

Datenzugang im Digital Services Act: Ein Blick aus der Forschungs- und Infrastruktur-Perspektive

Katrin Weller
katrin.weller@gesis.org

Vortrag bei @ DSC, organisiert von der Bundesnetzagentur
December 1, 2025

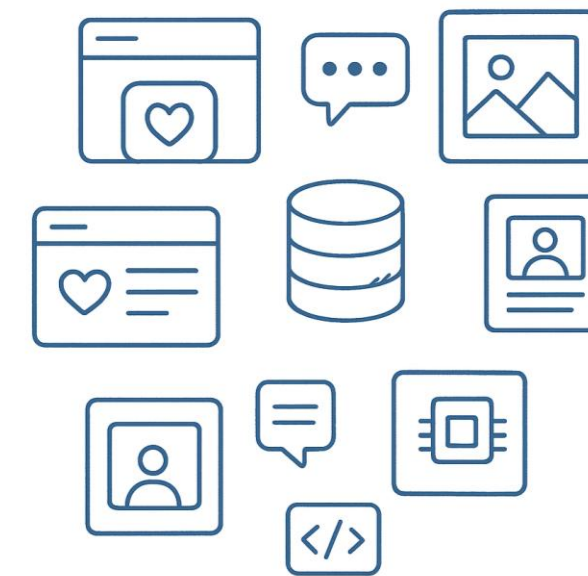


Eigener Hintergrund

Forschung

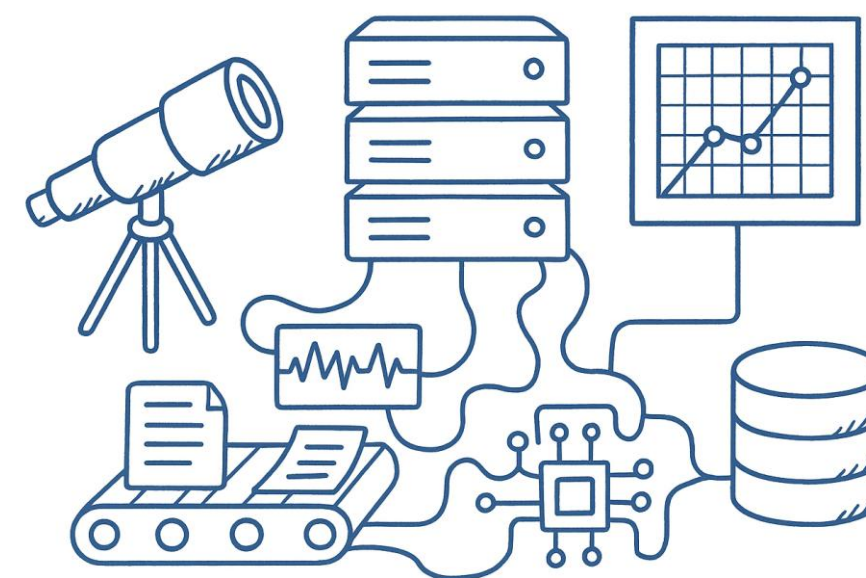
GESIS-Kolleg*innen arbeiten seit Jahren mit Online-Plattformdaten für verschiedene Forschungsfragen.

Derzeit sind erste Anträge im Rahmen des DSA Article 40(4) in Arbeit.



Infrastruktur

GESIS betreibt Infrastruktur für die (sozial-)wissenschaftliche Community. Dazu gehört u.a. ein Datenarchiv für Forschungsdaten, sowie Beratung, Guides und nachnutzbare Methoden.



RIDLOP – Research Infrastructure for Data from Large Online Platforms

Priorisierungsverfahren für große Forschungsinfrastrukturen (FIS) des BMFTR

- **Oktober 2024:** 32 Konzepte für FIS beim BMBF eingereicht
- **July 2025:** Shortlist mit 9 Konzepten veröffentlicht, inklusive RIDLOP
- **2026:** „Evaluative Begleitung“ durch Wissenschaftsrat
- **April 2026:** Begehung
- **2027:** Möglicher Start von RIDLOP

➤ Potentielle Finanzierung für 10 Jahre

➤ **Aber:** Bisher keine Finanzierungszusage durch BMFTR

Forschungsinfrastruktur (FIS) Priorisierungsverfahren

- **C** **CREATION**, Medizinische Hochschule Hannover, Universitätsmedizin Göttingen, Universitätsklinikum Leipzig, Fraunhofer-Institut für Zelltherapie und Immunologie (Leipzig/Potsdam-Golm/Halle (Saale)/Rostock/Würzburg)
- **D** **DALI**, Helmholtz-Zentrum Dresden-Rossendorf e.V.
- **E** **Vorphase für das Einstein-Teleskop**, RWTH Aachen, Universität Münster, Ruhr-Universität Bochum, Technische Universität Dresden
- **H** **HBS-I**, Forschungszentrum Jülich, Helmholtz-Zentrum Hereon (Geesthacht)
- **I** **IceCube-Gen2**, Deutsches Elektronen-Synchrotron (Hamburg/Zeuthen), Karlsruher Institut für Technologie
- **L** **LEGEND-1000**, Max-Planck-Institut für Kernphysik (Heidelberg), Technische Universität München
- **P** **PETRA IV**, Deutsches Elektronen-Synchrotron (Hamburg)
- **R** **RIDLOP**, GESIS - Leibniz-Institut für Sozialwissenschaften (Köln/Mannheim)
- **S** **SLICES-DE**, Technische Universität München

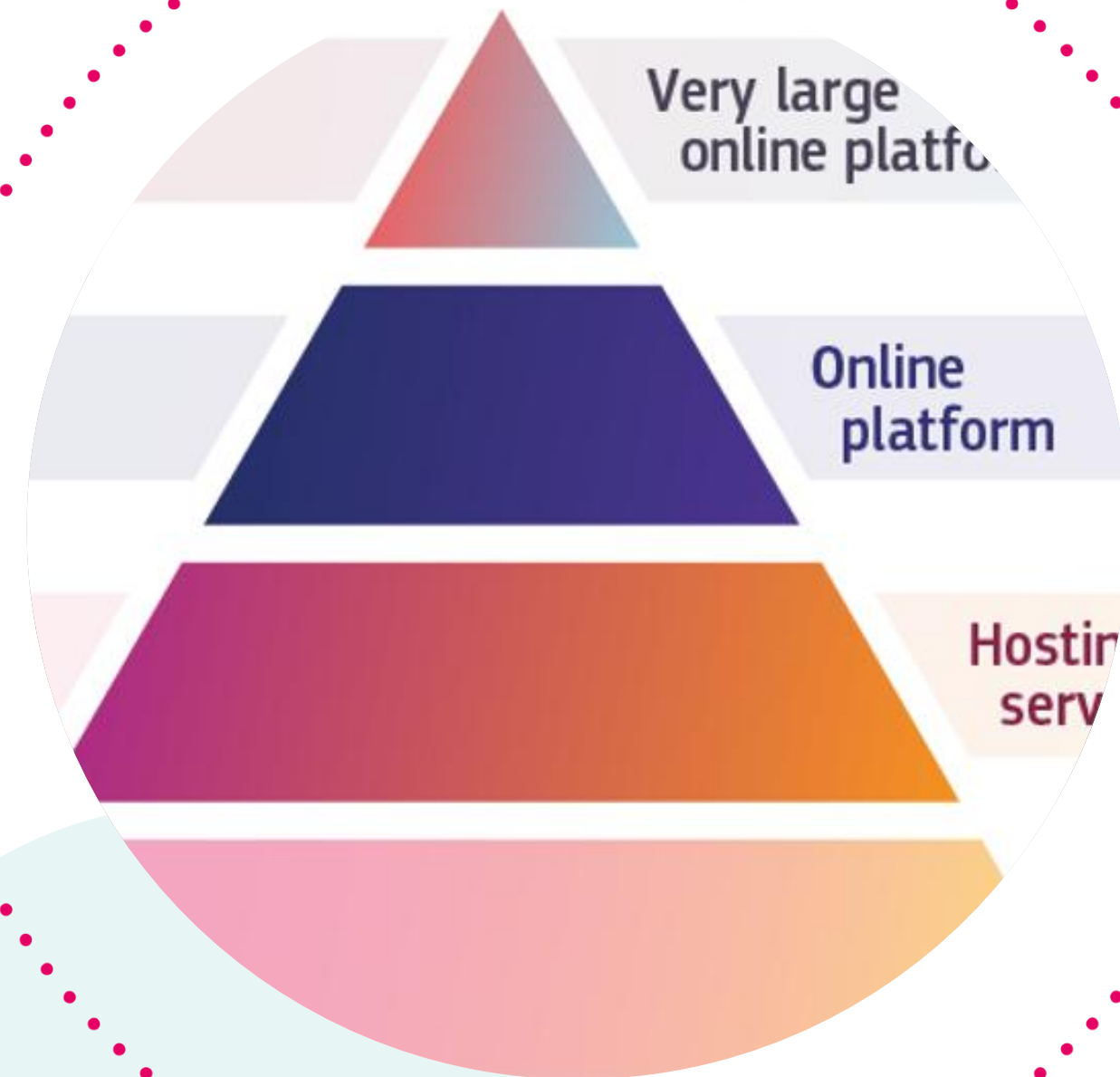


<https://www.forschungsministerium.de/fis>

Motivation für RIDLOP



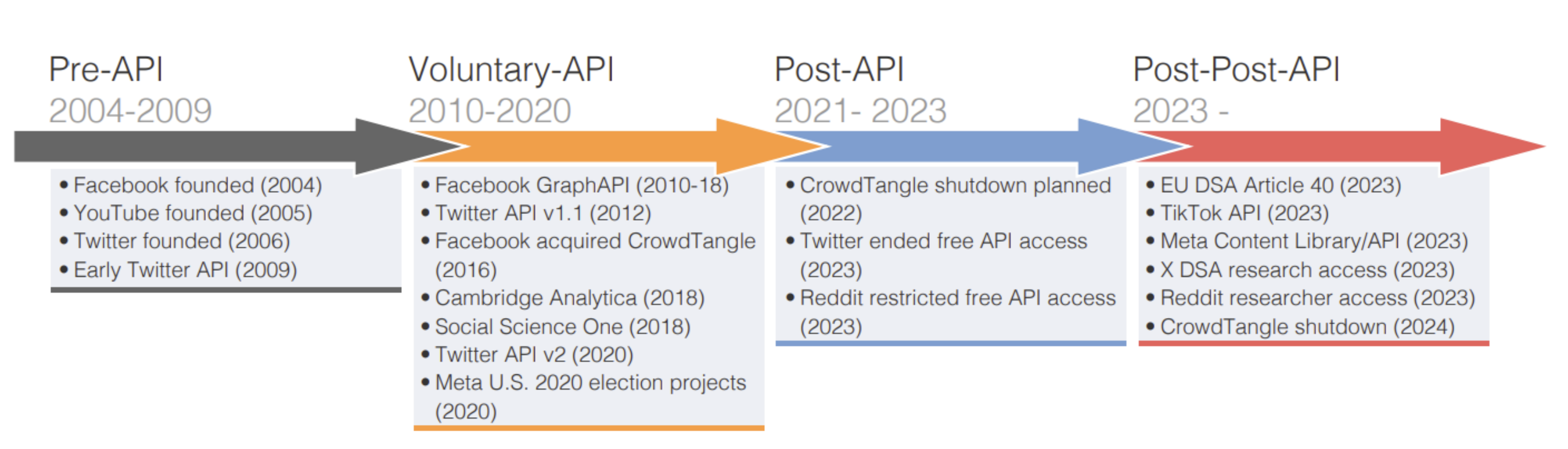
Es fehlt (weltweit) an spezialisierten Infrastrukturen für den **Zugang zu Daten aus Online Plattformen**



Durch den Digital Services Act (DSA) verändert sich der rechtliche Rahmen



Zeitalter des Datenzugangs



K. Mimizuka, M.A Brown, K-C. Yang, J. Lukito. 2025. Post-Post-API Age: Studying Digital Platforms in Scant Data Access Times. J. ACM 37, 4, Article 111 (May 2025).
<https://arxiv.org/abs/2505.09877>

Chancen und Risiken

Chancen

- DSA Article 40(4) eröffnet den Zugang zu Daten, die es in der Form bisher für die Forschung nicht gab – und damit auch zu ganz neuen Forschungsfragen. Ein Beispiel dafür sind Exposure Daten.
- Neue Standards im Umgang mit sensitiven, personenbezogenen Daten
- Neue Qualitätsstandards für Forschung mit Online-Plattformdaten

Risiken

- Potentiell fehlende Transparenz der Forschungsergebnisse
- Mögliche persönliche Risiken für Forschende, die zu besonders kritischen Themen arbeiten „Researchers at Risk“
- Vernachlässigung von Forschungsbereichen, die nicht (unmittelbar) als systemische Risiken eingestuft werden

Anträge 40(4) Datenzugang - Informationsstand

Was sind offene Punkte? Was ist noch unklar?

Formal:

Teilweise noch Unklarheiten im Antragsprozess, z.B.

- wie genau wird Funding bei Dauerfinanzierung angegeben? (Generelle Herausforderung wechselseitiger Abhängigkeit Fördergelder und Datenzugang)
- Wie geht man mit der Dynamik von Forschungsgruppen um (z.B. neue Personen kommen im Laufe eines Projektes hinzu)
- Für welchen Zeitraum beantragt man Datenzugang (DFG-Regelung: 10 Jahre aufbewahren)

Danke für vielfältige Informationsangebote und verschiedene Möglichkeiten zum Austausch und für Nachfragen!

Anträge 40(4) Datenzugang - Informationsstand

Was sind offene Punkte? Was ist noch unklar?

Inhaltlich:

- Prüfung von Systemic Risks in der Praxis
- Angemessene Formulierung von Datenbedarf/ konkrete Datenanforderung wird vermutlich ein iterativer Prozess
- Qualitätsprüfung der erhaltenen Daten

Infrastruktur-bezogen:

- Ablauf Prozesse zum tatsächlichen Datentransfer
- Dokumentation

Aus Forschungsperspektive

- Umdenken im Forschungsprozess erforderlich
- Das betrifft nicht nur die Art der Vorbereitung (Datenschutz, Ethics Approval, Forschungsdatenmanagement)
- Sondern auch die Art der Forschungsfragen (neue Datenverfügbarkeit)

Aber: Forschende sind daran gewöhnt, sich an die ständig wechselnde Landschaft im Zugangsprozess für Daten aus Online Plattformen anzupassen.

Kernpunkt aus Infrastrukturperspektive: Nachnutzbarkeit / Reproducibility

- Archivierung und Bereitstellung der Daten für die Sekundärnutzung ist nicht möglich.
- Das beeinträchtigt die Qualitätssicherung, insbesondere im Peer-Review-Verfahren, sowie die Reproduzierbarkeit und Wiederverwendung der Daten für weitere Forschung.
- Positiv ist, dass der DSA die (Open Access-)Veröffentlichung von Forschungsergebnissen vorschreibt. Während diese öffentliche Verfügbarkeit von Ergebnissen wichtig ist, sollten im weiteren Prozess auch nachhaltig höchste Standards bzgl. Methoden (z. B. Analyse-Skripte) und zugrundeliegenden Daten eingefordert werden.

Kernpunkt aus Infrastrukturperspektive: Nachnutzbarkeit / Reproducibility

Offene Fragen:

- Können zumindest aggregated statistics behalten/geteilt werden?
- Wie genau erfolgt die Übersicht über Anfragen im Data Access Portal? Was ist wenn eine Anfrage mehrere Datensätze oder Teilmengen erzeugt? Können DOIs oder andere eindeutige Kennungen vergeben werden.

Weiterführende Gedanken

- Feedback von GESIS zum Draft Delegated Act (DSA Article 40)
https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act/F3498981_en

Questions, ideas, or feedback?

I'd be happy to continue the conversation.

Katrin Weller

katrin.weller@gesis.org

Director „Data Services for the Social Sciences“, GESIS – Leibniz
Institute for the Social Sciences
Professor for Research Data Management, Heinrich Heine
University Düsseldorf