

u^{*b*}

b

**UNIVERSITÄT
BERN**

u^b

5 Tips on using Social Media Data in Research

TDM Week@vonRoll

Christine Krebs, Jennifer Morger

Open Science Team, University Library of Bern

Data Stewards

Support in Research Data Management



u^b

Content

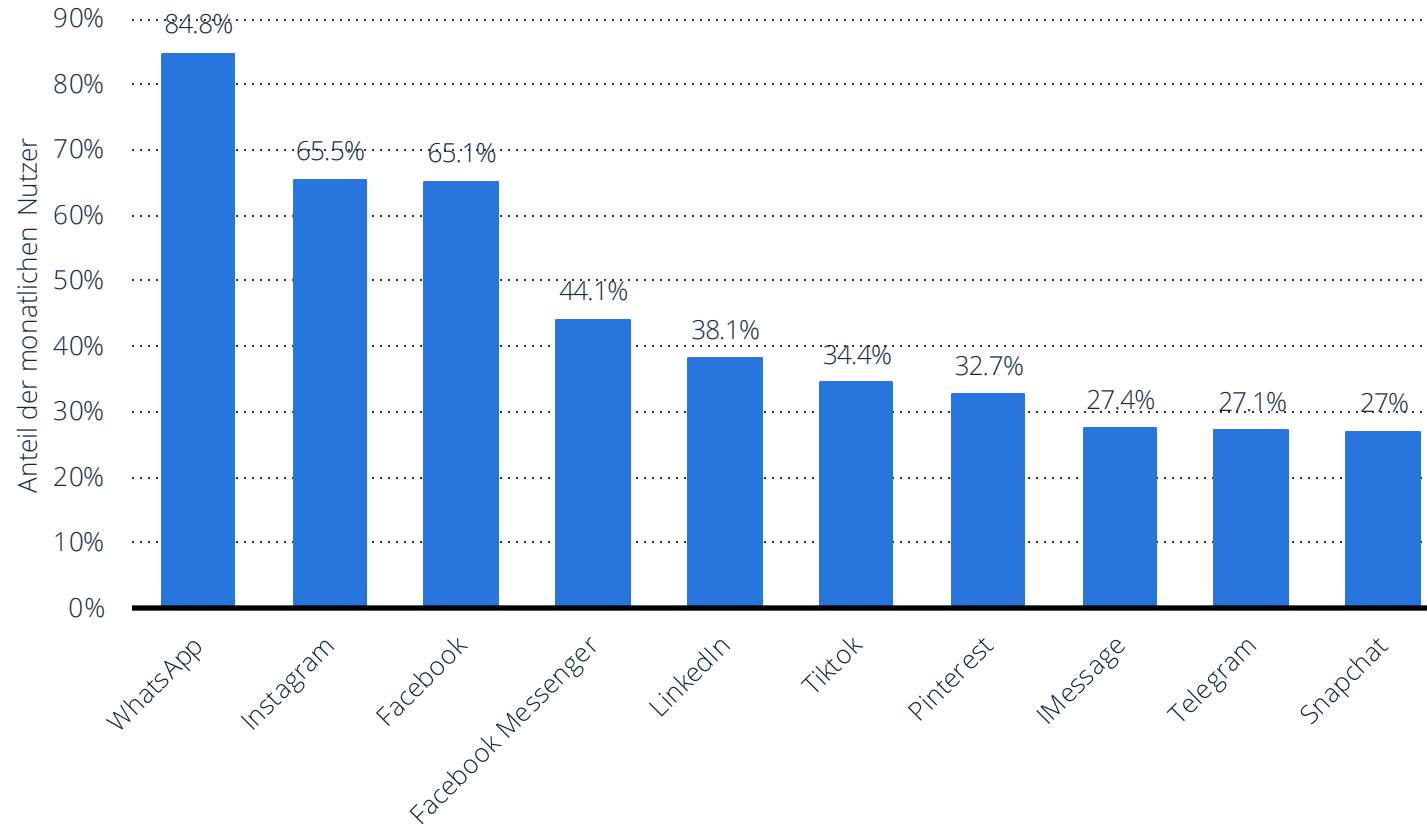
5 Tips on using Social Media Data in Research

- Tip 1: Identify platform(s)
- Tip 2: Get data from source: Collecting
- Tip 3: Get data from source: Platforms
- Tip 4: Getting data from the source is impossible, what now?
- Tip 5: Get help if needed

u^b

Tip 1: Identify platform(s)

Most used platforms in Switzerland



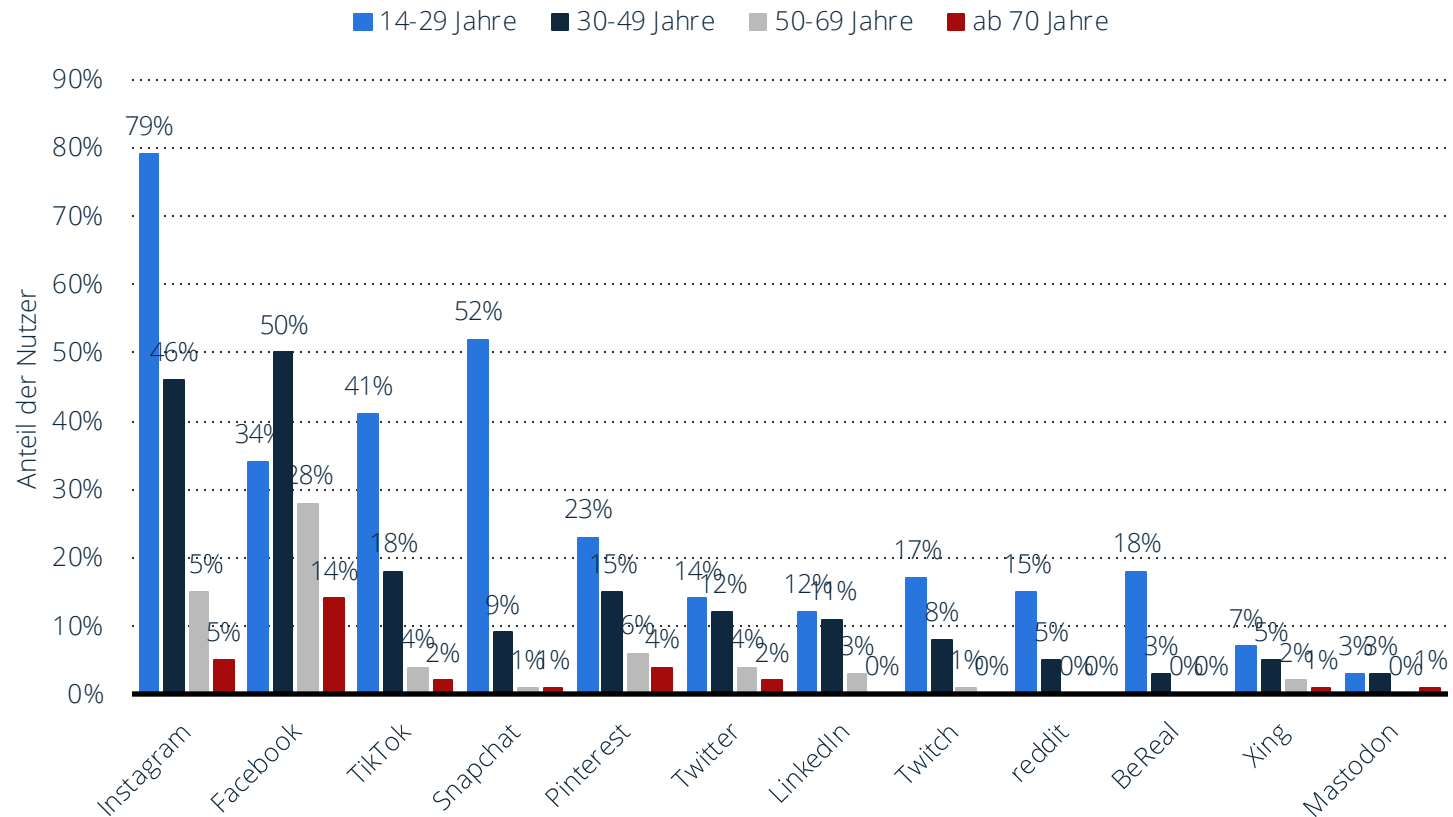
Switzerland; Q3 2023; 16 - 64 years; internet users

We Are Social, & DataReportal, & Meltwater. (23. Februar, 2024). Meistgenutzte Soziale Medien in der Schweiz nach monatlicher Nutzungsrate im Jahr 2023 [Graph]. In Statista, von <https://de.statista.com/statistik/daten/studie/467634/umfrage/bekanntheit-und-nutzung-von-ausgewaehlten-social-media-plattformen>

u^b

Tip 1: Identify platform(s)

Platforms by age groups in Germany



Germany; 2023; 14 years and older, n = 2000

ARD, & ZDF. (17. November, 2023). Anteil der Nutzer von Social-Media-Plattformen nach Altersgruppen in Deutschland im Jahr 2023 [Graph]. In Statista, von <https://de.statista.com/statistik/daten/studie/543605/umfrage/verteilung-der-nutzer-von-social-media-plattformen-nach-altersgruppen-in-deutschland/>

Tip 2: Get data from source: Collecting

Type of data collection	
Data donation	Users download data and share it with researchers, retrospective
Tracking	Capturing user logs mostly via plugins or apps, prospective
API	Often only official way to get data
Scraper, spider, crawler	Gathers data from websites
Manual collection by researcher	Manual collection of data

Legally complex

Adapted from: Davidson, B.I., Wischerath, D., Racek, D. et al. Platform-controlled social media APIs threaten open science. Nat Hum Behav 7, 2054–2057 (2023). <https://doi.org/10.1038/s41562-023-01750-2>

u^b

Tip 2: Get data from source: Collecting Scraping



YouTube Community Scraper

Soziale Medien

Extract user community posts on YouTube, including post text, time, likes, comments, attached video title, and description, etc.

Channel_name	Account	Subscribers	Bio	Community_U...	Post_date	Post_text	Post_URL	Like_count
National Geog...	@NatGeo	22.6M subscri...	Inspiring peop...	https://www.y...	6 days ago	On our planet ...	https://www.y...	464
National Geog...	@NatGeo	22.6M subscri...	Inspiring peop...	https://www.y...	8 days ago	Love is a biolo...	https://www.y...	2.5K
National Geog...	@NatGeo	22.6M subscri...	Inspiring peop...	https://www.y...	10 days ago	Paid content f...	https://www.y...	1.3K

u^b

Tip 2: Get data from source: Collecting Scraping: robots.txt file

- Generic name of a text file, indicates which parts of a website can be visited by bots/search engines
- If file exists, it is in the root directory of the domain

```
← ↻ 🏠 🔒 https://www.facebook.com/robots.txt
# Notice: Collection of data on Facebook through automated means is
# prohibited unless you have express written permission from Facebook
# and may only be conducted for the limited purpose contained in said
# permission.
# See: http://www.facebook.com/apps/site_scraping_tos_terms.php
```

```
User-agent: *
Disallow: /
```

➤ no scraping

```
User-agent: *
Disallow:
```

➤ full scraping

u^b

Tip 2: Get data from source: Collecting Tools

Social Media Data Collection & Analysis Tools



A no-code computational social science research tool for studying online communities and discourse.



A list of 50+ social media research



The Tweets Sampling Toolkit contains a



Netlytic is a text and social

From [Apps | Social Media Lab](#) (Toronto Metropolitan University)

u^b

Tip 3: Get data from source: Platforms Meta (Facebook/Instagram)

Meta Content Library and API

Meta Content Library and Content Library API provide comprehensive access to the full public content archive from Facebook and Instagram. Researchers can apply for access to the tools with the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan.

From [Meta Content Library and API | Transparency Centre \(fb.com\)](#)

Researcher Products

Researcher Platform

Researcher API

Researcher Data Sets

From [Facebook Open Research and Transparency Home \(fb.com\)](#)

u^b

Tip 3: Get data from source: Platforms X

As of 2023 the Twitter/X API is no longer freely available to academic researchers.

Options:

Basic X API package:

- \$100 per month – access to 10'000 posts a month

Pro package:

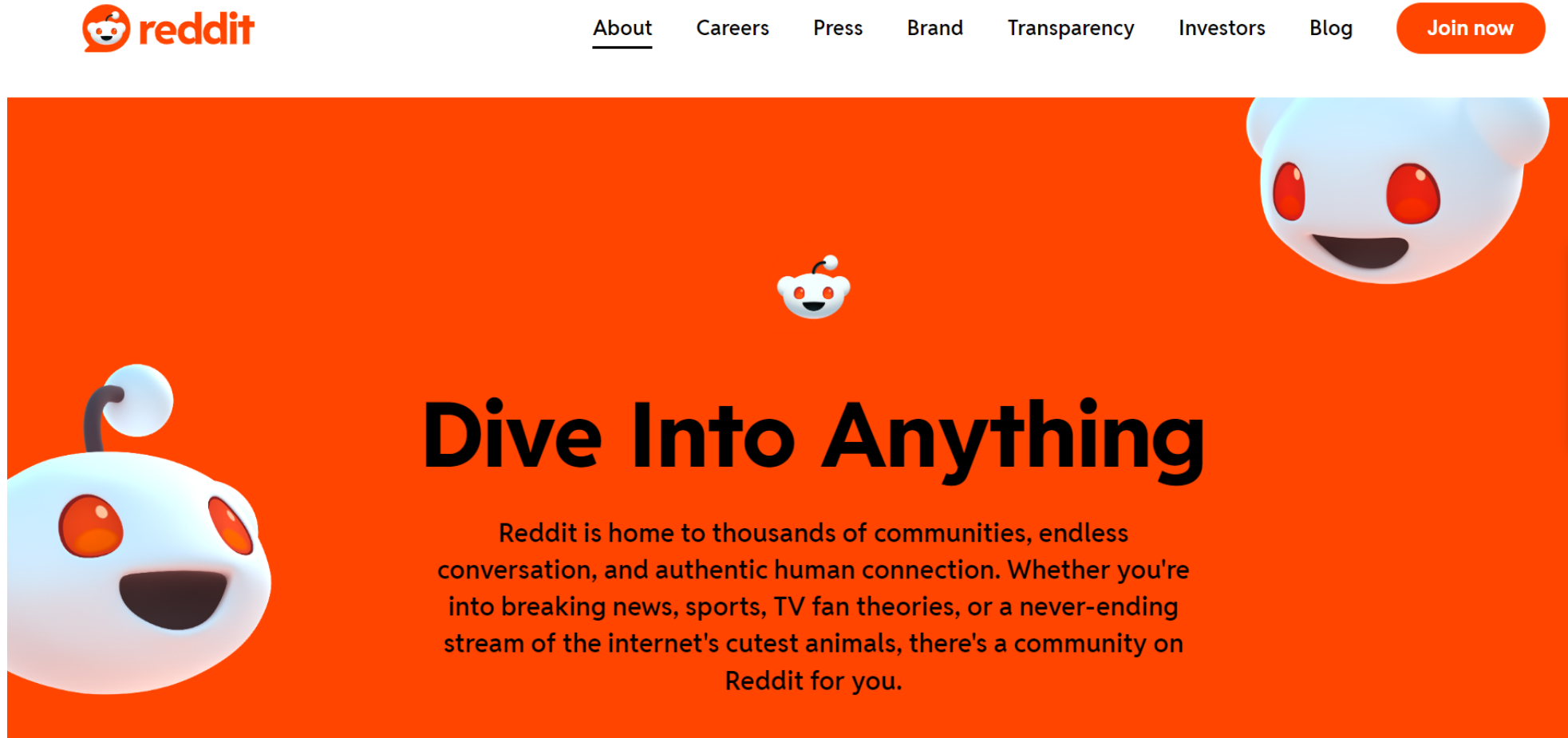
- \$5'000 per month – access to 1'000'000 tweets a month

[Link](#)

u^b

Tip 3: Get data from source: Platforms

Reddit



u^b

Tip 4: Getting data impossible, what now?

Social Media Archive



- 36 datasets
- login required
 - public & restricted datasets
- university of michigan

[SOMAR - Social Media Archive at ICPSR](#)

Tip 4: Getting data impossible, what now?

Tweet Datasets

TweetSets

(George Washington University)

Select source dataset

Select the source dataset from which to create a new dataset.

Other datasets that we've collected:

- Tax reform
- Hurricane Michael
- Solar Eclipse

Contact us if you'd like any of these loaded into TweetSets.

- 2016 United States Presidential Election: Democratic Candidate User Timelines **22,251 tweets**
- 2016 United States Presidential Election: Democratic Convention **8,340,688 tweets**
- 2016 United States Presidential Election: Democratic Party User Timelines **11,845 tweets**
- 2016 United States Presidential Election: Election Day Filter **3,826,254 tweets**
- 2016 United States Presidential Election: Election Filter **251,077,140 tweets**
- 2016 United States Presidential Election: First Presidential Debate Filter **3,182,243 tweets**
- 2016 United States Presidential Election: Republican Candidate User Timelines **62,395 tweets**
- 2016 United States Presidential Election: Republican Convention **6,248,688 tweets**

DocNow Tweet Catalog

143 Records comprising 7.665.613.088 tweets ADD RECORD

Subjects: All | Tweets Start: 20.03.2006 | Tweets End: 11.04.2024 | Search

Note: all metadata is shared under a CC0 license. Please read our [Code of Conduct](#) for more information about contributing datasets.

ADDED ↓	DATE RANGE	TITLE	TWEET COUNT	CREATORS	SUBJECTS	REPOSITORY
02/09/2023	04/08/2009 to 02/08/2023	TwitterDev	420.824	Ed Summers	Twitter	Github
01/18/2023	09/30/2019 to 04/26/2022	BillionGov: An enriched billion-scale collection of COVID-19 tweets for efficient hydration	1.410.446.121	Rabindra Lamsal, Maria Rodriguez Read, Shanika Karunasekera	COVID-19	IEEE

TweetsKB

(gesis)



u^b

Tip 4: Getting data impossible, what now?

Repositories for social sciences



[UK Data Service](#)

Limited collection of social media data, Facebook data, geo-coded Twitter data, and specific subsets of Wikipedia

gesis

Leibniz Institute
for the Social Sciences

gesis

Several Twitter datasets (20 collections of Twitter communication (tweets' IDs, timestamp, hashtags))

u^b

Tip 4: Getting data impossible, what now? Check general repositories

Limited but increasing number of social media datasets.



HARVARD

Dataverse

Be aware: no data curation - data obtained through self-archiving so be aware of potential data and metadata quality issues.

u^b

Tip 5: Get help if needed

Possible challenges:

- legal and data protection issues
- costs for data access
- often large amounts & rapidly growing data
- living data: data can be changed or deleted -> reproducibility problem

Get help:

- [Data Protection officer](#)
- At the Faculty:
 - Ethic commissions
 - FOMA
- [Research Management Office](#)
- [Digital Scholarship](#)
- [Data Science Lab](#)
- [Data Stewards](#) (Open Science Team)

NEXT BERN DATA TALKS

23.04.2024



Three researchers from different fields will share their insights on the **topic of sensitive data** in short 10-minute talks.

The meeting point is at the **UniS Schanzeneckstrasse 1, 3012 Bern, Room A022.**

Whether you're already a data pro or just getting started, the "Bern Data Talks" are open to everyone. No registration is required.

If you have any questions, feel free to contact openscience@unibe.ch. or check out the Uni Bern Website.

We look forward to seeing you!

5:15 pm

Nada El Makhzen

(Institute of Biochemistry and Molecular Medicine, Universität Bern)
Challenges in Generating, Sharing, and Storing Genetic Data for Patients with Rare Diseases in Africa: Insights from the Cystic Fibrosis Collaboration

5:30 pm

Regina Jenzer M.Sc.

(Soziale Arbeit, Fachhochschule Bern)
Research in civil law child protection: challenges in dealing with sensitive data (Talk in German)

5:45 pm

Prof. Dr. Michael Schulte-Mecklenbeck

(Institute of Marketing and Management, University of Bern)
Protecting personal information while sharing data: synthetic datasets

6:00 pm

Apéro

Transferable Skills Course

Introduction to Research Data Management

Kontakt Lageplan Jobs Bibliothek Medien Webmail KSL Uni intern DE Suche

AKTUELL UNIVERSITÄT STUDIUM **FORSCHUNG** WEITERBILDUNG FAKULTÄTEN & INSTITUTE

[Transferable Skills](#)

Workshop
Introduction to Research Data Management, spring 2024

Montag, 06.05.2024, 09:00 Uhr

Veranstaltende:	Vice-Rectorate International and Academic Affairs
Redner, Rednerin:	Research Data Management team staff, University Library Bern
Datum:	06.05.2024
Uhrzeit:	09:00 Uhr
Ort:	104 Main Building Hochschulstrasse 4 3012 Bern
Anmeldung:	here
Merkmale:	nicht öffentlich kostenlos

Dates Details

May 6, 2024, 9 a.m.-5 p.m.
May 7, 2024, 9 a.m.-5 p.m.

When:

May 6, 2024, 9 a.m.-5 p.m.

May 7, 2024, 9 a.m.-5 p.m.

Place:

Hochschulstrasse 4, Room 104
3012 Bern

Registration: Link is [here](#)

Open for PhD students, PostDocs
Free of charge

Credit Point: ECTS 1

u^b

Managing (sensitive) personal data in research

Target group: Researchers, data stewards

Goal: Knowledge exchange

When: 4 November 2024, 9:00 -17:15 h

Where: Hochschulstrasse 4 (Kuppelraum), Bern

Format: On-site only

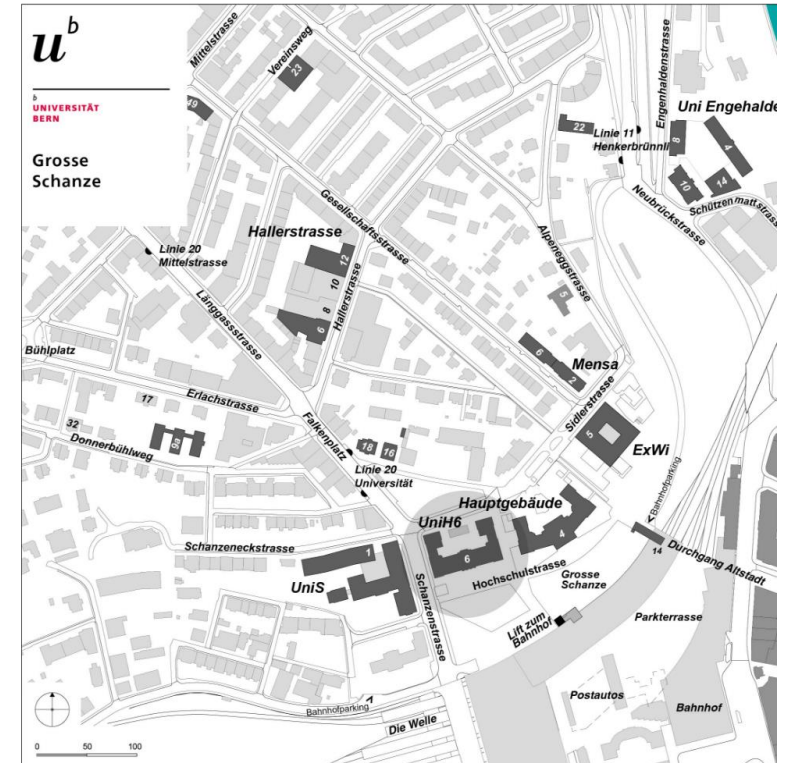
Registration: [Link](#), free of charge

Research Data Management Support Walk-In

When: Mondays 14:00-16:00 h

Where: Hochschulstrasse 6, Room 354, Bern

E-Mail: researchdata@unibe.ch



u^b

Service and Support

Webpage: [Research Data Management Support](#)



Newsletter: <http://www.unibe.ch/ub/osnews>



Survey: Your feedback on our workshops and suggestions <http://tiny.cc/rdmsurv>



Contact

Christine Krebs

E-Mail: christine.r.krebs@unibe.ch

Data Steward for Human Sciences



Jennifer Morger

E-Mail: jennifer.morger@unibe.ch

Data Steward for Business, Economics,
Social Science



u^b

Thank you

for your attention

Open Science Team

openscience@unibe.ch

