Inhalte?**
 - Wie gewährleisten wir die anonymisierte Verlinkung unterschiedlicher Daten?
 - Welches Aggregationsniveau benötigen wir?
- Wie testen wir für Verzerrungen?

Schritt II: Bereinigung

Zentrale Entscheidungen:

- Wie klassifizieren wir Inhalte?
- Wie matchen & aggregieren wir Inhalte?
- **Wie testen wir für Verzerrungen?**

Schritt II: Bereinigung - Verzerrungen

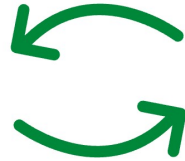
- Verzerrungen durch Fehler in Bezug auf Stichproben und Messungen, z.B.
 - z.B. durch systematisches Drop-Out ([Pak et al., 2022](#))
 - z.B. durch maschinelles Lernen für Klassifikation ([TeBlunthuis et al., 2023](#))

Schritt II: Bereinigung

Zentrale Entscheidungen:

- Wie klassifizieren wir Inhalte?
- Wie matchen & aggregieren wir Inhalte?
- **Wie testen wir für Verzerrungen?**
 - aktuell wenig adressiertes Problem in den Sozialwissenschaften
 - eher Quantifizierung als Adressierung

Schritt II: Bereinigung



1 Forschungsdesign

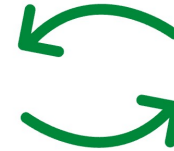
- Forschungsfragen
- Technische Umsetzung
 - Umfrage
 - Datenspende

2 Bereinigung

- Klassifikation (Links, Accounts)
- Matching & Aggregation
- Test auf Verzerrungen

Grafik. Beispiel eines Datenspende-Datensatzes

Schritt II: Auswertung



- 1** Forschungsdesign
 - Forschungsfragen
 - Technische Umsetzung
 - Umfrage
 - Datenspende
- 2** Bereinigung
 - Klassifikation (Links, Accounts)
 - Matching & Aggregation
 - Test auf Verzerrungen
- 3** Auswertung

Grafik. Beispiel eines Datenspende-Datensatzes

Schritt II: Auswertung

- deskriptive/inferenzstatistische Analysen, z.B.
 - um Unterschiede zwischen Befragungsdaten und digitalen Datenspuren zu verstehen (Parry et al., 2021; Scharkow, 2016)
 - um sozialwissenschaftliche Konzepte zu untersuchen (Thorson et al., 2021)

Zusammenfassung: Datenspenden-Studien aus Sicht der Forscher:innen

- **Zusammenfassung:** Zentrale Schritte umfassen...
 1. Forschungsdesign
 2. Bereinigung
 3. Auswertung
- **Weiterführende Literatur:**
 - Driel et al. (2022)
 - Boeschoten et al. (2022)
 - Haim et al. (2023)

Fragen? 🤔

Quellen

- Araujo, T., Ausloos, J., Atteveldt, W. van, Loecherbach, F., Moeller, J., Ohme, J., Trilling, D., Velde, B. van de, Vreese, C. de, & Welbers, K. (2022). OSD2F: An Open-Source Data Donation Framework. *Computational Communication Research*, 4(2), 372–387. <https://doi.org/10.5117/CCR2022.2.001.ARAU>
- Boeschoten, L., Mendrik, A., Van Der Veen, E., Vloothuis, J., Hu, H., Voorvaart, R., & Oberski, D. L. (2022). Privacy-preserving local analysis of digital trace data: A proof-of-concept. *Patterns*, 3(3), 100444. <https://doi.org/10.1016/j.patter.2022.100444>
- Driel, I. I. van, Giachanou, A., Pouwels, J. L., Boeschoten, L., Beyens, I., & Valkenburg, P. M. (2022). Promises and Pitfalls of Social Media Data Donations. *Communication Methods and Measures*, 1–17. <https://doi.org/10.1080/19312458.2022.2109608>
- Haim, M., Leiner, D., & Hase, V. (2023). Integrating Data Donations into Online Surveys. *Medien & Kommunikationswissenschaft*, 71(1-2), 130–137. <https://doi.org/10.5771/1615-634X-2023-1-2-130>
- Munzert, S., Ramirez-Ruiz, S., Watteler, O., Breuer, J., Batzdorfer, V., Eder, C., Wiltshire, D. A., Barberá, P., Guess, A. M., & Yang, J. (2023). *Publishing Combined Web Tracking and Survey Data* [Preprint]. Open Science Framework. <https://doi.org/10.31219/osf.io/y4v8z>
- Pak, C., Cötter, K., & Thorson, K. (2022). Correcting Sample Selection Bias of Historical Digital Trace Data: Inverse Probability Weighting (IPW) and Type II Tobit Model. *Communication Methods and Measures*, 16(2), 134–155. <https://doi.org/10.1080/19312458.2022.2037537>

- Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, 5(11), 1535–1547. <https://doi.org/10.1038/s41562-021-01117-5>
- Pfiffner, N., Witlox, P., & Friemel, T. N. (2022). *Data Donation Module*. <https://github.com/uzh/ddm>
- Reiss, M. V. (2022). Dissecting Non-Use of Online News – Systematic Evidence from Combining Tracking and Automated Text Classification. *Digital Journalism*, 1–21. <https://doi.org/10.1080/21670811.2022.2105243>
- Scharkow, M. (2016). The Accuracy of Self-Reported Internet Use—A Validation Study Using Client Log Data. *Communication Methods and Measures*, 10(1), 13–27. <https://doi.org/10.1080/19312458.2015.1118446>
- TeBlunthuis, N., Hase, V., & Chan, C.-H. (2023). *Misclassification in Automated Content Analysis Causes Bias in Regression. Can We Fix It? Yes We Can!* <https://doi.org/10.48550/ARXIV.2307.06483>
- Thorson, K., Cotter, K., Medeiros, M., & Pak, C. (2021). Algorithmic inference, political interest, and exposure to news and politics on Facebook. *Information, Communication & Society*, 24(2), 183–200. <https://doi.org/10.1080/1369118X.2019.1642934>

