

Registerforschung

Daten des öffentlichen Bereichs Das Datenökosystem in Österreich

12. September 2023

Michael Strassnig, WWTF / Plattform Registerforschung

W|W|T|F



Was ist Registerforschung?

Registerforschung ist die wissenschaftliche Nutzung (*secondary use*) von Daten, die der Staat / staatliche Institutionen durch Verwaltungshandeln produziert bzw. operativ benötigt.

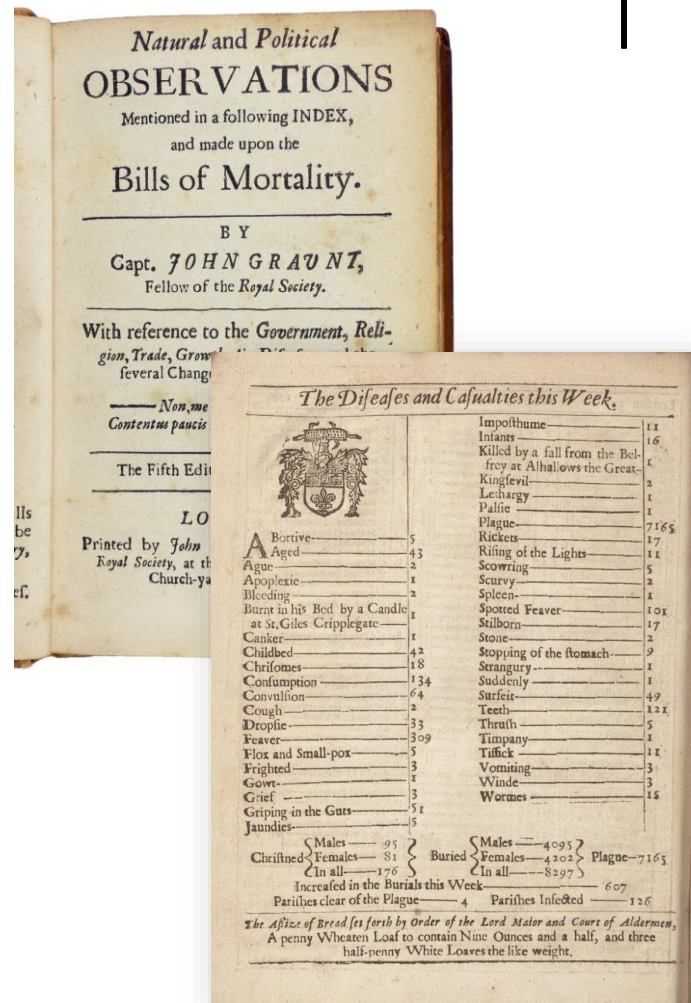
- Daten, die ein Staat braucht, um zu funktionieren. Z.B. Steuerdaten, Meldedaten, SV-Daten, Wählerevidenz ...
- Bundesgesetzlich vorgesehene Register / Administrativdaten (AMS, SV Datenbank)
- Dateneinheiten sind natürliche Personen, juristische Personen, aber auch viele andere (Arzneimittel, Rinder ...)
- Register verfolgen primär keine wissenschaftlichen Zwecke

Daten, die von der **amtlichen Statistik** erhoben werden, gehören in einem weiten Sinne auch dazu.

- Aber: für unmittelbares operatives Handeln nicht notwendig, eher für die Gesamtsteuerung des Staates → ermöglichen so der Politik, Maßnahmen zu setzen

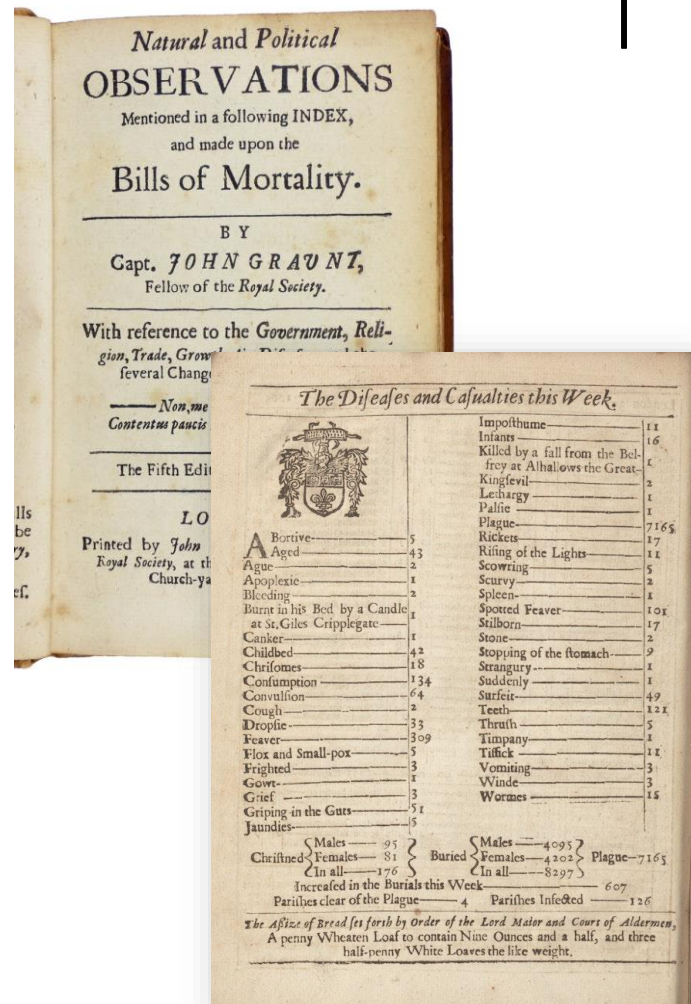
Ursprünge der Registerforschung

- Reicht zurück in die Ursprungsjahre der neuzeitlichen, experimentell orientierten Naturwissenschaft, die ihre Ursprünge in der Gründung der Royal Society 1660 hat.
- Um dieselbe Zeit: Londoner Kurzwarenhändler namens **John Graunt** hat begonnen, die wöchentlich veröffentlichten Sterbeverzeichnisse (Bill of Mortality) umfassender und mit von ihm neu entwickelten Methoden zu analysieren.



Ursprünge der Registerforschung

- Er bemerkte, dass die *Bills of Mortality* kaum genutzt werden abseits von Tratsch über die Anzahl von Begräbnissen
 - Er war einer der ersten, der **Sterbetafeln** produzierte und harte Evidenz über die Häufigkeit von Todesursachen lieferte.
 - Entwickelte **deskriptive statistische Methoden**, um aus den wöchentlichen Bills of Mortality jährliche, saisonale und regionale Statistiken zu erstellen
 - Er berichtete erstmals über **epidemiologische Trends** bei Krankheiten unter Berücksichtigung von Populationsgrößen
 - Er war der erste, der die **Bevölkerung** von London halbwegs vernünftig abschätzte und das Wachstum der Stadt aufzeigte und dass der Grund dafür in der Zuwanderung lag.
 - Er war der erste, der auch **Datenkritik** betrieb, indem er die Art und Weise der sogenannten „Searchers“ bei der Erhebung der Todesursachen hinterfragte.
 - → gilt als Begründer der **Demographie** und auch der **Epidemiologie**



Wozu Registerforschung?

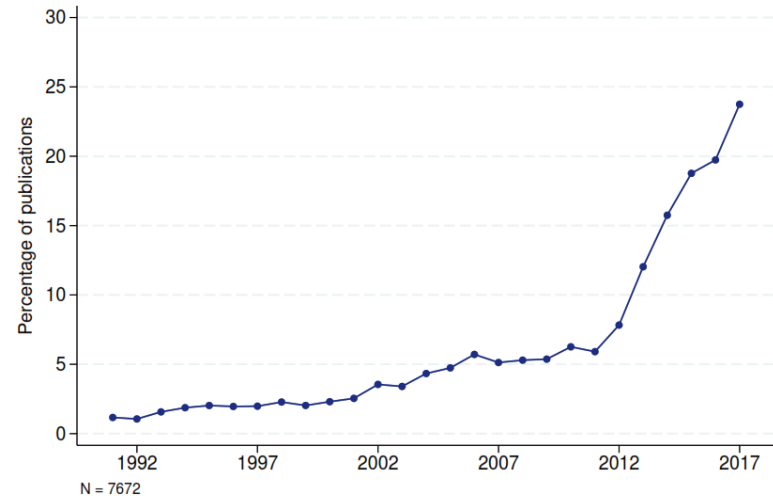
- **Enger Bezug der Daten zu Verwaltungshandeln/Politik** → in dem Sinne sind sie *real world data*
- → mit diesen Daten kann daher unmittelbar insbesondere staatliches Handeln erforscht werden (Maßnahme → Wirkung)
- Sowohl für Grundlagenforschung (Erkenntnisse über grundsätzliche Mechanismen in der Sozialpolitik / Wirtschaftspolitik, etc.) als auch in der angewandten Forschung / Evaluierungen (Auftraggeber will wissen, ob und wie eine bestimmte Maßnahme wirkt) relevant
- Möglicher Beitrag zu *evidence-based policy making* → vor / während / nach Maßnahmendurchführung

Relevanz der Registerforschung in der Wissenschaft

Zunehmende Bedeutung bei der Publikation
in den Sozial- und
Wirtschaftswissenschaften: Ökonomie,
Demographie, Epidemiologie & Public Health

Ökonomie-Zeitschriften (general interest) (Quelle: M. Halla 2020)

Average use of admin data in Top-5
3-year moving average



Vorteile der Forschung mit Registern

- **Es gibt sie schon!** D.h. wenige zusätzliche Kosten in der Erhebung im Vergleich zu eigenen Erhebungen/Umfragen
- Größe: Daten umfassen oft gesamte Populationen (und nicht Sample): **seltene Ereignisse** können beobachtet werden
- Daten in manchen Registern sind anderwärtig **unmöglich zu erheben** (weit zurückliegende Sachverhalte → historische Forschung (Matricula))
- IdR **Höhere Datenqualität**, da für Verwaltungszwecke erhoben und ev. bessere Qualitätssicherung
- **Datenschutzkonforme Verknüpfbarkeit** mit anderen Datenbeständen (bPK – bereichsspezifische Personenkennzeichen)
- Es werden einige **Verzerrungen in den Daten vermieden**, die insbesondere bei Befragungen schlagend werden
 - non-response bias; self-selection; drop outs bei Längsschnittstudien; social desirability bias; question order bias, ...

Nachteile der Forschung mit Registern

- **Unklarheit wie Daten generiert** werden bzw. was die **Datenqualität** ist (im Vergleich zur eigenen Erhebung)
- Notwendigkeit entsprechender **rechtlicher Rahmenbedingungen sowie Infrastruktur**
- **Monopol für Datenzugänge** liegt bei wenigen Institutionen; bürokratische Hürden
- **Mögliche Konflikte mit Datenschutz** (Re-identifikation von pseudonymisierten Daten; keine Einverständniserklärung, ...)
- Einschränkung bei der Datenanalyse (nicht jede Software möglich, da nur Remote Access, insbesondere bei non-Standard Software, z.B. AI)
- Einschränkung auf Variablen, die für die Verwaltung wichtig sind, aber möglicherweise forschungsrelevante Variablen nicht beinhalten

Registerdatenlandschaft in Österreich

Es gibt eine unglaubliche Vielfalt von Datensammlungen im öffentlichen Bereich

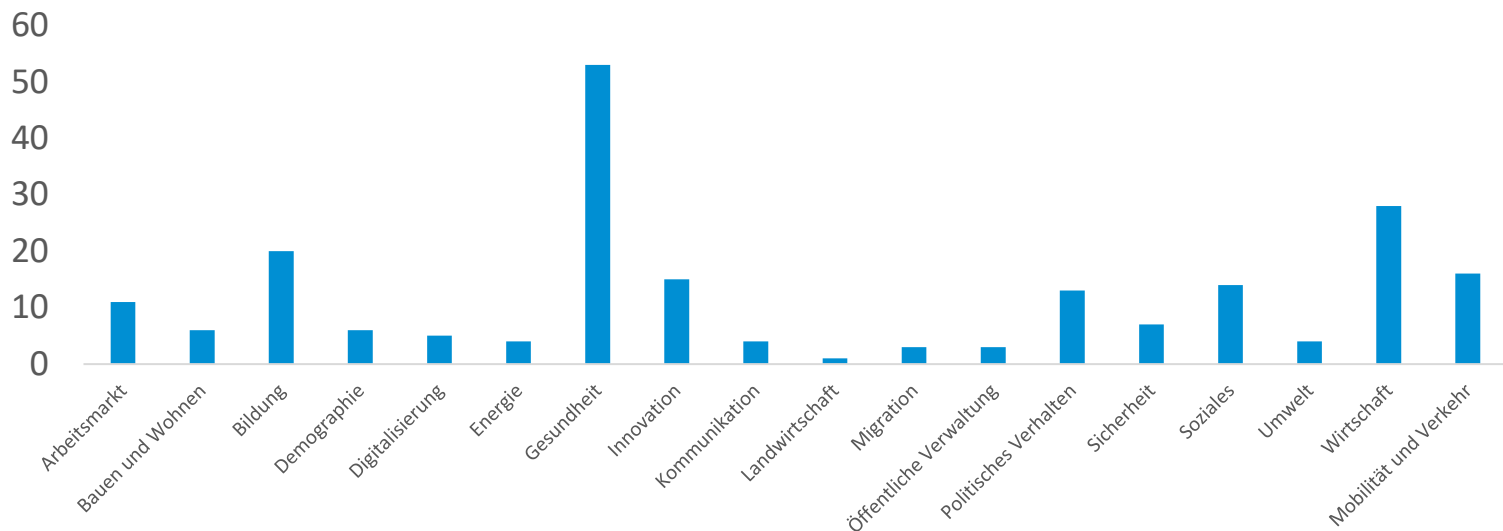
→ Frage ist, ob diese sinnvoll* für hochwertige Forschung genutzt werden können, weil

- ... man nicht weiß, welche Daten es überhaupt wo gibt
- ... die Anzahl der Merkmale gering ist
- ... Daten (open data Bereich) durch die Anonymisierung nicht mehr verknüpfbar sind
- ... Zeitreihen fehlen oder Daten veraltet sind
- ... Daten aus rechtlichen und anderen Gründen nicht für die Forschung zur Verfügung stehen
- ... Zugänge oft individuell erarbeitet werden müssen anstelle eines Rechtsanspruches
- ... Metadatenbeschreibungen fehlen / FAIR compliance?

Abhilfe durch AMDC, FOG, EU (DGA, DSA, European Data Spaces ...) → aber noch sehr langer Weg

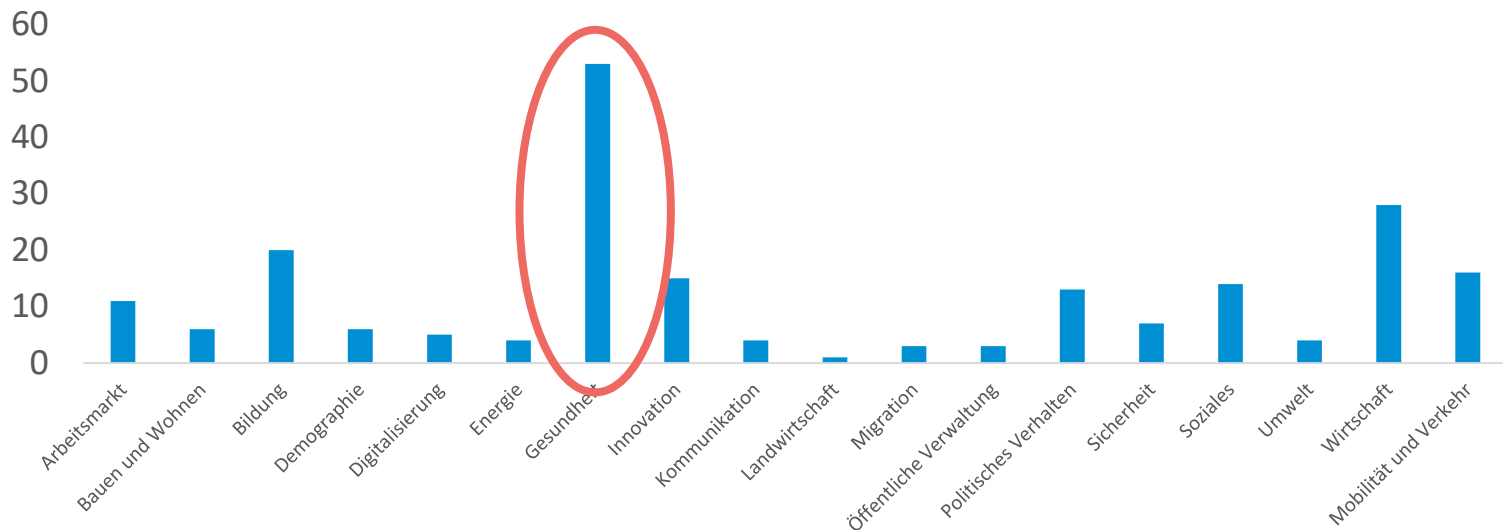
* Sinnvoll heißt, nicht nur einen Sachverhalt feststellen („xx Personen sind an Covid verstorben“), sondern Evidenzen für Zusammenhänge und kausale Ursachen produzieren

Register in Österreich



Basierend auf König T. et al. 2020. "Verfügbarkeit und Bedarf von administrativen Registern für die wissenschaftliche Forschung". IHS, <https://irihs.ihs.ac.at/id/eprint/5575/>

Register in Österreich



Basierend auf König T. et al., "Verfügbarkeit und Bedarf von administrativen Registern für die wissenschaftliche Forschung. IHS

Möglichkeiten zur Registerforschung in Österreich

- Seit Mitte 2022: Öffnung des **Austrian Microdata Center – angesiedelt bei der Statistik Austria** - für die Forschung
- Forscher*innen steht damit erstmals der Zugang zu den Mikrodaten der Statistik Austria offen
- Aktuell sind 93 Datensätze zugänglich; laufend kommen neue hinzu
- Einspielung von Registern, die über eine andere gesetzliche Grundlage “Forschungsorganisationsgesetz” freigegeben werden
 - Im Moment nur Daten des BMBWF (und zwar alle)
 - Andere Ministerien: unterschiedliche Signale: einige sind offener und es wird was kommen; andere verweigern sich total (Bereich Gesundheit)

Beschreibung und Details

Allgemeine Daten

Typ:	Register
Grundgesamtheit:	Österreichische Wohnbevölkerung mit ihren formalen Bildungsabschlüssen
Kurzbeschreibung:	Bildungsstand (höchste abgeschlossene Ausbildung) der in Österreich lebenden Personen im Alter von 15 Jahren und älter.
Themen:	Schulen, Hochschulen, Bildungsstand, Bildungsabschlüsse, Bildung
Statistische Einheiten:	Personen
Verknüpfungsschlüssel:	Personen_ID

Zeitliche und räumliche Abdeckung

Jahr:	2015 - 2020
Unterbjährig verfügbar:	Nicht unterjährig verfügbar
Periodizität:	Laufend (zumindest jährlich)
Datenerhebung:	Nicht laufend
Berichtszeitraum/Stichtag:	Stichtag des Bildungsstandes: 31. Oktober

Mikrodatenkatalog der Statistik Austria

<https://www.statistik.at/amdc-data/#!/product>

Methodische und rechtliche Informationen

Methodisches

Verwendete Verfahren:	Imputation
Weitere Methoden:	Bei allen über 15-jährigen Personen, für die im Bildungsstandregister keine Bildungsabschlüsse bekannt sind, wird ein Abschluss imputiert. Als Basis der Imputation dient eine vergleichbare Masse an Personen mit bekannten Bildungsabschlüssen aus dem Bildungsstandregister, wobei die Imputation mehrstufig erfolgt,

Codebook

Merkmalsname	Merkmalsbeschreibung	Merkmalsausprägung	
EDU_HAB_NAT	Höchste abgeschlossene Ausbildung	-1	entfällt unter 15 Jahren
		1	Pflichtschule (inkl. Personen ohne Pflichtschulabschluss)
		2	Lehre
		3	BMS
		4	AHS
		5	BHS
		6	Kolleg
		7	Akademie
		8	Hochschule
EDU_HAB_ABFELD2013_NAT	Ausbildungsfeld der höchsten abgeschlossenen Ausbildung seit 2015 (ISCED-Fields 2013) - nat.	-1	Entfällt
		0	Allgemeine Bildung
		1	Pädagogik
		2	Geisteswissenschaften
		3	Sozialwissenschaften
		4	Wirtschaft, Verwaltungswissenschaften
		5	Naturwissenschaften
		6	Informatik und Ingenieurwesen
		7	Ingenieurwesen
		8	Landwirtschaft
		9	Gesundheitswissenschaften
		10	Dienstleistungen
		99	Nicht bekannt
EDU_HAB_ABFELD_NAT	Ausbildungsfeld der höchsten abgeschlossenen Ausbildung bis 2014 (Fields of Education and Training 1999) - nat.	-1	Entfällt
		0	Allgemeine Bildung
		1	Erziehung
		2	Geisteswissenschaften
		3	Sozialwissenschaften
		4	Naturwissenschaften
		5	Ingenieurwesen
		6	Landwirtschaft
		7	Gesundheitswissenschaften
		8	Dienstleistungen
		9	Nicht bekannt
abdat_nat	Abschlussdatum		
zuord3	Zuordnungsqualität des Abschlusses	1	Zuordnung der Ausbildungsabschlüsse
		2	andere Zuordnungsmerkmale (Statistical Match, Gemeindekennziffer)
		3	Schätzung

Variablen

Bildungsabschlüsse

5 Variablen

EDU_HAB_NAT - Ebene des höchsten Abschlusses

EDU_HAB_ABFELD2013_NAT - Ausbildungsfeld der höchsten abgeschlossenen Ausbildung seit 2015 (ISCED-Fields 2013) - nat.

Klassifizierung

EDU_HAB_ABFELD_NAT - Ausbildungsfeld der höchsten abgeschlossenen Ausbildung bis 2014 (Fields of Education and Training 1999) - nat. Klassifizierung

abdat_nat - Abschlussdatum

zuord3 - Zuordnungsqualität des Abschlusses

Bildungsabschlüsse internationale Klassifikation

1 Variable

EDU_HAB_ABFELD2013_INT - Ausbildungsfeld der höchsten abgeschlossenen Ausbildung seit 2015 (ISCED-Fields 2013) - int.

Klassifizierung

Laufende Ausbildung

2 Variablen

EDU_WLAU - Ausbildungstyp der besuchten Ausbildung

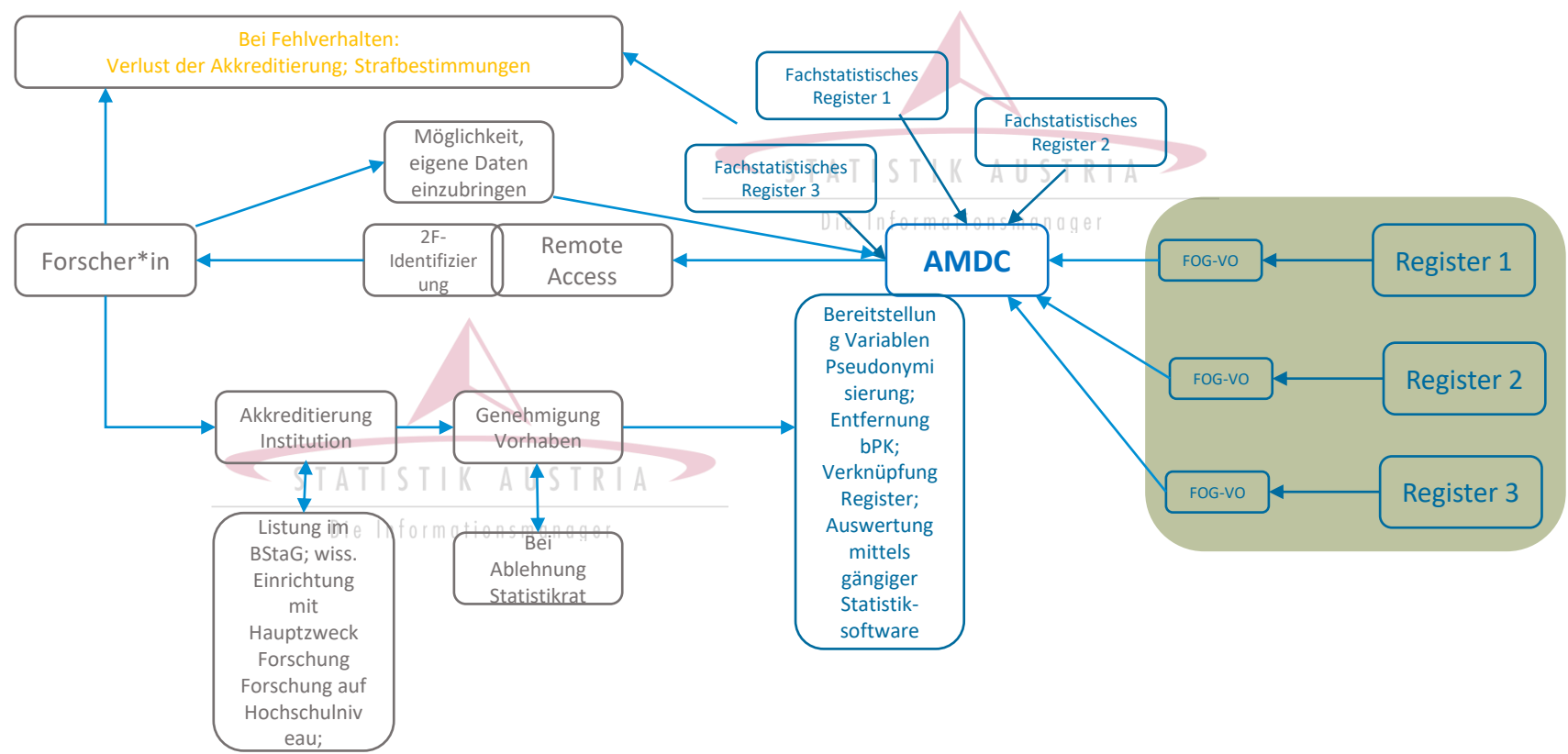
EDU_WLAU_ISCEDF_2013 - Ausbildungsfeld der besuchten Ausbildung

- Akkreditierung der Institution (erleichtert für im BStG genannte, Z.B. Unis).
 - Aktuell sind 50 Institutionen akkreditiert (davon 12 Unis in Österreich)
 - Bei im BStG genannten: Prüfung technischer Voraussetzungen zu Datenschutz
 - Bei anderen: Nachweis, dass Institution Forschung auf hohem Niveau betreibt
- **Jedes Forschungsprojekt muss eigens beantragt werden**
 - Braucht es die gewünschten Daten dafür? Sind die gewünschten Merkmale verfügbar?
 - Formaler Prozess: es erfolgt keine inhaltliche Qualitätsprüfung (z.B. ist die Frage sinnvoll? Ist sie state of the art?)
 - Kostenvoranschlag. Kosten sind nicht unerheblich (mehrere 10k EUR) . Günstiger als eigene Erhebungen, aber zusätzliche Finanzierung wohl notwendig (größere Drittmittel oder Aufträge)
- **Outputkontrolle: lässt sich aus dem Output auf Individuen schließen plus weitere Datenschutzmaßnahmen**

Was bekommt man?

- Nur die für die Forschungsfragen benötigten Daten werden zusammengestellt (ganzer Datensatz nicht zugänglich)
- Daten werden pseudonymisiert (direkte Identifier werden entfernt)
- Remote Access → Daten verbleiben bei der Statistik Austria
- ! Daten aus unterschiedlichen Datensätzen können verknüpft werden !
- Das Einbringen eigener Daten ist möglich (müssen mit bPK ausstattbar) sein
- Auswertung ist mit gängigen Statistikprogrammen möglich

Austrian Microdata Center (AMDC)



Wie ist das AMDC im Kontext einzuordnen?

- AMDC ist bedeutender Schritt für die Forschung in Österreich
 - Von Zero zum oberen Feld in Europa → DK, SE, FI sind weiter als AT, aber z.B. im Vergleich zu DE weit voraus
- Vielzahl von relevanten Daten liegen aber nicht bei der St.at, sondern in Registern bei Ministerien → hier gibt es noch kaum Schritte
- Notwendige Lernkurve auf allen Seiten
 - Weitere Register der Ministerien ins AMDC bringen
 - Aufbau operativer Expertise bei der Statistik Austria
 - Aufbau von Expertise in der Wissenschaft im Umgang mit den AMDC Datensätzen
 - Wie geht Politik mit den Ergebnissen um? → Evidence-based policy making / Wissenschaft-Politik Interface → noch viel Luft nach oben

Weitere Dynamiken auf Europäischer Ebene

Wesentliche Dynamiken kommen vor allem aufgrund europäischer Initiativen zustande

- Werden in den nächsten Jahren die Entwicklung in Österreich bestimmen
- Einige europäische Länder sind schon sehr weit (z.B. Finnland, Skandinavien)
- Politik in Österreich ist sehr zögerlich → kein Ministerium fühlt sich wirklich dafür zuständig
- Transparenz- und Datengovernance-Gesetze bedeuten nicht nur eine simple Freigabe von Daten, sondern Verwaltungspraktiken müssen neu gedacht werden; viele Ressourcen erforderlich (finanziell, personell)
 - Aber auch enorme Kosten bei der Nicht-Nutzung von Daten

Europäische Datenstrategie (2020)

- **Europäische Datenstrategie** sieht die Einrichtung gemeinsamer europäischer Datenräume vor
- Schaffung eines *single markets* für Daten in Europa
- *Ziele:*
 - *Innovation, Wachstum, Wettbewerb*
 - *Wissenschaft soll auch vom Datenzugang profitieren*

Viele Initiativen, Direktiven ...

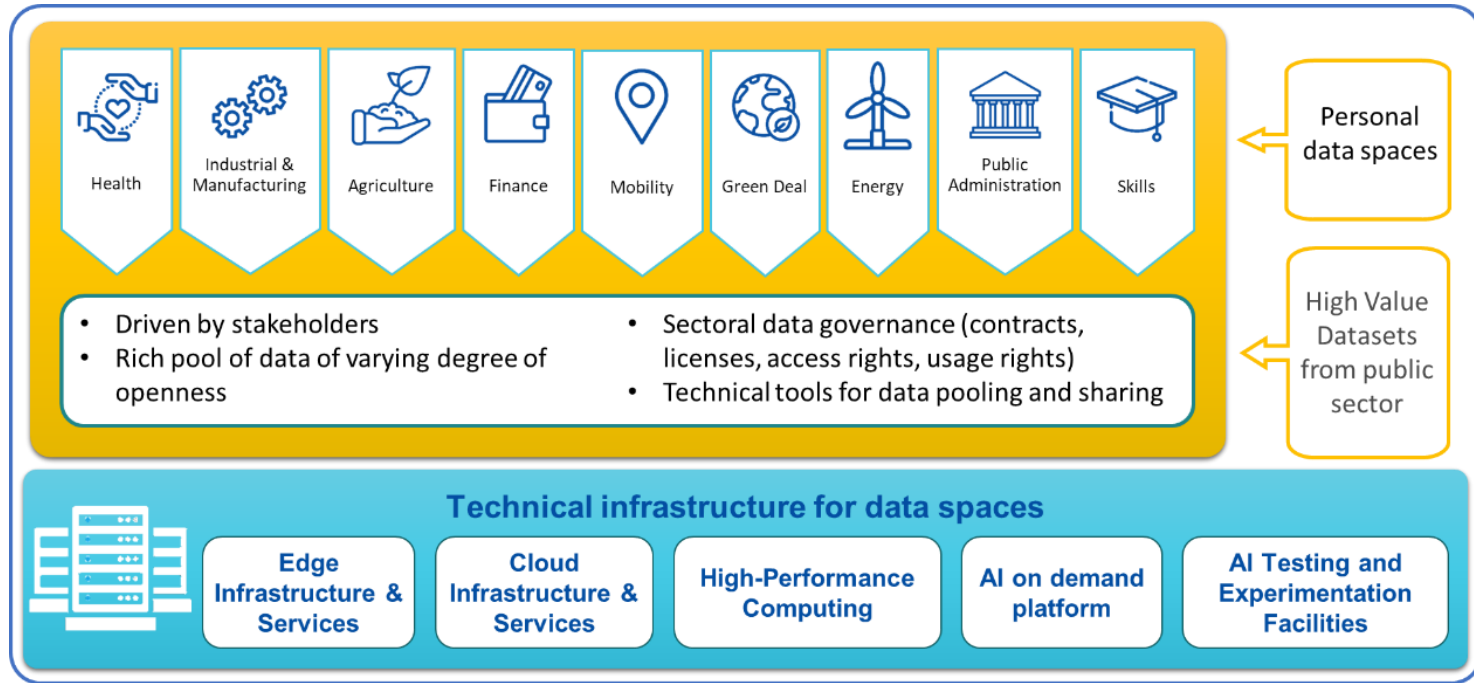
Data Strategy, AI Strategy, EOSC, Gaia-X, Digital Services Act, Data Governance Act, European Health Data Space ...

European Strategy for Data

A common European data space, a single market for data



Gemeinsame Europäische Datenräume



EU Data Governance Act

Data Governance Act → Anreize für die Weiterverwendung von Daten öffentlicher Stellen
EU will v.a. wirtschaftliche Impulse setzen (AI), aber insbesondere auch wissenschaftliche Forschung erleichtern

- Einrichtung einer sektorübergreifenden *zentralen Informationsstelle (ZI)* = *one stop shop*
 - Überblick über die verfügbaren Datenbestände der öffentlichen Hand
 - automatische Vermittlung von Anfragen für die Weiterverwendung von Daten
 - Standardisierung und Harmonisierung von Daten / Metadaten
 - Benennung zuständiger Stellen: Einrichtungen für bestimmte Sektoren, die den Datenzugang regeln (Beratung, technische Unterstützung, ...)

Aber: der DGA beinhaltet keine Verpflichtung, dass und welche Daten öffentliche Stellen zur Verfügung stellen müssen.

Tritt mit 24. September 2023 in Kraft; ZI noch unbekannt (wohl BRZ)

European Health Data Space

- Erster Health Data Space, der angegangen wird
- Vorschlag gerade bei den Mitgliedsstaaten
- EHDS stark geprägt vom finnischen Vorbild (*findata*)
- Skepsis in einigen Mitgliedsstaaten

Kernpunkte:

- Kontrolle / Zugang von Personen zu ihren elektronischen Gesundheitsdaten (EHR); grenzüberschreitende Infrastruktur
 - Sekundärnutzung der Daten: Forschung, Politikgestaltung, Statistik, öffentliche Verwaltung, Qualitätssicherung (grenzüberschreitend)
 - Im Moment sehr liberale Regelung, die auch die Sekundärnutzung der Daten für Unternehmen ermöglicht
 - Zugang erfolgt über eine/mehrere Zugangsstellen; entscheiden über Anträge; verarbeiten die Gesundheitsdaten für die Sekundärnutzung
- Wäre radikaler *regime change* in Österreich und ein Meilenstein für die Gesundheitsforschung

„Jeder Fortschritt ist es Wert dafür zu kämpfen“

„...hoch wer mas nimma gewinnen...“
(Toni Pfeffer)

Follow us on X / Twitter: @registerforsch1
Plattform Registerforschung: [https://www.registerforschung.at/
kontakt@registerforschung.at](https://www.registerforschung.at/kontakt@registerforschung.at)

Quellen und Links

- Halla, M. 2020. How Did Administrative Data Change Economic Research Output - A Descriptive Analysis. Annual Meeting of the Austrian Economic Association 2020. <https://sites.google.com/site/mhalla1980/>
- Austrian Microdata Center der Statistik Austria: <https://www.statistik.at/services/tools/services/amdc-mikrodaten-fuer-die-wissenschaft>
- Bundesgesetz über die Bundesstatistik (Bundesstatistikgesetz 2000): <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10006095>
- Plattform Registerforschung: <https://www.registerforschung.at/>
- Registerforschungsseiten des BMBWF: <https://www.bmbwf.gv.at/Themen/Forschung/Forschung-in-%C3%96sterreich/Strategische-Ausrichtung-und-beratende-Gremien/Leitthemen/Registerforschung.html>
- European Data Spaces: <https://digital-strategy.ec.europa.eu/en/library/building-data-economy-brochure>
- European Health Data Space: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en
- Data Governance Act: <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act>

Hinweis: Dieser Vortrag ist eine adaptierte Version eines Vortrags vom 26.1.2023 bei einer Veranstaltung der LBG/WU Executive Academy