

KOF Lab-Sektion Research Software Engineering And Economic Data Gestartet

Journal Article

Author(s):

Bannert, Matthias

Publication date:

2023-12

Permanent link:

<https://doi.org/10.3929/ethz-b-000651863>

Rights / license:

[In Copyright - Non-Commercial Use Permitted](#)

Originally published in:

KOF Analysen 2023(4)

KOF Analysen

Spezialanalysen:

Matthias Bannert

KOF Lab-Sektion Research Software Engineering
and Economic Data gestartet

2023, Nr. 4, Winter – SA1

IMPRESSUM

Herausgeberin

KOF Konjunkturforschungsstelle, ETH Zürich

Postadresse: LEE G 116, Leonhardstrasse 21, 8092 Zürich

Telefon: +41 44 632 42 39 E-Mail: kof@kof.ethz.ch Website: www.kof.ethz.ch

Chefredaktion KOF Analysen

Michael Graff

Spezialanalyse

Autor

Matthias Bannert:

KOF Lab-Sektion Research Software Engineering and Economic Data gestartet

Lektorat und Schlussredaktion

Günther Greulich

Layout

Vera Degonda, Nicole Koch

ISSN 1662-3517

Copyright © ETH Zürich, KOF Konjunkturforschungsstelle, 2023

Weiterveröffentlichung (auch auszugsweise) ist nur mit Bewilligung des Herausgebers und unter Quellenangabe gestattet.

KOF LAB-SEKTION RESEARCH SOFTWARE ENGINEERING AND ECONOMIC DATA GESTARTET



Matthias
Bannert

Zusammenfassung: Der Grossteil der soziökonomischen Daten wurde innerhalb des letzten Jahrzehnts generiert. In der Folge sind viele Forschungsfelder mit einer nie da gewesenen Datenflut konfrontiert. Die *KOF Lab-Sektion Research Software Engineering and Economic Data* möchte als Anlaufstelle für Forscher:innen mit einem programmatischen Ansatz zur Datenanalyse dazu beitragen, die Chancen zu nutzen die dieser Datenschatz für die Forschung bietet. Dabei soll nicht nur Wissen vermittelt werden und eigene Forschungsprojekte vorangetrieben werden, sondern auch einer breiteren Gruppe an Forschenden die Möglichkeit gegeben werden, mit KOF-Daten zu arbeiten.

Abstract: The majority of social and economic data has been created within the last decade. In turn, many fields of research are confronted with an unprecedented wealth of data. In an effort to make the most of this opportunity, *Research Software Engineering and Economic Data (RSEED) at KOF Lab* has set out to become a competence hub for research projects that embrace a data and programming driven approach to analysis. In addition, the new section looks to leverage KOF's unique datasets by enabling a broader group of researchers to work with the data.

JEL-Classification: A23, C63, C80.

Keywords: Research Software Engineering, Economic Data, Teaching, Open Source, Open Data.

In diesem Jahr startete das KOF Lab als neuer Forschungsbereich an der Konjunkturforschungsstelle der ETH, um neue Ansätze ökonomischer Forschung zu testen und zu implementieren. Mit den Sektionen *Mittel- und Langfristszenarien*, *Ungleichheit und öffentliche Wirtschaft* sowie der Sektion *Public and Social Policy Design* startete *Research Software Engineering and Economic Data* als eine von vier Sektionen.

Einführung

Ziel der neuen Sektion ist es, empirisch Forschenden eine Anlaufstelle und einen Anknüpfungspunkt zur aktiven Open Source und Open Data Community zu geben. Insbesondere möchte die Sektion Forscher:innen ausserhalb der Informatik helfen, das Potential von Quellcode als transparenten

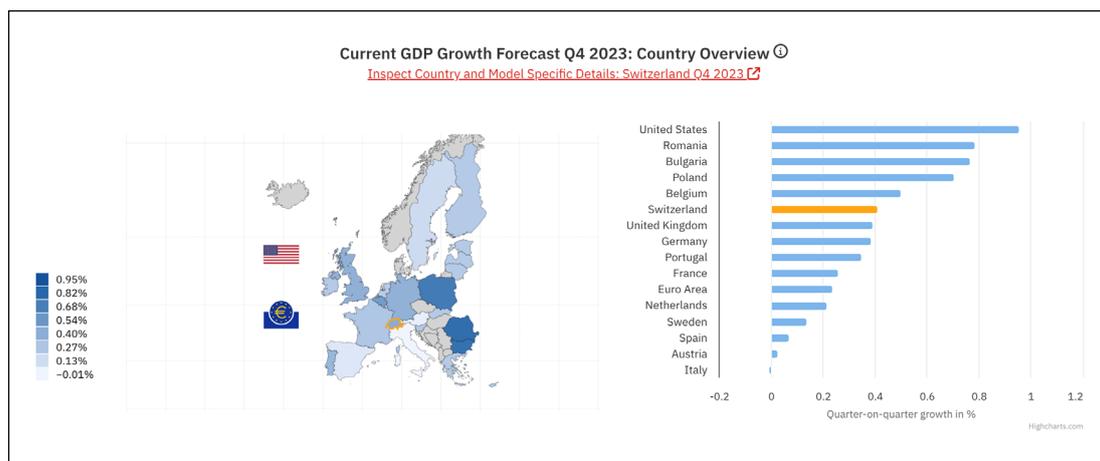
und reproduzierbaren Kommunikationskanal für methodische Fortschritte und Analysen zu nutzen. Zudem soll die neu geschaffene Sektion mehr Forschenden ermöglichen, mit den einzigartigen Daten der KOF zu arbeiten.

Ökonomische Daten

Bereits seit den 1970er Jahren erhebt die KOF Daten von Unternehmen, um ein Bild über die konjunkturelle Lage in der Schweiz erstellen zu können. Heutzutage beantworten mehr als 10 000 Schweizer Unternehmen regelmässig die Konjunkturumfragen der KOF. Diese Informationen sind nicht nur wertvoller Input für kurzfristige Prognosen und Indikatoren wie etwa das KOF-Konjunkturbarometer, sondern auch für die akademische Forschung.

Zeitreihen und Prognosen

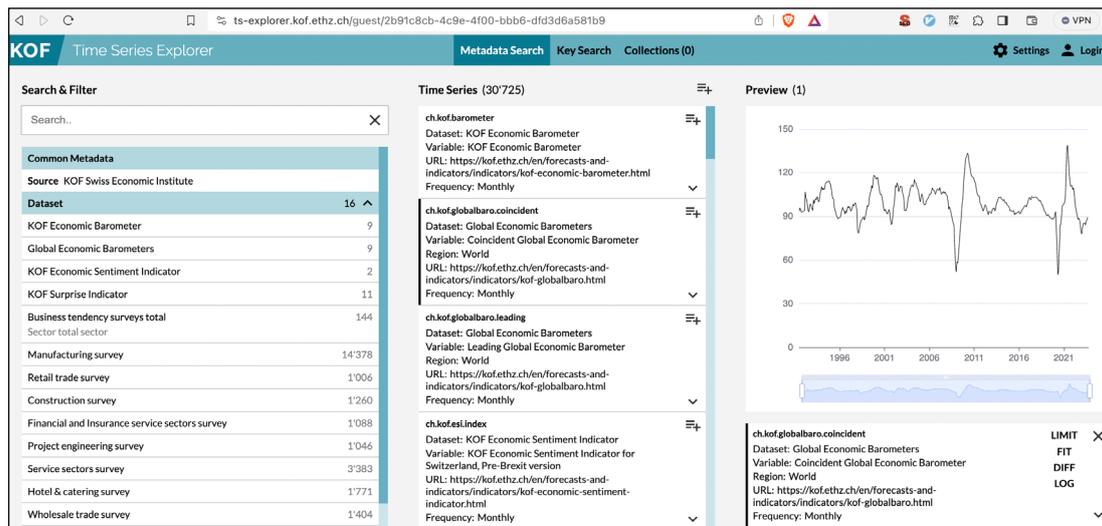
Durch die Arbeit mit den Daten können methodische Erkenntnisse gewonnen und neue Methoden und Modelle validiert werden. So entstand beispielsweise mit dem KOF Nowcasting Lab¹ eine Echtzeit-Testplattform für Schätzungen des BIP, mit der nicht nur Modelle und Informationsgewinn von Daten evaluiert werden, es wurde auch ein Ort, an dem sich Interessierte ein täglich aktualisiertes Bild zur wirtschaftlichen Lage vieler Länder machen können.



Das Nowcasting Lab ist eine Echtzeit-Testplattform für Schätzungen des BIP.

Ergänzend zu den bereits durch KOF-Forscher:innen vorarbeiteten Informationen möchte die KOF auch Forscher:innen ausserhalb der KOF den Zugang zu ihren ökonomischen Längsschnittdaten vereinfachen. Damit werden nicht nur Daten, deren Erhebung und Verwaltung der Daten öffentlich finanziert wird, breiter genutzt, sondern es kommt auch den wertvollen Beiträgen der teilnehmenden Personen und Firmen eine höhere Wertschätzung zu. Mit der Revision der KOF-Zeitreibendatenbank (Bannert, 2017) wurde die technische Grundlage für diese Art der Datennutzung geschaffen. Heute stehen bereits mehr als 30 000 Zeitreihen zur Schweizer Wirtschaft zur Verfügung.

¹ <https://nowcastinglab.org/map>.



Mit dem KOF Time Series Explorer können Interessierte die KOF Zeitreihendatenbank durchsuchen.

Forschung mit Mikrodaten

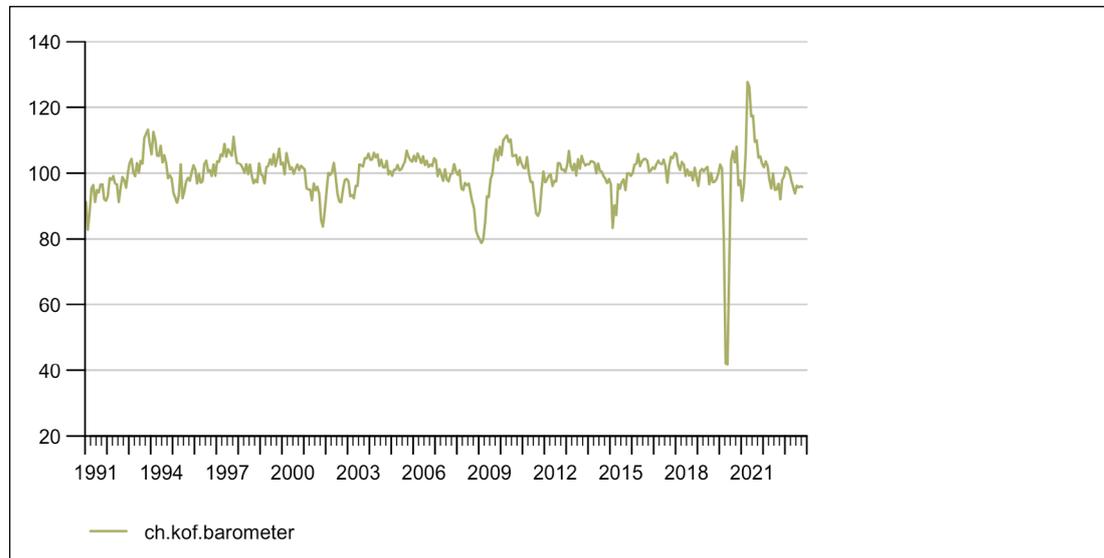
Während die in der Öffentlichkeit stark wahrgenommenen Prognosen und Indikatoren meist auf Zeitreihendaten basieren, sind die Mikrodaten der KOF, die über 50 Jahre zurückreichen, für die akademische Forschung besonders wertvoll. Etliche Arbeiten des Forschungsbereich Innovationsökonomik, der bereits seit Jahrzehnten die Strukturen der Schweizer Wirtschaft untersucht und Einblicke in Wachstumsprozesse und Strukturwandel bietet, basieren beispielsweise auf dem Unternehmenspaneldatensatz der KOF.

Um anderen Forscher:innen ebenfalls die Arbeit mit dieser einzigartigen Datenbasis zu ermöglichen und gleichzeitig den erhöhten Anforderungen des Datenschutzes gerecht zu werden, wurde das KOF Micro Data Center geschaffen. Diese Einrichtung ermöglicht es externen Forscher:innen, vor Ort an der KOF auf speziell isolierter Hardware statistische Analysen mit anonymisierten Daten durchzuführen. Im Rahmen der neuen Sektion *Research Software Engineering und Economic Data* soll auf diese Pionierarbeit aufgebaut werden und sowohl der Datenfundus als auch die Möglichkeiten, etwa durch Aussenstellen mit spezieller Hardware, erweitert werden.

Software Publikationen – Beitrag durch Forschungssoftware

Quelloffene Programmiersprachen, insbesondere die Sprachen R, Python und Javascript, zählen zu den wichtigsten Innovationen der Statistik in den letzten 30 Jahren. Diese durch technische Fortschritte im Hardwarebereich, Reproduzierbarkeit und Transparenz getriebene Entwicklung hat nicht nur zu einer Vielzahl von nützlichen, lizenzkostenfreien Softwarepaketen geführt, sondern hat auch mit dem Quellcode selbst einen neuen Kommunikationskanal geschaffen, der zu einem wichtigen Standard geworden ist. Die Sektion *Research Software Engineering und Economic Data* knüpft an diese Entwicklung an und möchte Forschenden helfen, ihre Ideen in State-of-the-art-Software zu verpacken und so den Einfluss ihrer Arbeit vergrößern. Zum einen werden auf diese Weise Beiträge und Methoden abstrahiert und wiederverwendbar, zum anderen bietet die Sektion Forscher:innen einen Einstiegspunkt in eine lebendige und globale Open Source und Data Science Community. Der folgende Auszug zeigt in einem einfachen Beispiel, wie mit der Sprache R Zeitreihendaten über die Applikationsschnittstelle (API) der KOF-Zeitreihendatenbank abgefragt und visualisiert werden können. Das Beispiel verwendet die an der KOF entwickelten R Bibliotheken *kofdata* (Bannert and Thoeni, 2022a) und *tstools* (Bannert and Thoeni, 2022b).

```
library(kofdata)
# just for viz
library(tstools)
tsl <- get_time_series("ch.kof.barometer")
tsplot(tsl)
```



Auf diese Weise können Abfragen parametrisiert und automatisiert werden, etwa für regelmässige Aktualisierungen, Publikationen oder Warnmeldungen. Die Grafik zeigt das KOF-Konjunkturbarometer im Herbst 2023.

Lehre

Auch wenn die oben diskutierte Entwicklung zahlreiche Fortschritte erst ermöglicht hat, stellt sie auch eine Herausforderung für viele Forscher:innen dar (Bannert, 2022). Neben feldspezifischer Expertise erfordern akademische Data Science-Projekte solide Statistik- und Programmierkenntnisse. Forschende wie Curriculae moderner Studiengänge sind gefordert, mit einer hohen Dynamik Schritt zu halten.

Hacking for Science

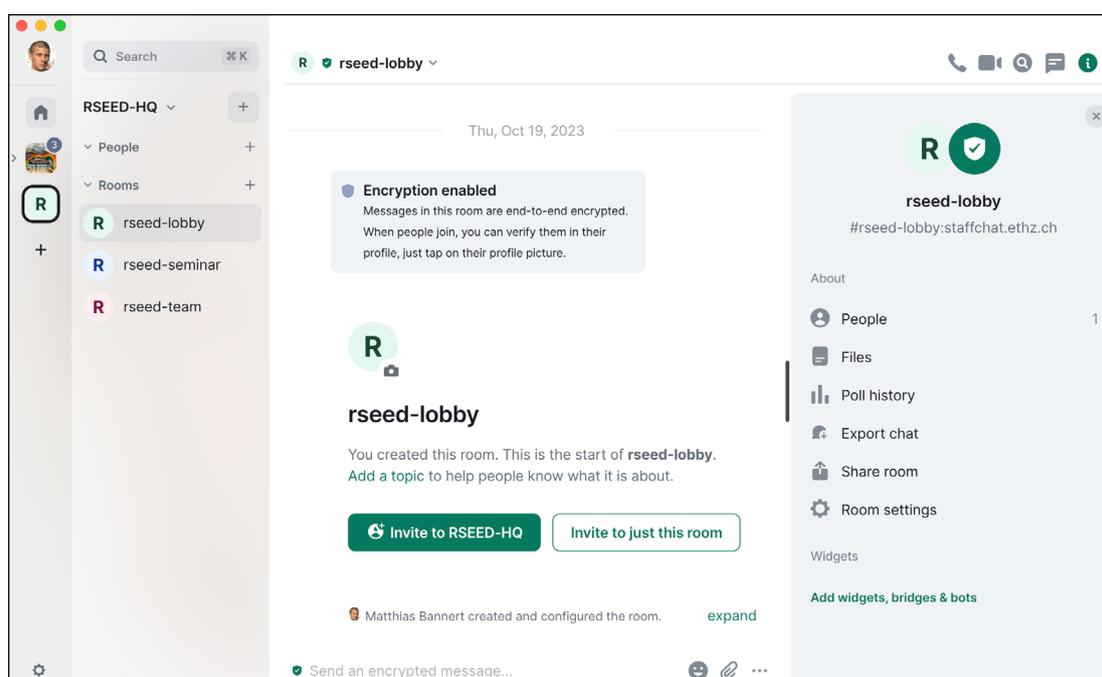
Daher wurde der Kurs *Hacking for Science* 2020 erstmals im Doktoratsprogramm angeboten und erfreute sich in den Folgejahren einem immer breiter werdenden, interdisziplinären Publikum. Nach einer organisatorisch bedingten Pause in diesem Jahr soll der Kurs im Jahr 2024 wieder als Teil des Angebotes der neuen KOF Lab-Sektion zurückkehren. Darüber hinaus sind speziell für studentische Mitarbeiter:innen (Research Assistants) entwickelte Schulungen geplant.



RSE-Seminar im Jahr 2024

Im Geist der Hacking for Science-Kurse soll ebenso ein Research Software and Data Engineering-Seminar angeboten werden, das Doktorierenden der ETH, aber auch anderen Interessierten, offensteht. Um die für dieses Feld entscheidende Interaktivität in der Lehre und im Austausch zu fördern, stellen Referierende ihr Thema bereits im Vorfeld in einem Blog Post vor, der bereits vor dem Vortrag zur Verfügung steht. Auf diese Weise steht während des Vortrags selbst detaillierter Austausch und Feedback im Vordergrund.

Zur Nachbereitung steht mit Matrix eine moderne Chat-Plattform zur Verfügung, die ebenfalls das Open Source-Konzept der Kurse verkörpert. Als *offenes und sicheres Netzwerk für dezentrale Kommunikation* können ETH Studierende und Mitarbeitende die Homeserver der ETH nutzen, gleichzeitig können Externe aber an den Diskussionen in ETH Matrixkanälen über öffentliche oder privat gehostete Homeserver teilnehmen.



Element Client für die Matrix Chat Plattform.

Mit dem frei verfügbaren Matrix Client *element* können Interessierte den Chat-Räumen des RSEED Chat Space beitreten und sich an allgemeinen sowie seminarspezifischen Diskussionen beteiligen. Detaillierte Information zur Verwendung von Matrix und Element befinden sich auf der Webseite der Sektion.

Projekte

Neben dem Daten-, Softwareentwicklungs- und Lehrangebot initiiert die KOF Lab-Sektion RSEED Forschungs- und Communityprojekten und beteiligt sich daran. Vorherige erfolgreiche Kollaborationen wie das SNF Projekt *Digital Lives* oder das in der Folge bearbeitete NRP 77 Projekt *Digital Transformation*² zeigen das Potenzial interdisziplinärer Zusammenarbeit zwischen Ökonomie, Data Science

² Gemeinsam mit dem KOF-Forschungsbereich «Schweizer Arbeitsmarkt» und Prof. Rafael Lalive d'Epinay-Gisin von der Universität de Lausanne wurde das Projekt «Was Arbeitskräfte wollen: Strategien für die Jobsuche» auf einer Online-Jobplattform im Rahmen des nationalen Forschungsprogramms 77 initiiert und durchgeführt. Weitere Information zum Projekt findet man auf <https://www.nfp77.ch/>.

und Data Engineering. Um diesen vielsprechenden, interdisziplinären Ansatz gezielt zu fördern, geht das KOF Lab in der ökonomischen Forschung neue Wege: Expertise in den Bereichen Data Science, Data Engineering und Softwareentwicklung soll mit dem KOF Lab in einen Forschungsbereich eingebunden werden, anstatt wie bisher meist über die Zentralen Dienste, also die Administration. Durch die Einbindung von Softwareentwickler:innen als Forscher:innen und die damit einhergehende Beteiligung am akademischen Betrieb mit Publikationen, Lehre sowie Konferenzen und Seminaren wird die KOF für Forscher:innen mit Expertise in Zukunftsfeldern wie Computational Statistics, Machine Learning, Data Science und Data Engineering attraktiver. So entstehen zudem neben den Publikationen in ökonomischen Fachzeitschriften auch Software- und Datenpublikationen.

Im laufenden Jahr 2023 gab es bereits eine Kollaboration mit der *Openwashdata* Community in einem durch das *Open Research Data Program* der ETH geförderten Projekt. Das Projekt hilft Forschenden im Water, Sanitation und Hygiene (WASH)-Sektor, bereits publizierte Daten in maschinenlesbarer Form nach *FAIR principles* (Wilkinson 2016) aufzubereiten und für andere nutzbar zu machen. Dabei übernehmen Mitarbeitende des Projekts nicht nur die Datenaufbereitung mittels Open Source-Programmiersprachen, sondern dokumentieren und teilen auch ihre Arbeitsschritte und Prozesse. Auf diese Weise gibt das *Openwashdata*-Projekt Expertinnen des WASH-Sektors einen Einstiegspunkt in die Werkzeuge und Workflows im Open Research Data (ORD)-Bereich.

Weiterführende Informationen

Website: www.rseed.ch

Matrix Chat Kanal: #rseed-lobby:staffchat.ethz.ch

Lehrveranstaltung Hacking for Science: <https://u.ethz.ch/dJnxB>

Quartalsweise informiert bleiben – Anmeldung zum Newsletter:



Literaturverzeichnis

- Bannert, Matthias (2017): Revision der KOF-Zeitreihendatenbank. *KOF Analysen* 2017 (3): 115–20. <https://doi.org/10.3929/ethz-b-000226855>.
- Bannert, Matthias (2022): Zauberer oder Zimmermann – Welche technischen Fähigkeiten braucht eine moderne Ökonom:in? *KOF Analysen* 2022 (1): 90–95. <https://doi.org/10.3929/ethz-b-000538838>.
- Bannert, Matthias und Severin Thoeni (2022a): *Kofdata: Get Data from the 'KOF Datenservice' API*. <https://CRAN.R-project.org/package=kofdata>.
- Bannert, Matthias und Severin Thoeni (2022b): *Tstools: A Time Series Toolbox for Official Statistics*. <https://github.com/mbannert/tstools>.
- Wilkinson, Mark D. et al. (2016): The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* 3. <https://doi.org/10.1038/sdata.2016.18>.