

BOOK OF ABSTRACTS



Journées suisses de la statistique

Schweizer Statistiktage

Giornate svizzere della statistica

Dis svizers de la statistica

Swiss Statistics Meeting

11./12. November 2019

Veranstaltungsort / Lieu du congrès

Lorzensaal Cham

Dorfplatz 3

6330 Cham

Key Note

Das grosse Rauschen

Stefan Pabst
W.I.R.E. (Web for Interdisciplinary Research & Expertise)
Think Tank for Business, Society & Life Sciences
Hallwylstrasse 22
8004 Zürich, Schweiz

Thesen zu einer datenbasierten Wirtschaft und Gesellschaft

Die Digitalisierung ist in aller Munde und wird gerade vom Kopf auf die Füße gestellt. Gleichzeitig wurde das Themenfeld durch seine Omnipräsenz in Medien und auf Konferenzen in den letzten Jahren immer unschärfer, was Entscheidungen zu konkreten Handlungsfeldern und Massnahmen erschwert – für Individuen, Unternehmen und öffentliche Institutionen.

In seinem Referat wird der Zukunftsforscher Stefan Pabst die Digitalisierung entschlüsseln und damit den Begriff wieder greifbarer machen, er wird differenziert zwischen klaren Entwicklungen und diffusen Spekulationen unterscheiden und er wird Handlungsempfehlungen formulieren, um die Chancen einer datenbasierten Welt durch einen souveränen Umgang – auch mit ihren Herausforderungen – nutzen zu können.

Lightning talk 1

Stupid Statistics?!

Servan Grüninger
EPFL/reatch - research and technology in Switzerland
Aarbergstrasse 95
Biel-Bienne, Schweiz

Kampagne gegen statistische Desinformation

Alle brauchen sie, kaum einer braucht sie richtig: Statistiken. Statistik ist die Basis für seriöse wissenschaftliche Analysen, für politische Entscheidungen, für ökonomische Berechnungen. Und mit der jüngsten Revolution im Bereich des maschinellen Lernens und der Digitalisierung hat sich die Bedeutung der Statistik als Disziplin noch einmal massgeblich erhöht.

Doch die wissenschaftliche und gesellschaftliche Bedeutung von Statistik steht in krassem Widerspruch zu ihrem Verständnis innerhalb von Wissenschaft und Gesellschaft. Auch in der Wissenschaftsvermittlung findet sie kaum einen Platz. Zu trocken, zu komplex scheint die Materie. Denn Statistik bietet weder atemberaubende Explosionen noch beeindruckende Maschinen.

Das ist gefährlich. Denn mit dem Mangel an statistischem Verständnis wächst die Gefahr von Missbräuchen. In den Wissenschaften ist die sogenannte "Reproduzierbarkeitskrise" eng mit statistischem Unwissen verknüpft. Und in der Politik werden Statistiken längst schamlos zu Stimmungsmache jenseits faktischer Grundlagen genutzt. Was tun?

Das Projekt «Stupid Statistics» möchte in der Bevölkerung und insbesondere unter Bevölkerungsgruppen, die kein statistisches Vorwissen besitzen, ein Grundverständnis für statistische Sachverhalte schaffen und dabei auf leichte und unterhaltsame Weise aufzeigen, welche Fragen und Überlegungen dabei eine Rolle spielen. Es geht nicht in erster Linie darum, mathematische Kenntnisse zu vermitteln, sondern ein Grundverständnis für das Lesen und Verstehen von Statistiken, statistischen Argumenten und den zugrundeliegenden Annahmen zu schaffen.

Um das zu erreichen, lanciert die wissenschaftliche Ideenschmiede «[reatch - research and technology in switzerland](#)» zusammen mit Partnern aus Wissenschaft, Politik und Wirtschaft eine Informationskampagne gegen statistische Desinformation. Basierend auf dem aktuellen Stand der Forschung und im Austausch mit Expert*innen aus den Kommunikationswissenschaften und dem Campaigning identifizieren wir jene Elemente, die (statistische) Falschinformationen viral machen und nutzen sie, um eine möglichst grosse Breitenwirkung für die Kampagne gegen Desinformation zu erzielen.

Zu diesem Zweck wird im Rahmen in regelmässigen Abständen eine "dumme Statistik" veröffentlicht, allgemein verständlich korrigiert und aktiv in den sozialen Medien verbreitet und beworben. Statistische Laien sollen auf spielerische Art lernen, falsche von verlässlichen Informationen zu erkennen. Die damit einhergehende Sensibilisierung für statistische Themen fördert die Informationskompetenz in der breiten Bevölkerung. Die Kampagne ist ein Pilotprojekt, um aufzuzeigen, wie effektive und breitenwirksame Wissenschaftskommunikation im digitalen Raum funktionieren kann.

Lightning talk 2

Estimating confidence interval for the change using partially overlapping samples

Swetlana Renner
Agroscope
Tänikon 1, Ettenhausen/Switzerland

The evaluation of the economic situation of farmers requires reliable information on changes of agricultural income. The major data source for the income monitoring in Switzerland is a longitudinal survey collecting the accountancy data from a random sample of commercial agricultural holdings, carried out by Agroscope.

Generally, the estimates based on samples are associated with uncertainty. Thus, simply comparing the estimated means of the income between two years might lead to an over-interpretation, especially in the case of large sampling variances. Better praxis requires providing the variance (or confidence intervals) of change in order to judge whether or not the observed differences statistically significant.

In contrast to the repeated cross-sectional surveys, in which different units are selected at different time points, a longitudinal survey collects data from the same sample elements on multiple occasions over time. In the case of cross-sectional survey, there is no overlap between samples in two occasions. Thus, samples are statistically independent. The survey collects data from the same farms on different years. survey some participants stay in the sample for two or more occasions, while some other members leave the sample and/or are replaced by the new respondents (depending on the sampling design, for example rotating or split panels). If samples include both paired (matched) and independent observations, they can be referred to as “partially overlapping” (Derrick et al. 2017).

In this contribution, we present the variance estimator of change for a complex survey with unequal sampling probabilities, taking into account the overlap between data from two periods (Qualité and Tillé, 2008). The proposed approach enables to calculate confidence intervals for the change, that can be provided additionally to the point estimate, and offers a useful tool to judge whether the difference of means between two periods is significant different from zero or not.

In the empirical example, we estimate the change in the average agricultural income per farm between the accounting years 2017 and 2016 using the random sample of 2100-2200 Swiss farms. Using the standard tests for comparing of two samples (t-test for independent samples or paired observations) might produce misleading results. If we consider the covariance of the overlapping sample in calculating the confidence interval for the change, we find that the estimated increase in the average agricultural income was statistically significant.

1) Derrick, B., Russ, B., Toher, D., & White, P., 2017. Test statistics for the comparison of means for two samples that include both paired and independent observations. *Journal of Modern Applied Statistical Methods*, 16(1), pp.137-157.

2) Qualité, L. and Tillé, Y., 2008. Variance estimation of changes in repeated surveys and its application to the Swiss survey of value added. *Survey Methodology*, 34(2), pp.173-181.

Lightning talk 3

Rethinking Human Well Being: a multidimensional exploration across Europe

Josue Aduna
Smart Social Software Studio
Strada Transilvaniei 13
Brasov, Romania

Human well-being is a complex construct that has been addressed from different

perspectives. Increasingly, contributions embrace the concept of flourishing to capture the breadth and depth of human life. This perspective usually relies on Aristotelian notions and includes a combination of positive *feelings* and psychosocial *functionings*, yet morality is seldom considered despite the strong moral component in Aristotle's account. This study explores the main constituents of well-being under a flourishing perspective that includes moral attitudes. Accordingly, cross-sectional data from the European Social Survey is analysed by performing Exploratory Factor Analysis with the statistical software R. A *eudaimonic* approach that includes moral attitudes is proposed and nine features are determined via literature review. The results reveal three irreducible dimensions of human well-being: hedonic, *eudaimonic* and moral. Thus, it is shown that current indicators of well-being fall short to provide a holistic picture of well-being and that the *eudaimonic* dimension provides a promising pathway to enhance the human condition, both individually and at a societal level.

1) Huppert, F.A. and So, T.T., (2013). *Flourishing across Europe: Application of a new conceptual framework for defining well-being*. *Social Indicators Research*, 110(3), pp.837-86.

2) Keyes, C. L. M. (2002). *The mental health continuum: From languishing to flourishing in life*. *Journal of Health and Social Behavior*, 43, 207222.

Lightning talk 4

Vom Zählen zum Erzählen: Ereignisse im Lebensverlauf analysieren

Tina Schmid
Statistik Stadt Zürich SSZ
Napfgasse 6
8001 Zürich, Schweiz

Das Beispiel der Transformation vom Zwei-Eltern- zum Ein-Elternhaushalt

In der amtlichen Statistik wird häufig gezählt: Die Bevölkerung, die Anzahl Geburten, Scheidungen oder Todesfälle. Jede/r von uns wird somit immer wieder gezählt. Einzelne Ereignisse im Lebensverlauf werden jedoch selten zueinander in Beziehung gesetzt, obwohl die Daten des Bevölkerungs- und des Gebäude- und Wohnungsregisters diese Möglichkeit bieten. Werden Personen oder Haushalte über mehrere Zeitpunkte betrachtet, lassen sich spannende Erkenntnisse über Veränderungen und Entwicklungen gewinnen.

Der vorliegende Beitrag zeigt dies am Beispiel der Transformation von Zwei-Eltern-Haushalten zu Ein-Elternhaushalten. In der Stadt Zürich gibt es knapp 50 000 Haushalte, in denen zwei Elternteile und mindestens ein minderjähriges Kind lebt. Knapp 500 Haushalte werden jedes Jahr zu Ein-Elternhaushalten.

Der Beitrag zeigt, welcher Elternteil bei den Kindern bleibt und welcher Elternteil unter welchen Bedingungen in der Familienwohnung bleibt – und wohin der andere Elternteil zieht. Analysiert wird auch die Wohnentfernung zwischen Kindern und dem Elternteil, der nach der Trennung nicht mehr bei den Kindern wohnt, sowie die Veränderung der finanziellen Situation aller Familienmitglieder. Danach wird aufgezeigt, welche technischen Voraussetzungen für die Analyse nötig sind.

Lightning talk 5

Role of Markets for the Early Warning of Outbreaks in Cattle Transport Networks

Beatriz Vidondo
University of Bern, Veterinary Public Health Institute
Schwarzenburgstrasse 161
Bern, Switzerland

Early Warning in Networks

We consider the spreading of a phenomenological disease on the cattle transport network of Switzerland over a 4-year period that involves 49,497 agricultural holdings (nodes) and 927,289 trade contacts (links). We use this empirical and time-resolved network as a substrate of the interaction between different types of holdings. In particular, we distinguish between farms and markets. The latter are the places of periodic events where many animals are gathered for trade. Based on our numerical simulations, we confirm that the potential of disease spreading from markets to other holdings is high. Then, we study the predictive power for early warning of disease outbreaks by selecting different subsets of nodes for surveillance: (i) markets and (ii) farms with a high risk of infection. We show that a detection probability of 86% can be achieved by monitoring all 137 markets in the network.

An additional number of 250 farm sentinels -- selected according to their risk - - would need to be placed under surveillance so that the probability of first hitting a farm is at least as high as the probability of first hitting a market. Combining markets and 1000 farms with highest risk of infection will lead to a detection probability of 99%.

Atelier 1: Pilotprojekte der BFS-Dateninnovationsstrategie

Arealstatistik "Deep Learning" (ADELE)

Claudio Facchinetti
Bundesamt für Statistik
Espace de l'Europe 10
2010 Neuchâtel, Schweiz

Im Rahmen der BFS-Dateninnovationsstrategie wurden fünf besonders inspirierte Pilotprojekte ausgewählt, die in den letzten zwei Jahren die Grenzen der Machbarkeit in der öffentlichen Statistik getestet haben. Im Zentrum des Interesses stand die Anwendung komplementärer Analysemethoden (z. B. prädiktive Analytik mittels erweiterter Statistik, Datenwissenschaft und/oder maschinelles Lernen) auf BFS-interne Primärdatenquellen sowie bereits verknüpfte identifizierbare Sekundärdatenquellen, die in der aktuellen Statistikproduktion des BFS bereits verwendet werden. Es handelt sich hierbei um Pilotprojekte, die das Potenzial haben die BFS-Statistikproduktion zu erweitern oder zu ergänzen. Das Projekt Arealstatistik Deep Learning (ADELE) ist ein davon.

Für die langfristige Raumbewertung trägt dieses Projekt zum Aufbau des nötigen Knowhows beim Einsatz von Technologien der künstlichen Intelligenz bei. Auf lange Sicht ist das Ziel eine Teilautomatisierung der Luftbildinterpretation bei der Identifizierung und Klassifizierung von Veränderungen.

Atelier 1: Pilotprojekte der BFS-Dateninnovationsstrategie

Automatisierung der NOGA-Kodierung (NOGauto)

Claude Macchi
Bundesamt für Statistik
Espace de l'Europe 10
2010 Neuchâtel, Schweiz

Im Rahmen der BFS-Dateninnovationsstrategie wurden fünf besonders inspirierte Pilotprojekte ausgewählt, die in den letzten zwei Jahren die Grenzen der Machbarkeit in der öffentlichen Statistik getestet haben. Im Zentrum des Interesses stand die Anwendung komplementärer Analysemethoden (z. B. prädiktive Analytik mittels erweiterter Statistik, Datenwissenschaft und/oder maschinelles Lernen) auf BFS-interne Primärdatenquellen sowie bereits verknüpfte identifizierbare Sekundärdatenquellen, die in der aktuellen Statistikproduktion des BFS bereits verwendet werden. Es handelt sich hierbei um Pilotprojekte, die das Potenzial haben die BFS-Statistikproduktion zu erweitern oder zu ergänzen. Das Projekt Automatisierung der NOGA-Kodierung (NOGAuto) ist ein davon.

Automatisierung der Kodierung der wirtschaftlichen Tätigkeit von Unternehmen anhand von Methoden des maschinellen Lernens, die an bereits im BFS vorhandene Daten angewandt werden (z. B. Erhebungsdaten, Beschreibung im Handelsregister, Schlüsselwörter, Erläuterungen der Nomenklaturen).

Atelier 1: Pilotprojekte der BFS-Dateninnovationsstrategie

Machine Learning SoSi (ML_SoSi)

Thomas Ruch
Bundesamt für Statistik
Espace de l'Europe 10
2010 Neuchâtel, Schweiz

Im Rahmen der BFS-Dateninnovationsstrategie wurden fünf besonders inspirierte Pilotprojekte ausgewählt, die in den letzten zwei Jahren die Grenzen der Machbarkeit in der öffentlichen Statistik getestet haben. Im Zentrum des Interesses stand die Anwendung komplementärer Analysemethoden (z. B. prädiktive Analytik mittels erweiterter Statistik, Datenwissenschaft und/oder maschinelles Lernen) auf BFS-interne Primärdatenquellen sowie bereits verknüpfte identifizierbare Sekundärdatenquellen, die in der aktuellen Statistikproduktion des BFS bereits verwendet werden. Es handelt sich hierbei um Pilotprojekte, die das Potenzial haben die BFS-Statistikproduktion zu erweitern oder zu ergänzen. Das Projekt Machine Learning» SoSi (ML_SoSi) ist ein davon.

Gruppierung typischer prospektiver Verlaufsmuster bezüglich Leistungsbezügen im System der sozialen Sicherheit und Erwerbsarbeit sowie Schätzung der Gruppenzugehörigkeit durch Nutzung individueller Merkmale und retrospektiver Verlaufsdaten unter Anwendung von Methoden des maschinellen Lernens.

Atelier 1: Pilotprojekte der BFS-Dateninnovationsstrategie

Plausibilitätsprüfung mit Machine Learning (Plausi ++)

Christian Ruiz
Bundesamt für Statistik
Espace de l'Europe 10
2010 Neuchâtel, Schweiz

Im Rahmen der BFS-Dateninnovationsstrategie wurden fünf besonders inspirierte Pilotprojekte ausgewählt, die in den letzten zwei Jahren die Grenzen der Machbarkeit in der öffentlichen Statistik getestet haben. Im Zentrum des Interesses stand die Anwendung komplementärer Analysemethoden (z. B. prädiktive Analytik mittels erweiterter Statistik, Datenwissenschaft und/oder maschinelles Lernen) auf BFS-interne Primärdatenquellen sowie bereits verknüpfte identifizierbare Sekundärdatenquellen, die in der aktuellen Statistikproduktion des BFS bereits verwendet werden. Es handelt sich hierbei um Pilotprojekte, die das Potenzial haben die BFS-Statistikproduktion zu erweitern oder zu ergänzen. Das Projekt Plausibilitätsprüfung mit Machine Learning (Plausi++) ist ein davon.

Algorithmen aus dem maschinellen Lernen bieten vielversprechende Anwendungsmöglichkeiten um die Plausibilitätsprüfungen im BFS zu erweitern, zu beschleunigen und gleichzeitig die Datenqualität zu steigern. Die Datenlieferanten erhalten dabei nicht nur eine automatisierte Fehlermeldung, sondern auch die Erklärung – ein Feedback – dazu.

Atelier 2: Statistik als Steuerungsinstrument

Neue Wohnformen und Wohnraumbedarf. Simulationen möglicher Trends für den Kanton Waadt

Reto Schumacher
Statistique Vaud
Rue de la Paix 6
1014 Lausanne

Haushalts- und Wohnraumszenarien basieren einerseits auf Bevölkerungsszenarien, andererseits auf Annahmen zum zukünftigen Verlauf des Wohnverhaltens. Statistique Vaud hat sich in den letzten Jahren auf ein Modell gestützt, welches die in der Vergangenheit beobachteten Entwicklungen des Wohnverhaltens mehr oder weniger fortschreibt. Dabei wurde der Fokus vor allem auf die alters- und geschlechtsspezifische Verteilung der in einem Privathaushalt lebenden Personen nach deren Haushaltsgrösse gelegt (z.B. 15% der 30 bis 34-Jährigen Frauen leben alleine, 30% leben zu zweit, usw.). Kombiniert mit Annahmen zur altersabhängigen Entwicklung von Personen in Kollektivhaushalten, ergeben sich aus verschiedenen Hypothesen zum Verlauf obiger Verteilung verschiedene Szenarien zur Anzahl und Grösse zukünftiger Privathaushalte. Deren Verteilung nach Haushaltstypen (z.B. Paarhaushalte, Familienhaushalte usw.) und Wohnungsgrösse erfolgt erst in weiteren Rechenschritten. Aufgrund der Sequenzierung dieser Rechenschritte eignet sich dieses klassische *membership rate model* jedoch schlecht zur Simulation von Verläufen in welchen alternative Wohnformen eine bedeutendere Rolle als heute spielen. Gemäss unserer Mission, verschiedenste Zukunftsverläufe zu betrachten, haben wir das Modell neu konzipiert. Der erste Rechenschritt basiert neu auf einer alters- und geschlechtsabhängigen Verteilung, welche Haushaltsgrösse und Haushaltstyp kombiniert (insgesamt 12 Kategorien), und aus Daten aus dem kantonalen Bevölkerungsregister und aus der Strukturhebung berechnet wird. Simulierte Verschiebungen innerhalb einer oder mehrerer Altersklassen zwischen diesen Kategorien können dann als Änderungen im Wohnverhalten interpretiert werden. In unserem neuesten Bericht¹⁾ zum zukünftigen Wohnraumbedarf im Kanton Waadt haben wir insbesondere Trends zu mehr Wohngemeinschaften unter jungen Erwachsenen, zu getrennt lebenden Paaren (d.h. in verschiedenen Wohnungen), sowie zu einfachen Mehrgenerationenhaushalten (Hochbetagte mit einem - auch schon älteren - Kind) simuliert. Für den stark wachstumsgeprägten Kanton Waadt hat sich dabei gezeigt, dass solche Trends den Wohnbedarf nicht bedeutend beeinflussen werden können.

1) Martin, M., Bologna, C., R. Schumacher et al. (2018). *Besoins en logement à l'horizon 2040*. Lausanne : Statistique Vaud.

Atelier 2: Statistik als Steuerungsinstrument

Schülerprognosen Basel-Stadt

Tobias Erhardt
Statistisches Amt des Kantons Basel-Stadt
Binningerstrasse 6
Basel, CH

Prognosen für das Erziehungsdepartement Basel-Stadt

Die Abteilung Raum und Anlagen des Erziehungsdepartements Basel-Stadt (ED) ist mit der Planung des Schulraums für die Stadt Basel betraut. Dabei steht sie vor der Herausforderung, den Schulraum für künftige Schülerinnen und Schüler (SuS) planen zu müssen, ohne die künftigen Schülerzahlen zu kennen. Das Statistische Amt Basel-Stadt beliefert das ED mit Prognosen zur künftigen Anzahl Schülerinnen und Schüler (SuS) nach Schulstufen und –standorten. Diese Schätzungen sollen dem ED die Planung des künftigen Schulraumbedarfs erleichtern.

Seit dem Schuljahr 2000/2001 prognostiziert die grossräumige Schülerprognose die Schülerzahlen in öffentlichen Schulen des Kantons vom Kindergarten bis zum Ende der Gymnasialzeit. Für die Detailplanung des Schulraums bestand jedoch seitens ED der Bedarf für eine kleinräumige Schülerprognose. Diese wird seit 2019 erstellt und prognostiziert die Zahl der SuS für 15 vom ED vorgegebene Perimeter, die ersten 8 Schulstufen (also Kindergarten und Primarschule) und 5 Schuljahre (2019/20-2023/24). Die Prognose liefert auch Schätzungen von Prognoseintervallen, womit die Unsicherheit bezüglich den künftigen Schülerzahlen besser in die Planung miteinbezogen werden kann.

Das Prognosemodell basiert auf der Annahme, dass uns die Ein- und Übertrittsquoten der Vergangenheit Informationen zu den künftigen Ein- und Übertritten liefern können. Wir schätzen diese Quoten mithilfe eines Panel-Regressionsmodells. Zur Evaluation der Prognosequalität wurden die Schätzergebnisse der Ein- und Zweijahresprognosen anhand des Jahres 2018/2019 getestet. Die Einjahresprognosen treffen die Schülerzahlen mit einer mittleren Abweichung von weniger als 4 SuS, die Zweijahresprognosen haben eine mittlere Abweichung von weniger als 6 SuS. Für diese Prognosehorizonte sind auch die geschätzten Prognoseintervalle von zufriedenstellender Qualität.

Verbesserungspotential besteht bei der Berücksichtigung von städtischen Entwicklungsgebieten, sowohl in der Abbildung des Ex-Post-Verlaufs als auch in der Prognose. Weiter wird es künftig möglich sein, Vorhersageintervalle aus den empirisch beobachteten Prognoseabweichungen zu schätzen, was die Verlässlichkeit der Konfidenzintervalle weiter steigern sollte.

Atelier 2: Statistik als Steuerungsinstrument

Statistiques de la justice et de la sécurité. Compléments indispensables aux statistiques officielles

Daniel Fink, Christophe Koller, Luisella Demartini
Université de Lausanne, ESEHA,
Service de probation du canton du Tessin

Depuis 2012, l'association ESEHA développe des activités statistiques dans le domaine de la justice et de la sécurité dont les données récoltées sont hébergées sur le portail CHStat.ch, en complément à celles réalisées par l'OFS dans le domaine des statistiques de la criminalité. ESEHA réalise l'enquête sur l'efficacité de la justice pour le compte des tribunaux cantonaux et du Tribunal fédéral, enquête reliée aux travaux de la CEPEJ du Conseil de l'Europe. De plus, depuis 2015, ESEHA développe, pour le compte du Réseau national de sécurité (collaboration cantons-Confédération), une statistique des forces sécurité publique et privée en Suisse. Ainsi, un corpus important de sources externes aux statistiques officielles sont disponibles pour les autorités via CHStat. Il est prévu de combiner ces sources diverses afin de de construire un monitoring de la chaîne pénale qui débouchera sur un Atlas suisse de la justice et de la sécurité. Notre contribution vise à présenter quelques-unes de ces données et le projet de statistique générale de la chaîne pénale.

Atelier 2: Statistik als Steuerungsinstrument

Gemeinsame Gesundheitsregion Basel

Matthias Minke
Statistisches Amt Basel-Stadt
Binningerstrasse 6
4001 Basel, CH

Entwicklung und Implementierung eines Planungs- und Wirkungsmodells für die Gemeinsame Gesundheitsregion beider Basel

Am 09.02.2019 haben die Stimmbevölkerungen der Kantone Basel-Stadt und Basel-Landschaft den Staatsvertrag über die gemeinsame Planung, Regulation und Aufsicht in der Gesundheitsversorgung angenommen. Ziel des Staatsvertrags ist unter anderem eine gemeinsam erstellte Analyse des Versorgungsbedarfs in den beiden Kantonen sowie die Etablierung einheitlicher Kriterien für die Aufnahme auf die gemeinsame Spitalliste und der Vergabe von Leistungsaufträgen.

Sowohl die Bedarfsanalyse als auch das Planungs- und Wirkungsmodell sollen mit statistischen Daten abgestützt sein. Aus diesem Grund haben das Gesundheitsdepartement Basel-Stadt und die Volkswirtschafts- und Gesundheitsdirektion Basel-Landschaft das Statistische Amt Basel-Stadt mit der Entwicklung eines entsprechenden Modells beauftragt. Die Entwicklung dieses Modells erfolgt in enger Abstimmung mit den Auftraggebern.

Die Aufgabe des Statistischen Amts Basel-Stadt besteht darin eine Reihe von Datenquellen über die Leistungserbringung und -inanspruchnahme aus dem Gesundheitsbereich aber auch Bevölkerungsdaten zusammenzuführen und modellgeleitete Aussagen herzuleiten. Eine an die gesundheitsökonomische Literatur angelehnte Regressionsanalyse bildet den Ausgangspunkt für eine Bedarfsprognose, welche mit unterschiedlichen Szenarien der Bevölkerungsentwicklung unterlegt wird. Die so ermittelte Bedarfsbandbreite liefert einen wichtigen Input für das Planungsmodell, welches, gestützt auf einen Algorithmus, die Spitallandschaft bei einer bedarfsgerechten Planung nach einheitlichen Kriterien simuliert. In einem weiteren Projektschritt werden die Ergebnisse des Planungsmodells in eine Wirkungsanalyse einfließen, die neben der Inanspruchnahme auch die Kosten und Finanzierung von Gesundheitsleistungen abbildet.

Perspektivisch werden wichtige Teile des Planungs- und Wirkungsmodells in einer Web-Applikation eingebettet sein, sodass die Auftraggeber eigenständige Berechnungen vornehmen können, ohne selbst über vertieftes Wissen über die dahinterliegenden Methoden haben zu müssen.

Atelier 3: Wege in der Methodik...

«Small Area Estimation» von Mietpreisen in der Stadt Zürich

Andreas Papritz
Statistik Stadt Zürich SSZ
Napfgasse 6
8001 Zürich, Schweiz

Die Daten der Strukturerhebung des Bundesamts für Statistik (BFS) werden zur Beantwortung vieler Fragestellungen verwendet. Oft müssen bei solchen Auswertungen Aussagen nicht nur für die gesamte Zielpopulation (z. B. Wohnbevölkerung eines Kantons) gemacht werden, sondern auch für Teile derselben (z. B. Personen in bestimmten Gemeinden). Die Stichprobengrösse in diesen «Domains» ist dann oft zu klein, um mit Standardverfahren (Horvitz-Thompson-Schätzer, HT) die erforderliche Genauigkeit zu erreichen. Durch «Pooling» der Daten mehrerer Jahre kann die Stichprobengrösse in den «Domains» zwar erhöht werden, aber dadurch verringert sich die zeitliche Auflösung, mit der Entwicklungen verfolgt werden können.

Für kleine «Domains» können Schätzungen statt mit dem HT-Schätzer mit «Small Area Estimation»-Verfahren (SAE, 1) berechnet werden. Neben modell-unterstützten Domain-Schätzern, basierend auf verallgemeinerten Regressionsschätzern (GREG), werden dazu modellbasierte Verfahren («Area»- und «Unit»-Level-Modelle) verwendet. In beiden Ansätzen wird die Zielgrösse einer Erhebung durch ein statistisches Modell mit Daten verknüpft, die in Registern vorhanden sind. Damit lässt sich die Genauigkeit von Schätzungen in vielen Fällen verbessern, es besteht aber auch die Gefahr, dass die Schätzungen verzerrt werden.

In der Schweiz setzen statistische Fachstellen SAE-Methoden noch nicht routinemässig ein. Das BFS hat im Rahmen des Projekts [Experimental Statistics](#) methodische Grundlagen erarbeitet. Um Vor- und Nachteile von SAE-Methoden im Hinblick auf deren praktische Anwendung weiter auszuloten, hat SSZ in einer Pilotstudie über Mietpreise in der Stadt Zürich die Genauigkeit zweier SAE-Methoden («Domain-GREG»-, Fay-Herriot-Schätzer FH) mit dem HT-Schätzer verglichen. Die Aufgabe bestand darin, die durchschnittliche Miete von 3- und 4-Zimmer-Wohnungen in den 12 Stadtkreisen und rund 200 statistischen Zonen zu schätzen. Entgegen den Erwartungen war der Domain-GREG-Schätzer für die Stadtkreise nicht genauer als der HT-Schätzer. Die FH-Methode schätzte sowohl für die Kreise wie für die statistischen Zonen die mittleren Mietpreise am genauesten. Mit den FH-Schätzer können die Mietpreise zudem für statistische Zonen vorhergesagt werden, die mit der Erhebung gar nicht erfasst wurden. Allerdings sind die Vorhersagefehler dann beträchtlich. Im Vortrag stellen wir die Resultate aus dieser Pilotstudie im Einzelnen vor.

1) Rao, J. & Molina, I. 2015. *Small Area Estimation*. 2. Auflage. Wiley

Atelier 3: Wege in der Methodik...

Resultate der Pilotprojekte der BFS-Dateninnovationsstrategie: Ermittlung des Potenzials von Kleingebietsschätzung im Rahmen der Beschäftigungsstatistik (BESTA)

Francis Saucy
Bundesamt für Statistik

Im Rahmen der BFS-Dateninnovationsstrategie wurden fünf besonders inspirierte Pilotprojekte ausgewählt, die in den letzten zwei Jahren die Grenzen der Machbarkeit in der öffentlichen Statistik getestet haben. Im Zentrum des Interesses stand die Anwendung komplementärer Analysemethoden (z. B. prädiktive Analytik mittels erweiterter Statistik, Datenwissenschaft und/oder maschinelles Lernen) auf BFS-interne Primärdatenquellen sowie bereits verknüpfte identifizierbare Sekundärdatenquellen, die in der aktuellen Statistikproduktion des BFS bereits verwendet werden. Es handelt sich hierbei um Pilotprojekte, die das Potenzial haben die BFS-Statistikproduktion zu erweitern oder zu ergänzen. Das Projekt Ermittlung des Potenzials von Kleingebietsschätzungen im Rahmen der Beschäftigungsstatistik (BESTA) ist ein davon.

Es wird geprüft, inwiefern die Methode der «Kleingebietsschätzungen» sich für die Beschäftigungsstatistik eignet und zuverlässige Schätzungen der Gesamtbeschäftigung und der Vollzeitäquivalente für Kantone, Grossstädte und im Stichprobenplan nicht berücksichtigte Wirtschaftszweige liefert.

Atelier 3: Wege in der Methodik...

Une fois au top, toujours au top ? Développement des athlètes dans le classement FIS de ski de fond

Anne Renaud et Hippolyte Kempf
Haute école fédérale de sport de Macolin HEFSM
Route principale 247
2532 Macolin, Suisse

La concurrence est rude entre les nations sur le circuit des compétitions internationales de ski de fond de la Fédération internationale de ski (FIS). La sélection des athlètes en début de carrière et le choix des athlètes pouvant participer aux compétitions sont tous deux complexes. Dans cette étude, nous avons comme objectif d'étudier le développement des athlètes représentant leurs nations dans les compétitions internationales en fonction de leur âge. On vise notamment à déterminer s'il y a des différences entre messieurs et dames, entre distance et sprint, et également entre les athlètes ayant atteint le plus haut niveau durant leur carrière (« once top 30 ») et ceux n'ayant pas atteint ce niveau. Les points FIS (distance et sprint) ont été choisis comme mesure de performance, et différents modèles ont été définis pour expliquer la performance en fonction de l'âge. Les résultats indiquent que les athlètes ayant atteint le plus haut niveau avaient globalement également de bons résultats en début de carrière. Les carrières individuelles des meilleurs athlètes à Pyong Chang 2018 montraient cependant que plusieurs d'entre eux faisait partie des meilleurs seulement à partir d'env. 20 ans et que la stabilité des performances était plus grande en distance qu'en sprint. Les résultats pourraient intéresser la FIS et les fédérations nationales de ski de fond. Ils présentent en effet de nouvelles connaissances sur les points FIS et un outil pour comparer des athlètes avec les meilleurs, en fonction de leur âge.

Atelier 3: Wege in der Methodik...

Anonymisierung von Aggregatdaten

Rolf Schenker
Statistik Stadt Zürich
Napfgasse 6
8001 Zürich

Wenn Tabellen zu viel kreuzen

Statistische Ämter verfügen über Einzeldaten, die nur so publiziert werden dürfen, dass keine Rückschlüsse auf einzelne (natürliche oder juristische) Personen möglich sind. Oft werden statistische Kennzahlen in Form von Kreuztabellen publiziert. Bei Kreuztabellen mit feingliedrigen Klassifizierungen sind die ausgewiesenen Gruppen zum Teil aber so klein, dass Rückschlüsse möglich wären, beispielsweise wenn in einer Gemeinde nur eine Bäckerei existiert. Dann müssen die Werte von Hand anonymisiert werden, damit eine Veröffentlichung möglich ist. Wenn die Gruppierungen – wie Branchen, Berufe oder geografische Gliederungen – hierarchisch aufgebaut sind, wird die Anonymisierung einer Tabelle zu einer echten Knobelei.

Diese Arbeit manuell zu erledigen ist aufwendig und braucht hohe Konzentration. Doch ein von Statistik Stadt Zürich entwickelter Algorithmus führt die Anonymisierung auch in hierarchisch gegliederten Tabellen automatisch durch. Dies verringert die für die Anonymisierung benötigte Zeit und die Wahrscheinlichkeit eines Fehlers deutlich.

Das Referat zeigt die Vorgehensweise des Algorithmus auf und erklärt, welche Bedarfe zur Anonymisierung damit abgedeckt werden und welche noch offen sind.

Atelier 4: Interaktive Informationsvermittlung

Echtzeitdaten am Abstimmungssonntag – automatisierte Prozesse und die Zusammenarbeit von Bund und Kantonen als Erfolgsfaktoren für ein innovatives Projekt

Madeleine Schneider, Corinne Straub
Bundesamt für Statistik
Espace de l'Europe 10
2010 Neuchâtel, Schweiz

Daniel Bierer,
Statistisches Amt des Kantons Zürich
Schöntalstrasse 5
8090 Zürich

Die mobile App Votelfo als Ausgangspunkt für einen Modernisierungsschub und die Umsetzung des Once-only-Prinzips in der Abstimmungsstatistik

Bei der Abstimmung vom 10. Februar 2019 kam erstmals die mobile App «Votelfo» zum Einsatz. Aufbauend auf der vom Statistischen Amt des Kantons Zürich entwickelten kantonalen App, hat die Bundeskanzlei zusammen mit dem Statistischen Amt des Kantons Zürich und dem Bundesamt für Statistik ein App-Projekt für die Bundesebene angestossen. Ziel war, sämtliche Abstimmungsinformationen (Erläuterungen und Resultate) von eidgenössischen und kantonalen Abstimmungen digital und aus einer Hand für mobile Geräte anzubieten. Die für die erfolgreiche Lancierung nötigen neuen Prozesse der Datenübermittlung, Resultate-Aufbereitung und automatisierte Diffusion werden dargelegt. Thema sind ebenfalls die Herausforderungen wie die kantonalen Staatskanzleien als Datenlieferanten ins Boot geholt werden konnten. Das gemeinsame Projekt der drei Partner ist nicht nur ein gelungenes Beispiel für die Zusammenarbeit verschiedener Verwaltungsstellen, sondern ebenso ein Anwendungsfall für die Umsetzung des Once-Only-Prinzips, indem die Kantone ihre Abstimmungsergebnisse künftig nur noch einmal an den Bund übermitteln müssen. Die kurze Laufzeit von 15 Monaten kann als besonderer Erfolg des Projekts angesehen werden. Last but not least profitiert die Öffentlichkeit von den im OGD-Portal zur Verfügung gestellten maschinenlesbaren Datensätzen, die parallel zur Veröffentlichung in der App allen gleichzeitig für eigene Anwendungen zur Verfügung stehen und von Medien und Wissenschaft bereits rege benützt werden.

Atelier 4: Interaktive Informationsvermittlung

Statistische Informationsvermittlung – Wo und wie positioniert sich die öffentliche Statistik in der digitalisierten Datengesellschaft?

Oliver Thommen Dombois, Nathalie Grillon
Statistisches Amt des Kantons Basel-Stadt
Binningerstrasse 6
4001 Basel, Schweiz

Mehr Sichtbarkeit dank neuer Visualisierungsformen?

Wie lässt sich die Aufmerksamkeit der breiten Öffentlichkeit in einer zunehmenden medialen Informationsflut gezielt auf statistische Informationen lenken? Müssen wir immer eine Datengeschichte erzählen, um unsere Zahlen interessant zu machen?

Wir werfen einen Blick zurück auf die ersten Gehversuche mit interaktiven Grafiken. Wir zeigen verschiedene Beispiele, wie wir Statistik-Ergebnisse mit einfachen Mitteln anschaulicher gestalteten, um dadurch ein grösseres Publikum zu gewinnen.

Erfahrungen und Reaktionen zeigen, dass die öffentliche Statistik ihre Rolle stärken kann, wenn Informationen durch visuelle Akzente hervorgehoben werden, ohne dass dadurch Faktentreue verlorenght. Das Erzählen von «Datengeschichten» sei dabei den Journalisten überlassen. Die öffentliche Statistik hingegen soll weiterhin vor allem ihre Rolle als verlässliche Partnerin und «Rohstofflieferantin» ausfüllen.

Atelier 4: Interaktive Informationsvermittlung

Interaktive Grafiken und Datenverarbeitung für ein webbasiertes Indikatorensystem

Oliver Zumbrunnen
Schweizerisches Gesundheitsobservatorium - Obsan
Espace de l'Europe 10
Neuchâtel, Schweiz

Die Umsetzung eines Indikatorensystems mit Statistiksoftware und JavaScript

Im Auftrag für das Bundesamt für Gesundheit BAG erstellt das Schweizerische Gesundheitsobservatorium das Monitoringsystem MonAM¹⁾. Die rund 120 Indikatoren zeigen im Web statistische Daten zu nichtübertragbaren Krankheiten und Suchtthemen aus verschiedenen Datenquellen. Für die Präsentation der Daten wurden dynamische und interaktive Grafiken in JavaScript entwickelt, die die Darstellung verschiedener Aspekte wie Zeit, Raumgliederung, Geschlecht und Alter ermöglichen. Dafür wird im Browser unter anderem die Open Source Bibliothek Data Driven Documents d3.js²⁾ verwendet, um aus den Daten benutzerfreundliche Grafiken, Webtabellen und Excel-Dokumente erstellen zu lassen. Aus diesen Erfahrungen werden Tipps für den Aufbau von automatisierten Indikatorensystemen gegeben. Weitere Themen sind die Lösung der Accessibility für grafische Daten und Veröffentlichung der Daten als Open Government Data.

1) Schweizer Monitoring-System Sucht und nichtübertragbare Krankheiten (MonAM)
<https://www.obsan.admin.ch/de/MonAM>

2) Data Driven Documents <https://www.d3js.org>

Atelier 4: Interaktive Informationsvermittlung

Make Information Accessible Again

Patrick Roelli, Léo Picard
Uthink, Université de Lausanne
Anthropole
Lausanne, Suisse

Le 21ème siècle ouvre une nouvelle ère de la digitalisation et du "big data" dans la vie quotidienne des citoyens et citoyennes Suisses.

Bien que la Suisse met à disposition un budget conséquent dans l'acquisition de données auprès de sa population, de ses entreprises et de ses institutions, l'accessibilité et la compréhension de ces données reste limitée pour une personne non initiée.

La quantité d'information augmentant de façon exponentielle, il nous appartient de trouver de nouvelles manières de présenter et d'exposer ces dernières aux citoyens et citoyennes suisses.

Cette transition vers une digitalisation globale engendre également de nouveaux problèmes tel que les "fake news". Il arrive également que certains partis politiques usent de données statistiques d'une manière biaisée pour influencer l'opinion de la population. Un des piliers fondamentaux de notre démocratie étant l'accès à une information de qualité et non biaisée pour toute personne capable de voter, il devient alors crucial de la rendre accessible, factuelle et visible au plus grand nombre.

Afin de répondre à ces défis, nous proposons d'utiliser des outils permettant de créer des applications interactives à partir de ces données. Celles-ci permettent une accessibilité accrue, une facilité d'utilisation et apporte un côté ludique, permettant aux personnes non initiées à l'analyse de ces données, d'interagir avec elles et d'en apprendre plus sur notre Suisse et ses enjeux.

Nous illustrons cette capacité dans deux projets basé sur Shiny, un package du langage de programmation R, permettant de créer des interfaces interactives à partir de données traitées dans ce même langage.

Nous avons préparé deux applications interactives consultables directement sur internet à travers un navigateur web.

Uthink Migration: En préparation d'une potentielle initiative visant à sortir de l'espace Schengen, cette application permet de tester ses propres croyances et affiche ensuite les données de l'OFS concernant la migration en Suisse afin d'avoir une idée plus claire sur le sujet. L'application est hébergée ici: https://hoohm.shinyapps.io/Uthink_Migration/

Uthink Vote blanc: Cette application propose d'explorer des données de votations fédérales Suisses entre 1980 et 2017 avec une proposition d'initiative: Donner une voix au vote blanc en Suisse. L'application est hébergée ici: https://hoohm.shinyapps.io/Uthink_Vote_blanc/

Atelier 5: Verknüpfung im Bundesamt für Statistik

Appariements de données à l'OFS

Magnus Fink, Regina Scartazzini
Bundesamt für Statistik
Espace de l'Europe 10
2010 Neuchâtel

Le principe de base d'un appariement de données consiste à relier plusieurs jeux de données existants, en général au moyen d'un identifiant unique tel que le numéro AVS, afin de générer une plus-value pour des analyses. L'Office fédéral de la Statistique (OFS) utilise les appariements dans le cadre de la production statistique, dans le but de décharger les répondants, de générer des synergies et d'éviter les redondances. Les jeux de données ainsi constitués offrent de nouvelles possibilités d'analyse sans qu'il soit nécessaire de mettre en place des relevés supplémentaires.

L'OFS effectue aussi des appariements de données pour des tiers tels que des universités et hautes écoles, des instituts de recherche, ou encore des Offices statistiques régionaux et cantonaux. Les utilisateurs de données déposent une demande contenant tous les éléments nécessaires pour statuer sur la faisabilité de l'appariement d'un point de vue technique, méthodologique et juridique.

L'OFS et les instances cantonales ont la compétence d'apparier des jeux de données, pour autant qu'au moins l'un d'entre eux ait été relevé dans le cadre de la loi sur la statistique fédérale et que le but ne se rapporte pas à une personne. Mais même si ces appariements restent dans un cadre exclusivement statistique/de recherche/de planification, ils sont délicats d'un point de vue de la protection des données. Les acteurs, les processus et les conditions cadres sont ainsi définis de manière détaillée dans les bases légales y relatives.

Atelier 5: Verknüpfung im Bundesamt für Statistik

Trajectoires sociales basées sur les données appariées de l'aide sociale, de l'assurance invalidité et de l'assurance chômage (SHIVALV)

Nora Meister
Bundesamt für Statistik
Espace de l'Europe 10
2010 Neuchâtel

En Suisse, l'intégration durable de divers groupes de personnes au sein du marché du travail s'avère être un défi considérable de la politique sociale pour les prochaines années à venir. Par la fragmentation des différents types de prestations du système de sécurité sociale, les individus y ont recours de manière diverse. Ainsi, suivant les cas, les personnes peuvent percevoir de manière simultanée ou successive l'une ou l'autre de ces prestations.

Traditionnellement, la statistique analyse la situation des bénéficiaires par prestation et par année. Dès lors, l'intérêt du set de données SHIVALV réside dans le fait qu'il se compose d'un ensemble d'informations qui lie trois domaines de prestations sociales sur plusieurs années : l'aide sociale, l'assurance-invalidité et l'assurance-chômage. Ces données contiennent donc un vaste potentiel d'analyses des trajectoires individuelles entre ces différents types de prestations ; elles permettent aussi des couplages avec des données supplémentaires qui rendent possible le suivi des individus au-delà de ces prestations et du système de sécurité sociale.

La présentation explique comment l'ensemble de données est structuré et montre quelles approches innovatrices peuvent être appliquées pour réduire la complexité des informations à disposition et pour les diffuser de manière attractive.

Atelier 5: Verknüpfung im Bundesamt für Statistik

LABB (Longitudinal analyses in the field of education)

Jacques Babel
Bundesamt für Statistik
Espace de l'Europe 10
2010 Neuchâtel

The introduction of the AHV number (Personal Identification Number) around 2010 in many official statistics of FSO leads to tremendous improvements in Switzerland in particular in the field of education. The FSO launched the LABB (Longitudinal analyses in the field of education) program in 2014 to take profit of the new potentialities and deal also with the numerous challenges.

LABB follows two axes: firstly, to produce harmonized longitudinal files and make them available to the research community; secondly, to communicate results on transitions, paths and key-indicators for the education domain, as well as between education and the labour market by the means of aimed publications.

The data cover:

- 1) The whole population of pupils or students at the considered degree (thus with N often above 100'000).
- 2) Entire and multi-cohorts starting from school year 2012.

The project links many data sources from FSO in the field of education with the population and household statistics, the structural survey, the CDC («compensation offices») register and the unemployment data from the SECO. In the near future, it will also linked data from the STATENT and the ESS and include data in connection with the incomes.

Until now, 3 parts of the educational system have been covered: the paths after compulsory school, the paths within the upper secondary and the paths in the tertiary degree. By now, the data permit to cover several transitions (the paths after compulsory school, the paths within the upper secondary, the paths between the upper secondary and the tertiary and the paths in the tertiary degree). The projects covers also the transition to the labour market after an upper secondary degree.

These data covering the whole post-compulsory school will be renewed each year, with new cohorts as well as a longer follow-up of the first cohorts.

Atelier 5: Verknüpfung im Bundesamt für Statistik

Career patterns in Swiss Higher Education Institutions (HEI)

Elena Zafarana, Mehmet Aksözen & Vera Herrmann
Swiss Federal Statistical Office
Espace de l'Europe 10
2010 Neuchâtel

First results of a longitudinal analysis

The Swiss Higher Education Area is characterised by three types of higher education institution: universities including the Federal Institutes of Technology in Zurich and Lausanne (UNI), universities of applied sciences (UAS), and universities of teacher education (UTE). All types are managed and financed differently, and have different characteristics and this might be reflected in different career patterns.

Data is based on the Swiss University Information System (SUIS), an institution in which representatives of the universities, the university cantons and the Swiss Confederation work together in order to compile statistics on higher education throughout Switzerland. The same abbreviation also stands for a database containing data on students, diplomas, staff (and even costs) in higher education.

The unique identifier for each and every person in the Swiss staff statistics at Higher Education Institutions was introduced in 2013. With 2018 as the latest year for data collection, the data comprises five years. The anonymised dataset of each staff member allows us to follow the development of each person's status.

The comparison of the changes in status may highlight career paths. At UNI career patterns are relatively well defined for many years; for this reason, first results of a longitudinal analysis on UNI staff with the focus on a gender perspective have been presented at the UNECE work session on gender statistics in May 2019. At UAS and UTE career patterns are less clearly defined; therefore, this analysis will integrate data on UAS/UTE staff and will shed light on potential differences in career patterns between the three types of HEI. In addition, the approach on FSO data collection, its method and parameters for further data linkage and analyses will be discussed.

Atelier 6: Kooperation

KORSTAT 4.0

Matthias Mazenauer
Sekretär KORSTAT
Schöntalstrasse 5
Zürich, Schweiz

Die neue Austauschplattform der regionalen statistischen Ämter

Die KORSTAT hat dieses Jahr ein neues Zeitalter der Zusammenarbeit in den Gremien eingeläutet. Die Präsentation gibt einen kurzen Überblick über die bereits bestehenden Inhalte und die Möglichkeiten auf der neuen Austauschplattform.

Unter anderem sollen folgende Fragen beantwortet werden:

- Wie wird die Arbeit in den Gremien der KORSTAT erleichtert?
- Wie kann die Plattform die Zusammenarbeit unter den regionalen Statistikstellen unterstützen?
- Was kann jeder von uns beitragen, um die öffentliche Statistik in der Schweiz zu stärken?

Die Präsentation soll zudem Gelegenheit bieten, Rückmeldung zur Plattform zu geben und Bedürfnisse zu sammeln und zu diskutieren.

Atelier 6: Kooperation

Öffentliche Statistik im Alltag – Weiterbildung für Verwaltungsstellen

Lukas Mohler
Statistisches Amt des Kantons Basel-Stadt
Binningerstrasse 6
4001 Basel

Samuel Wegmann
LUSTAT Statistik Luzern
Burgerstrasse 22
6002 Luzern

Verwaltungsstellen brauchen statistische Daten und Analysen als Informations- und Planungsgrundlagen. Die Statistikstellen der Kantone Basel-Stadt und Luzern bieten beide im Rahmen von verwaltungsinternen Weiterbildungsprogrammen Einführungskurse in die öffentliche Statistik an.

In den Kursen werden beispielsweise die Aufgaben sowie die rechtlichen und ethischen Grundlagen der öffentlichen Statistik erläutert und deren Bedeutung hervorgehoben. Es werden weiter Tipps zur Datensuche und zur Nutzung und Interpretation von Statistiken gegeben sowie statistische Grundbegriffe erklärt.

Ein Co-Referat der beiden Statistikstellen geht folgenden Fragen nach: Welche Überlegungen haben zu diesen Angeboten geführt? Welche Erfahrungen wurden mit ihnen gemacht? Wie sollen sie in Zukunft weiterentwickelt werden?

Atelier 6: Kooperation

La cartographie au service de la statistique publique avec la plateforme Statatlas

Olaf König
Office federal de la statistique
Espace de l'Europe 10
Neuchâtel, Suisse

Retour d'expérience et perspectives futures

L'Office fédéral de la statistique (OFS) exploite depuis 2009 un outil d'édition de visualisations graphiques – la plateforme Statatlas – qui couvre plus de 90% des besoins de l'OFS en matière de production et de diffusion cartographique.

Cet outil est une application en ligne, un *Content Management System* (CMS), qui permet d'importer et de gérer des données spatiales, d'éditer des cartes thématiques, des systèmes d'indicateurs régionaux, et d'autres produits sur la base de données régionales, et d'exporter et de diffuser ces produits dans différents formats et pour différents médiums. La majorité de cette diffusion passe par l'offre en ligne d'atlas statistique et de cartes interactives.

L'OFS met aujourd'hui à la disposition du public trois atlas statistiques, qui concourent à notre mandat d'information et disposent de bases légales assurant la pérennité de cette offre.

L'Atlas statistique interactif de la Suisse est l'atlas de référence de l'OFS et couvre la quasi-totalité des thèmes statistiques et des nomenclatures spatiales, avec une mise à jour en continue au fur et à mesure de la disponibilité de nouveaux résultats.

L'Atlas politique de la Suisse donne accès à l'ensemble des résultats politiques à l'échelon fédéral – votations depuis 1866 et élections au Conseil national depuis 1919 – et offre des cartes mises à jour en direct lors des dimanches de votations et d'élections.

Enfin, l'Atlas des villes regroupe les résultats du projet City Statistics, qui regroupe des indicateurs de condition de vie en ville, et visualise des résultats à différents niveaux géographiques, dont des niveaux infra-communaux.

Cet outil de production et de diffusion très complet est également mis à disposition d'acteurs tiers, et la statistique du canton de Lucerne – LUSTAT – utilise depuis 2014 notre CMS pour produire et diffuser l'Atlas statistique de Lucerne ainsi que des cartes interactives.

Après plus d'une décennie de service rendu, et des dizaines de milliers de cartes produites et diffusées, l'OFS devra se doter dans les prochaines années d'une nouvelle application qui succédera à la plateforme actuelle. Il s'agit de proposer un outil comparable tout en améliorant la productivité et en élargissant l'offre de produits et de leurs fonctionnalités, avec l'adoption des standards actuels et futurs du *webmapping* et plus largement de la *data visualisation*.

Dans cette optique, l'OFS souhaite développer d'autres partenariats au sein des acteurs suisses de la statistique publique, afin de partager et de développer un outil innovant à même d'apporter au public par le moyen de la représentation graphique une information éclairée et facilement appréhendable au service de notre démocratie.

Atelier 6: Kooperation

Kleinräumige Arbeitsplatzszenarien für Kanton und Stadt Zürich

Andrea Schnell
Statistisches Amt Kanton Zürich
Schöntalstrasse 5
8090 Zürich

Rolf Schenker
Statistik Stadt Zürich
Napfgasse 6
8001 Zürich

Eine Zusammenarbeit der statistischen Ämter von Kanton und Stadt Zürich

Szenarien zur zukünftigen Entwicklung der Bevölkerung haben sich als Planungs- und Entscheidungsgrundlage seit Jahren etabliert – auch in Zürich. Daneben ist seit einigen Jahren ein Bedürfnis nach Entwicklungsszenarien für die Zahl der Arbeitsplätze feststellbar. Da der Wirtschaftsstandort Zürich nicht an der Stadtgrenze endet, haben sich Statistik Stadt Zürich und das Statistische Amt des Kantons Zürich gemeinsam Gedanken zur zukünftigen Entwicklung der Arbeitsplätze in Stadt und Kanton Zürich gemacht. Als Ausgangspunkt dienen die Szenarien zur künftigen Bevölkerungsentwicklung sowie die zur Verfügung stehenden Flächenreserven.

Die Präsentation soll einerseits die Modalitäten der Zusammenarbeit darlegen, andererseits aber auch einen ersten Blick auf die Resultate geben und mögliche Weiterentwicklungen des Modells aufzeigen.

Atelier 7: Open!

Open Government Data als wichtiger Beitrag zu einem starken Daten-Ökosystem

Juan Pablo Lovato
Bundesamt für Statistik BFS
Espace de l'Europe 10
2010 Neuchâtel, Switzerland

Am 30. November 2018 wurde die zweite Open Government Data Strategie (2019 – 2023) vom Bundesrat verabschiedet. Damit will der Bundesrat die OGD-Politik stärken und den Strukturwandel, den die Digitalisierung auslöst, mitgestalten und vorantreiben. Die neue Strategie ist deshalb auf die Strategien «Digitale Schweiz» und «E-Government» abgestimmt. Ein freier Zugang zu Verwaltungsdaten fördert Transparenz und Partizipation, stärkt die Verantwortlichkeit und trägt zur Steigerung der Wertschöpfung bei, etwa indem Innovationen gefördert oder neue Geschäftsmodelle ermöglicht werden. Daher sollen ab 2020 alle publizierten Daten von Bundesstellen als offene, frei und maschinell nutzbare Verwaltungsdaten publiziert werden. Ausgenommen bleiben Daten, wenn überwiegende legitime Schutzinteressen dies erfordern, etwa der Datenschutz oder der Schutz von privaten Geschäftsgeheimnissen.

Die publizierten Daten müssen möglichst effizient ins Daten-Ökosystem integriert werden. Dafür sind unterschiedliche Anforderungen zu erfüllen. Die Open Government Data müssen sowohl für Personen als auch für technische Systeme leicht auffindbar und nutzbar sein und unter einer offenen Lizenz verfügbar sein. Es geht aber weiter: Verschiedene organisatorische und kulturelle Massnahmen sind umzusetzen, wie die Unterstützung der Stärkung der Datenkompetenz oder die Aufbereitung direkterer Kanäle zwischen Datenanbietern und Datennutzern, um eine systemische Kooperation zu ermöglichen und fördern. Nur so kann der kulturelle Wandel, der sich in dem OGD-Paradigma manifestiert, unterstützt und das Potential der Open Government Data realisiert werden.

Zum Aufbau einer effizienten Dateninfrastruktur gehört zudem eine systematische Zusammenarbeit zwischen den involvierten publizierenden Stellen: Bund, Kantone, Gemeinden oder Organisationen, welche Daten von öffentlichem Interesse publizieren.

Welche Massnahmen sind aber wirklich effektiv? Und welche sind konkret realisierbar? Und wer sollte sie umsetzen?

1) Strategie für offene Verwaltungsdaten in der Schweiz 2019–2023 (Open-Government-Data-Strategie, OGD-Strategie) <https://www.admin.ch/opc/de/federal-gazette/2019/879.pdf>

Atelier 7: Open!

Die neue OGD-Plattform des Kantons Basel-Stadt

Jonas Eckenfels und Jonas Bieri
Fachstelle Open Government Data Basel-Stadt
Binningerstrasse 6
4001 Basel, Schweiz

Vorstellung von Inhalten und Funktionalitäten

Seit Januar.2019 besteht die Fachstelle Open Government Data (OGD) am Statistischen Amt des Kantons Basel-Stadt. Der Aufbau der Fachstelle folgte auf ein erfolgreich durchgeführtes zweijähriges Pilotprojekt. Seit Anfang November ist nun die durch die Fachstelle OGD betriebene Datenplattform live.

Die OGD-Plattform Basel-Stadt zeigt sämtliche kantonalen Daten, welche als OGD publiziert sind und damit Daten aus unterschiedlichen thematischen Bereichen und von verschiedenen kantonalen Dienststellen wie dem Amt für Umwelt und Energie, dem Grundbuch- und Vermessungsamt, der Kantonspolizei oder auch dem Statistischen Amt. Die Plattform bietet sowohl Echtzeitdaten wie auch aggregierte, historisierte Daten zur freien Nutzung in verschiedenen Formaten an. So können zum Beispiel Echtzeitdaten der Rhein-Überwachungsstation oder der baselstädtischen Parkhäuser ebenso genutzt werden wie Migrationsdaten zu Zu- und Wegzögern in den Kanton Basel-Stadt nach verschiedenen Merkmalen oder diverse Geo-Basis-Datensätze.

Die Plattform erlaubt den Nutzern, die Daten bereits vor dem Download anzuschauen, zu filtern und zu sortieren. Auch können einfache Visualisierungen (Diagramme, Karten) direkt erstellt werden und auf baselstädtischen OGD basierende Projekte (Visualisierungen, Applikationen) können verlinkt werden. Für Programmierer stehen die Daten auch als API zur Verfügung, was die Entwicklung von Websites und Apps stark vereinfachen kann.

Über eine Kommentarfunktion können Datenkonsumenten mit den Dateneignern in Kontakt treten und Feedback zu den angebotenen Daten geben.

Sämtliche Daten, welche auf der OGD-Plattform Basel-Stadt zur Verfügung gestellt werden, können auch über den OGD-Katalog des Bundes auf opendata.swiss gefunden werden.

Atelier 7: Open!

Managing a modern R environment

Bernhard Meindl
Statistics Austria
Guglgasse 13
Vienna, Austria

Experiences from the last few years at Statistics Austria

In this contribution we want to discuss the experiences we have made over the last few years in setting up and maintaining the R infrastructure for R users within Statistics Austria. In 2018 the transition from a desktop-based to a server-based infrastructure was started. We give some insights on the steps we took and the lessons that we have learned in the process.

To minimize the impact of changes for users, we set up a two-phase infrastructure consisting of a testing- and a production environment. In order to properly scale the servers, we aimed for a smooth transition and not trying to migrate all users at once. Thus, monitoring of the workload on the servers was an issue that we kept in mind. Furthermore, we also created a new training course that was aimed specifically at our existing R users which wanted to move to the server infrastructure. In the process, we have developed several new R packages for internal usage (for example to connect to databases or to mount shares) that help our staff in the transition phase.

We also show how we manage different R Installations, when we update and/or freeze packages and how we use RStudio Connect to provide APIs, documents and apps for internal usage and give some insights on the remaining problems that need to be solved and outline some future work.

Atelier 7: Open!

Offene Daten zur Nutzung offener Daten publizieren

Thomas Lo Russo
Statistisches Amt des Kantons Zürich
Schöntalstrasse 5
8090 Zürich, Schweiz

Die Fach- und Koordinationsstelle OGD (Statistisches Amt des Kantons Zürich) greift auf Webstatistiken zur Nutzung der offenen Behördendaten (OGD) des Kantons Zürich zurück, um kontinuierlich zu prüfen, wie das OGD-Angebot genutzt wird. Das Ziel ist es, Indizien dafür zu finden, wie das Datenangebot bedarfsgerechter und attraktiver gestaltet werden kann. Seit Anfang 2019 sind die monatlichen Webstatistiken (ab Februar 2018) als OGD auf opendata.swiss verfügbar.

<https://opendata.swiss/de/dataset/web-analytics-der-open-government-data-des-kantons-zuerich>

Ausgewählte Metriken werden im öffentlich zugänglichen [openzh-Dashboard](https://openzh.shinyapps.io/openzh_dashboard/) (beta) dargestellt:

https://openzh.shinyapps.io/openzh_dashboard/

Die Webstatistik-Daten erlauben es im Detail nachzuvollziehen, wie hoch das Interesse an den offenen Behördendaten des Kantons ist, welche Datensätze besonders häufig nachgefragt werden und wie sich die Zugriffszahlen über die Zeit entwickeln.

Quelle der Daten sind zwei unterschiedliche Schnittstellen. Mittels der offenen Software R werden die notwendigen Informationen zusammengezogen und in Form eines einzigen Datensatzes publiziert.

Atelier 8: Steuerdaten und Einkommen

Steuerstatistik – Grundlage für eine datenbasierte Steuerpolitik

Roberto Frisullo
LUSTAT Statistik Luzern
Burgerstrasse 22
Postfach 3768
6002 Luzern

Mit dem Bundesgerichtsentscheid vom 10. April 2019 wurde die eidgenössische Volksabstimmung vom 28. Februar 2016 zur Volksinitiative «Für Ehe und Familie - gegen die Heiratsstrafe» aufgehoben. Unter anderem wurde die Nennung von 80'000 betroffenen Zweiverdiener-Ehepaaren in der Medienmitteilung des Bundesgerichts vom 10. April 2019 als fehler- und lückenhafte Information bezeichnet.

Eine externe Überprüfung der Schätzmethoden und des statistischen Materials der Eidgenössischen Steuerverwaltung (ESTV) kam zum Schluss, dass wegen fehlender Individualdaten der Steuerpflichtigen die Bereitstellung der Grundlagen für die nationale Steuerpolitik erschwert wird. In der Tat sind bei fehlenden Individualdaten zu viele Annahmen über die Grundgesamtheit zu treffen, was die Genauigkeit der Schätzung stark beeinflusst.

Die Steuerstatistik von LUSTAT Statistik Luzern enthält die wichtigsten Ziffern der Luzerner Steuererklärungen ab der Steuerperiode 2005. Im Auftrag der Dienststelle Steuern werden auf dieser Individualdatenbasis verschiedene Szenarien zu unterschiedlichen Anpassungen des Steuergesetzes berechnet oder modelliert. Dabei können auch die Anzahl betroffener Steuerpflichtige und deren soziodemografische Zusammensetzung ausgewertet werden. Diese empirische Vorgehensweise ermöglicht genauere Schätzungen der Mehr- bzw. Mindereinnahmen aufgrund von Steuergesetzesanpassungen. Des Weiteren können die Modellparameter einfach variiert und so auch die Sensitivität der Anpassung analysiert werden.

In diesem Beitrag werden die Möglichkeiten und Einschränkungen der Steuerstatistik aufgezeigt. Es wird auch kurz auf mögliche Erweiterungen der Steuerstatistik eingegangen und die weiteren Fragestellungen, die damit beantwortet werden können.

Atelier 8: Steuerdaten und Einkommen

Appariement du relevé structurel avec les données fiscales: vers une statistique cantonale du revenu des ménages

Matti Langel
Office cantonal de la statistique (OCSTAT), Genève

La mise à disposition d'indicateurs statistiques fiables sur le revenu des ménages au niveau cantonal répond à un besoin récurrent, qui émane notamment des autorités politiques genevoises, mais aussi de plusieurs autres acteurs. Afin de répondre à cette demande l'OCSTAT a opté pour un appariement des données fiscales avec le relevé structurel, dont l'échantillon est densifié pour le canton de Genève.

Une fois réalisé, un tel appariement permet de disposer de résultats solides relatifs au revenu des ménages, ainsi que d'utiliser la notion de revenu d'équivalence pour produire des estimations pour les indicateurs classiques de pauvreté et d'inégalité (taux de risque de pauvreté, indice de Gini, rapport S80/S20, etc.). De plus, la richesse du relevé structurel en termes d'informations sur les ménages et leurs membres (types de ménage, niveau de formation, statut d'occupation, informations sur le logement, etc.), offre de nombreuses nouvelles possibilités d'analyse.

La présentation a également pour objectif de résumer la démarche méthodologique ainsi que les difficultés liées à un tel appariement entre des données d'enquête et des données de registre.

Atelier 8: Steuerdaten und Einkommen

Le revenu des ménages vaudois: un exemple d'utilisation et d'appariement de données administratives

Benjamin Gay
Statistique Vaud
Rue de la Paix 6
Lausanne, Suisse

Le revenu de ménages vaudois est l'aboutissement de plusieurs projets distincts basés sur l'utilisation de sources administratives pour obtenir une information statistique multithématique, composite et d'un haut niveau de complexité. Le premier est la mise en place d'un système d'appariement (base de données socio-économiques vaudoises, BDSEV) entre données fiscales et prestations sous conditions de ressources qui permet de calculer un revenu « amélioré » tenant compte à la fois des revenus fiscalisés et des revenus exonérés (prestations sociales). Dans un second temps, StatVD a utilisé les données du contrôle des habitants pour créer des ménages. Il s'agissait de rassembler les personnes habitant un même logement et de caractériser les relations qui les unissent afin de définir un type de ménage. L'ajout de ces données ménages à la BDSEV a permis de calculer un niveau de vie à l'échelle du ménage. Cette présentation reviendra sur les principaux résultats de ces travaux.

L'utilisation de sources multiples permet de compenser, au moins partiellement, les manques des données administratives. Par exemple, la prise en compte des revenus non imposables des aides sociales permet une meilleure description des bas revenus et une mesure de la pauvreté. On peut aussi signaler que la prise en compte des données de ménages permet de traiter les cas de personnes partageant un même logement, mais remplissant une déclaration d'impôt indépendante (en particulier les concubins et les enfants majeurs). Les données administratives sont intrinsèquement moins précises que des données statistiques collectées avec un objectif précis (par exemple l'enquête sur le budget des ménages). Néanmoins, la masse d'informations disponibles et les possibilités d'appariement permettent de déduire une information statistique sophistiquée et multithématique.

En plus des principaux résultats et apports de cette démarche, nous reviendrons aussi de manière plus générale sur les défis liés à la mise en place de ce type de statistique. La principale difficulté est la quantité de ressources, essentiellement en termes de temps, nécessaire à leur réalisation. La recherche de synergie entre services statistiques et administratifs ou entre services statistiques est une piste intéressante pour rassembler des ressources et atteindre les objectifs. A l'interne des services statistiques, il est aussi intéressant de favoriser les synergies entre personnes et entre équipes afin de pouvoir dégager les compétences et ressources nécessaires.

Atelier 8: Steuerdaten und Einkommen

Die Statistik zur finanziellen Lage der Luzerner Haushalte – Potential und Herausforderungen der Datenverknüpfung für die öffentliche Statistik

Nathalie Portmann
LUSTAT Statistik Luzern
Burgerstrasse 22
Postfach 3768
6002 Luzern

Zur Analyse der finanziellen Lage der Luzerner Bevölkerung erstellte LUSTAT Statistik Luzern vor einigen Jahren erstmals einen neuartigen verknüpften Datensatz, der seither jährlich aktualisiert wird und aktuell die Jahre 2009 bis 2016 umfasst. Verknüpft werden Informationen aus dem Personen- und Wohnungsregister, Steuerdaten sowie Daten zu bedarfsabhängigen Sozialleistungen (u.a. Stipendien, IPV, EL, Sozialhilfe). Der Datensatz ermöglicht umfassende, kleinräumige und valide Analysen zur Einkommenssituation der Haushalte (u.a. zum Erwerbseinkommen und zu den Renten), zum Vermögen und zum Beitrag staatlicher Transferleistungen (Armutsbekämpfung).

Inzwischen verfolgen mehrere andere Statistikbehörden ähnliche Projekte, und auf nationaler Ebene bestehen Bemühungen zur Erstellung eines kantonsübergreifenden Datensatzes. Auch andernorts werden komplexe Datensätze, die unterschiedlichste Administrativdaten und öffentliche Statistiken miteinander verknüpfen, vermehrt genutzt. Vor diesem Hintergrund schildert der Beitrag aktuelle und zukünftige Herausforderungen, die sich LUSTAT bei der Aufbereitung, Weiterführung und der Auswertung dieses Datensatzes stellen. Ausgeführt werden soll insbesondere:

- Wie werden die einzelnen Datenquellen mit unterschiedlichen Bezugszeiträumen, Untersuchungseinheiten und Grundgesamtheiten harmonisiert? (Stichworte Haushaltsbildung, quellenbesteuerte Personen)
- Wie kann die langfristige Aktualisierung und Kontinuität des verknüpften Datensatzes sichergestellt werden, sowohl in organisatorischer als auch in methodischer Hinsicht?
- Wie soll mit unvollständigen oder fehlenden Informationen umgegangen werden?

Das einzigartige Erkenntnispotential des Datensatzes soll abschliessend anhand einer beispielhaften Analyse illustriert werden, die aufzeigt, wie sich je nach Typ des armutsbetroffenen Haushalts die Relevanz der einzelnen staatlichen Sozialleistungen bei der Armutsbekämpfung unterscheidet.

Atelier 9: Innovationen in der Datenerhebung

Modernisierung des eCensus

Julia Voronkova
Bundesamt für Statistik
Espace de l'Europe 10
2010 Neuchâtel, Schweiz

Eine Antwort auf die Ansprüche des digitalen Zeitalters

Der Online-Fragebogen der Strukturerhebung der Volkszählung (eCensus) wurde für die Erhebung 2018 neu programmiert. Hintergrund war die Absicht, die befragten Personen zur verstärkten Nutzung des eCensus zu motivieren. Es wurde eine benutzerfreundliche und moderne Lösung umgesetzt, welche sich automatisch an die Grösse des Bildschirms des verwendeten Endgeräts (PC, Notebook, Tablet, Smartphone) anpasst (Responsive Web Design). Damit einher ging ein neues Online-First-Vorgehen beim Erstversand des Erhebungsmaterials an die befragten Personen.

Die befragten Personen haben die Möglichkeit, ihre Antworten online über den eCensus oder auf dem Papierfragebogen anzugeben. Bis zur Erhebung 2017 war der Rücklauf über den eCensus-Kanal ca. 30%. In dieser Präsentation wird aufgezeigt, wie das Verfahren für die Erhebung 2018 angepasst werden konnte, um einen Rücklauf über den eCensus-Kanal von über 70% zu erreichen. Auf die begleitenden Massnahmen, welche ein solches Verfahren ermöglichen, wird ebenfalls eingegangen.

Mit diesem neuen Vorgehen konnte die Datenqualität der Erhebung erheblich gesteigert werden. Die wichtigsten Kennzahlen zur Erhebung werden präsentiert.

Atelier 9: Innovationen in der Datenerhebung

Wie Cloud-Lösungen Erhebungsprozesse verbessern

Cuno Lanz & Raphael Vogel
LUSTAT Statistik Luzern
Burgerstrasse 22
Luzern, Schweiz

Vereinfachte Prozesse und höhere Datenqualität am Beispiel Gemeindefinanzstatistik

Den bei den Luzerner Gemeinden aktuell stattfindenden Wechsel der Rechnungslegungs-Vorschriften von HRM1 zu HRM2 hat LUSTAT Statistik Luzern zum Anlass genommen, die Erhebung der Gemeindefinanzstatistik vollständig neu zu konzipieren. Es war das Ziel, einerseits den Erhebungsprozess zu vereinfachen und zu digitalisieren und andererseits die Kontinuität der Statistik sowie die direkte Verwendung für Drittaufgaben (wie kommunaler Finanzausgleich, Quellensteuer, Strassenrechnung, Bildungsfinanzstatistik, Kulturfinanzstatistik etc.) zu gewährleisten.

Entwickelt wurde eine massgeschneiderte Cloud-Lösung als sicherer, vertrauenswürdiger Intermediär zwischen den Beteiligten. Sie ist das zentrale Tool eines simplen, dreiphasigen Erhebungsprozesses und erlaubt den Gemeinden, ihre Dateneinreichung autonom sowie in einem Durchgang zu erledigen. Das Erhebungstool prüft dabei die über eine standardisierte Schnittstelle (ED-OEFIN gemäss BFS/EFV) importierten Daten und meldet Mängel (fehlerhafte oder unplausible Daten) sofort. Diese werden in der Folge mit Unterstützung des Tools direkt durch die Gemeinden behoben. Korrigiert oder begründet die Gemeinde Mängel, so aktualisiert das Erhebungstool die damit neu erzielte Datenqualität und visualisiert sie mittels eines einprägsamen Ampelsystems. Das Erhebungstool ist dabei auf eine einfache Bedienbarkeit hin ausgelegt, um den Gemeinden eine möglichst einfache Übermittlung und ggf. Korrektur der Daten zu ermöglichen.

Der Vortrag fokussiert auf drei für die öffentliche Statistik zentrale Vorteile der Verwendung von Cloud-Lösungen: Sie ermöglichen (a) eine umfassende Prozessübersicht, d.h. einen Überblick über den Status und die Qualität der Erhebung, sowohl für die Erhebungsstellen als auch für den Empfänger der Daten. Sie ermöglichen (b) einen verstärkten Fokus auf Daten und Datenqualität durch Übernahme eines erheblichen Anteils an administrativer Arbeit und Qualitätstests. Und schlussendlich generieren sie (c) eine erhöhte Validität der Daten, da fehlerinduzierte Änderungen durch diejenige Instanz vorgenommen werden, welche am nächsten bei den Daten sind. Diese drei Plädoyers für Cloud-Lösungen werden um konkrete Erfahrungen aus dem Entwicklungsprozess ergänzt.

Atelier 9: Innovationen in der Datenerhebung

Twitter as a potential complementary source to traditional survey data: the case of Switzerland

Maud Reveilhac & Stephanie Steinmetz
Lausanne University (LINES Institute)
Géopolis building
Lausanne, Switzerland

An increasing number of people is active on various social media platforms, where they voluntarily share information and discuss topics of interest. In the meantime, survey response continues to decrease (Ernst-Stähli 2011; Kreuter 2013; Williams & Brick 2017) and administrative data complements survey data in order to produce statistics of sufficient quality at an efficient cost (Diekmann 2018). Against this background, we investigate the potential usability of social media, most notably Twitter, as a complementary data source for official statistics in Switzerland. We focus on three main components of social media data: data collection, population bias and topic classification.

Following Daas et al. (2012) we use the “snowball” sampling collection technique. The collection of Twitter users starts with collecting information (such as location and profile description) about followers of a Swiss politician who has the highest number of followers. The process continues with all not readily included followers of these followers and is repeated until no further new user is identified. We select only Swiss users by relying on the location and description fields. It results in a dataset of unique Swiss Twitter users for which we can estimate gender, age group and political affiliation by relying on natural language models. This enables us to examine overall population bias of Twitter data as compared to standard survey samples.

For each Twitter user, we then collect the most recent 200 tweets in order to identify the topics discussed on Twitter. The most prevalent topics indicate us for which research areas Twitter data can most valuably complement survey data. The identification of the topics is done in two stages. First, the topics in all the hashtag containing messages are determined and messages are classified accordingly by clustering techniques. Next, a random sample of the non-hashtag containing messages is classified based on a classification model developed on the hashtags containing messages. We present the prevalence of topics found in Twitter conversation and detail the distribution of topics according to the estimated gender, age groups and political affiliation in order to compare it with survey data at the aggregated level of analysis: the prevalence of topics is compared with the proportion of valid item answers (non-neutral and non-“don’t know” survey answers).

The paper concludes with a discussion of challenges and potentials for using Twitter in the context of official statistics. The findings will be not only relevant for official statisticians but also for sociologist and other scholars interested in the study of social media.

- 1) Ernst-Stähli, M. 2011. «Switzerland». In: Häder S., Häder M., Kühne M. (eds.). *Telephone Surveys in Europe*. Berlin: Springer.
- 2) Kreuter, F. 2013. «Facing the Nonresponse Challenge». *The Annals of the American Academy of Political and Social Science*, 645(1), 23–35.
- 3) Williams, D. & Brick, J. M. 2017. «Trends in U.S. Face-to-Face Household Survey Nonresponse and Level of Effort». *Journal of Survey Statistics and Methodology*, (September), 1–26.
- 4) Diekmann, A. 2018. «New Perspectives for Survey Research. Data Linkage, Georeferenced Data, Causal Analysis». [PowerPoint slides]. Presentation given at the University of Lausanne, Switzerland, on September 12th 2018 at FORS 10th anniversary symposium «The future of data collection in the social science». Retrieved May 1, 2019, from <https://www.ethz.ch/content/dam/ethz/special-interest/gess/chair-of-sociology-dam/documents/news/FORS%2010%20New%20Perspectives%20for%20Survey%20Research.pdf>
- 5) Daas, P., Roos, M., van de Ven, M., & Neroni, J. 2012. «Twitter as a potential data source for statistics». (Discussion paper 201221). Den Haag/Heerlen, The Netherlands: Statistics Netherlands.

Atelier 9: Innovationen in der Datenerhebung

PendelIndendaten unter der Lupe

Walter Eichhorn
Statistik Stadt Bern
Junkerngasse 47
3000 Bern 8, Schweiz

Wohn- und Arbeitsort: Nichts leichter als das! – Oder?

Wer sich für Arbeitspendelnde interessiert, kann sich nur schon im Bereich der öffentlichen Statistik auf verschiedenste Datenquellen stützen. Es liegt in der Natur der Dinge (der Statistik?), dass unterschiedliche Untersuchungen mit unterschiedlichen Ansätzen, unterschiedlichen Methoden und unterschiedlichen Zeitreferenzen zu unterschiedlichen Ergebnissen führen.

Doch wenn die feststellbaren Differenzen zwischen den einzelnen Quellen ein übergrosses Mass annehmen, gilt es sorgfältig hinzusehen und sich die Frage zu stellen, was die einzelnen Statistiken eigentlich ausweisen. Genau dieser Frage werden wir uns widmen.

Mit der hier präsentierten Arbeit sollen verschiedene verfügbare Datenquellen für die Stadt Bern betrachtet und die möglichen Interpretationsspielräume ausgeleuchtet werden. Ziel ist es, (etwas) Klarheit zu schaffen und aufzuzeigen, was hinter den verschiedenen PendelIndendaten steckt, welche Aussagen möglich sind und welche Grenzen und Einschränkungen zu beachten sind.

Wir werden uns weder dem Geschlecht, dem Alter oder gar dem sozialen Hintergrund der Personen annehmen, noch ihrer Ausbildung, ihrem Beschäftigungsgrad oder der Branche in welcher sie tätig sind. Auch interessiert uns hier nicht die Weglänge oder -dauer und auch nicht, welche Verkehrsmittel zum Einsatz kommen. Es geht somit 'einzig' um Wohn- und Arbeitsort der Erwerbstätigen: Bereits so ein Unterfangen mit ausreichend Tücken.