The Internet as a Data Source for Advancement in Social Sciences

Askitas, Nikos; Zimmermann, Klaus F.

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

Askita, Nikolaos; Zimmermann, Klaus F.

Working Paper

The Internet as a Data Source for Advancement in Social Sciences

RatSWD Working Paper, No. 248

Provided in Cooperation with:
German Data Forum (RatSWD)

Suggested Citation: Askita, Nikolaos; Zimmermann, Klaus F. (2015) : The Internet as a Data Source for Advancement in Social Sciences, RatSWD Working Paper, No. 248, Rat für Sozial- und Wirtschaftsdaten (RatSWD), Berlin

This Version is available at:
http://hdl.handle.net/10419/108978

RatSWD Working Paper Series

www.ratswd.de

RatSWD.
German Data Forum

248

The Internet as a Data Source for Advancement in Social Sciences

Nikolaos Askitas
Klaus F. Zimmermann

April 2015

SPONSORED BY THE
Federal Ministry
of Education
and Research

Working Paper Series of the German Data Forum (RatSWD)

The *RatSWD Working Papers* series was launched at the end of 2007. Since 2009, the series has been publishing exclusively conceptual and historical works dealing with the organization of the German statistical infrastructure and research infrastructure in the social, behavioral, and economic sciences. Papers that have appeared in the series deal primarily with the organization of Germany's official statistical system, government agency research, and academic research infrastructure, as well as directly with the work of the RatSWD. Papers addressing the aforementioned topics in other countries as well as supranational aspects are particularly welcome.

RatSWD Working Papers are non-exclusive, which means that there is nothing to prevent you from publishing your work in another venue as well: all papers can and should also appear in professionally, institutionally, and locally specialized journals. The *RatSWD Working Papers* are not available in bookstores but can be ordered online through the RatSWD.

In order to make the series more accessible to readers not fluent in German, the English section of the *RatSWD Working Papers* website presents only those papers published in English, while the German section lists the complete contents of all issues in the series in chronological order.

The views expressed in the *RatSWD Working Papers* are exclusively the opinions of their authors and not those of the RatSWD or of the Federal Ministry of Education and Research.

The Internet as a Data Source for Advancement in Social Sciences

Nikolaos Askitas
IZA

Klaus F. Zimmermann
IZA and Bonn University

This Paper has first been published in

International Journal of Manpower
Vol. 36 No. 1 - Special Issue

and

IZA Discussion Paper No. 8899

IZA
P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 8899
February 2015

ABSTRACT

The Internet as a Data Source for Advancement in Social Sciences*

This paper advocates the use of Internet data for social sciences with a special focus on human resources issues. It discusses the potentials and challenges of Internet data for social sciences and presents a selection of the relevant literature to establish the wide spectrum of topics, which can be reached. Such data represent a large and increasing part of everyday life, which cannot be measured otherwise. They are timely, perhaps even daily following the factual process, they typically involve large numbers of observations, and they allow for flexible conceptual forms and experimental settings. Internet data can successfully be applied to a very wide range of human resource issues including forecasting (e.g. of unemployment, consumption goods, tourism, festival winners and the like), nowcasting (obtaining relevant information much earlier than through traditional data collection techniques), detecting health issues and well-being (e.g. flu, malaise and ill-being during economic crises), documenting the matching process in various parts of individual life (e.g. jobs, partnership, shopping), and measuring complex processes where traditional data have known deficits (e.g. international migration, collective bargaining agreements in developing countries). Major problems in data analysis are still unsolved and more research on data reliability is needed. Current research is highly original but also exploratory and premature. Our article reviews the current attempts in the literature to incorporate Internet data into the mainstream of scholarly empirical research and guides the reader through this Special Issue. We provide some insights and a brief overview of the current state of research.

JEL Classification: J00, C80, C81, C83

Keywords: World Wide Web, web data, internet data, forecasting, human resources and the internet

Corresponding author:

Klaus F. Zimmermann
IZA
P.O. Box 7240
53072 Bonn
Germany
E-mail: Zimmermann@iza.org

* This paper is the Editorial to a forthcoming special issue of the International Journal of Manpower on "Using internet activity data to analyze human resources issues".
(http://www.emeraldinsight.com/journal/ijm)

1. Introduction

The digital revolution, a loose synonym of digital age or information age marks the use of digital information as opposed to analog or mechanical. Starting in the 1990s, it is impacting our world in a completely new and different way and will continue to be doing so in the foreseeable future. It places networked computing into an increasing number of objects that are neatly integrated into our daily lives thus creating a data driven society and a data driven economy.

This offers the perspective of a complete recording of all aspects of our life. Since demand and supply, all activities of companies and individuals, as well as the matching process are documented in the Internet, not only the complete market economy but also all social aspects are reflected in a huge data cloud (big data). When it is accessible by social scientists, this provides a universe of research potentials. The past cannot only be not forgotten, it can be restudied again and again giving historical analysis a completely new perspective. This suggests a new survey design, since the Internet provides answers to questions before they are asked.

The real life of data analysts and social scientists is not yet so far, although we are getting there much faster than we think. We already see the rapid evolution of online markets for all kinds of goods and services, in particular jobs. And the social media generate huge amounts of information about individual preferences and behavior (Askitas, 2014). The social component will accelerate and become more important as technology becomes miniaturized and produces affordable life accessories that are tightly integrated into daily life.

The impact of the advances in digital technology and the economics of ICT (Information and Communication Technology) components is best understood in the increase of the size of the so-called 'second economy' (Arthur, 2011) in the macro sense and in the proliferation of social media in the micro sense. The size of the second economy in the US, which is defined to be the invisible maze of routers, switches, servers, cabling and so on, is expected to surpass that of the physical economy by 2025[1]. The second economy introduces a neural system underlying the physical world and is the core of the digital age.

The Internet is the most natural place where the second economy surfaces into reality and it is where social media take place, nowadays supported by products like Facebook, Twitter, Google+, LinkedIn or YouTube. The data that will result from the second economy, the Internet, social media and technology miniaturization will be indispensable for social sciences and will more than complement more traditional data sources such as official statistics.

[1] A vivid impression of how extensive the second economy worldwide is may be obtained by taking a look at the maze of submarine cables wiring the planet at http://www.submarinecablemap.com/.

In the paper, we first discuss the type of data that are available on the internet and researchers have started using in their analyses. We reveal the particular chances and challenges these data bring to deal with human resources questions. We then introduce key literature in the subfields related to manpower issues in the social sciences. Finally, we present an overview of the contributions of this special issue.

2. Internet data: Chances and challenges

In the 1980s, as the Internet was in its infancy, social scientists first saw it as a medium over which one could "build and field"[2] surveys with ease, in an unprecedented scale, price and speed. In the 1990s, the Internet started entering the homes and everyday lives of individuals, via email communication, 'surfing' and 'askjeeves' for specific questions, to name a few options. In the 2000s, as web technologies became more involved, via increasingly more effective techniques, and as individuals used the Internet more intensively, tons of data just started piling up. These data, at least in the beginning, existed without even people knowing it. The beauty of these data is that, unlike traditional survey data that are collected upon the consent of the individual and may suffer from several biases, they reveal the visceral and logical choices people make in the privacy of their home, and while they think they are under no observation. With the entry of Google in the industry all kinds of personal information "broke lose"; now these data could be, for example, information about the political preferences of people during an election year under recession, or about the xenophobia of natives after 9/11.

The distributed nature of the Internet allowed surveyors to close the geographic gap while the practically zero marginal costs of email- or web-based surveys made repeating surveys possible not only in an affordable manner but also in an unprecedented cross-sectional size, frequency and scale. Naturally, the Internet as a survey platform brought with it both new potential and new methodological challenges. However, as the Internet becomes ubiquitous there are a priori and in principle no artificial limits in creating truly random and representative data samples. At the same time, as ICT advance sampling becomes less and less of an unavoidable fact. By connecting an ever-larger part of the population we progressively eliminate selection bias because the online population tends to become equal to the general population thus allowing us to have truly random and representative samples, at least when there is full access to the data. At the same time progress in ICT makes sampling unnecessary since we are able to deal with practically unlimited amounts of data.

A prime example of the Internet as a means for carrying out large-scale surveys is the Wage Indicator Survey of the Wage Indicator Foundation,[3] which provides wages on over 60 countries in over 20 languages, producing harmonized and hence comparable data on wages in a large cross section. It also suffers from the aforementioned issues of selection bias, which need more research in order to get a handle on them. Other data

[2] The story of this early development is nicely told in (Das et al., 2011).

[3] See for more information about the Wage Indicator Foundation: http://www.wageindicator.org/ . The data bank center of IZA, IDSC, hosts the Wage Indicator Survey and makes it publicly accessible.

sources used from the Internet include data from Google, Wikipedia, Facebook, Twitter, Google+ and LinkedIn, to mention a few.

As the contacting of surveys over the Internet matured and turned into an indispensable instrument the proliferation of the Internet brought about even more data which do not come from voluntary surveys. This type of data arises because more of what used to be "offline life" is being transferred online. One of the things that ICT and the Internet can do well is to reduce frictions in any kind of matching task in almost any type of market hence many businesses, which deal with searching and matching in various contexts and markets now take place online. Matching is fundamental to life in general and to economics in particular and a large portion of economics research is dedicated to understanding matching problems and finding optimal solutions.

Whether it is about matching city passengers to taxis, long distance travellers to airplane seats,[4] individuals in the marriage market (Hitsch et al., 2010) or in the job market (Kuhn, 2014) the Internet has reduced search frictions and hence opened new business opportunities, where business like Online Dating Services or Job Board Services flourish. Naturally, this creates a new data source for understanding matching in the various contexts of economic behavior and to use it to investigate old questions.

In reality, the Internet has replaced many labor markets. If people are looking for a carpenter, a handyman, a medical doctor or a lawyer they go online and they type these keywords. "Free" services such as "zocdoc," "askJenny" or the "craigslist" immediately provide people with hundreds of options. Likewise, employers often use the Internet to screen and hire employees (maybe via the career network LinkedIn), while cutting down on hiring costs by using Skype. The power of the Internet revolution was particularly evident during the global financial and economic crisis of 2008-2011. Internet data show us that the unemployed searched the web extensively to find a job locally or globally.

The core search-and-match market in the Internet of course is the market of the Internet search engines themselves, which match the supply and demand for documents. They match the demand for information with the supply of documents that contain this information. Google Trends data are an interesting, if aggregate, form of data exactly because of that. Knowing how demand for certain types of information varies in time reveals information about the state of the agents seeking this information. This fact is at the core of Google's business model and Google Trends provide us with an aggregate impression of this demand. This is at the core of our own work with Google Trends data (Askitas and Zimmermann 2009, 2011 & 2015) and technology data in general (Askitas and Zimmermann, 2013).

Internet data are born digital so it is easy to store, manage and curate. They are time-stamped and geo-tagged allowing in principle precise measurements in the longitudinal and cross-sectional dimensions at any preferred scale. While geo-location is today still largely based on the geo-location of IP addresses (precise only at the country level, but increasingly imprecise as we drill down to smaller geographic units) it will eventually

[4] For example: a search for "taxi" in the Travel Category of the Apple App Store returns over 1200 apps, while a search for "cheap flights" returns 290 apps.

become precise as we substitute IP based geo-location with true GPS (Global Positioning System) based geo location.

New data will come to exist out of the so-called Internet of Things. An increasing array of affordable embedded sensors will transmit precise geo-located measurements on time and will cover everything from individual vital signs to the emotional state and well-being of individuals as well as to any economic or other human activity. Obviously these developments will make the economy more data dependent and will hence open new research opportunities but also bring along new challenges. On the one hand new technologies and their recombination produce new data with new potential. On the other hand we continue to inherit some of the old issues as we go along.

Selection bias continues to be an issue due to the fact that adoption differs and will continue to do so with every technological wave both across countries and across individuals. Internet penetration in some countries is now over 90% so that for those countries coverage is eliminated as a potential cause of selection bias but others are still lagging behind. In 2013, the penetration rate has been 95.0% in Norway, 86.2% in Germany, 84.2% in the US, and 74.8% in Spain.[5] However, even in countries in which everybody is online, not everybody uses a smart phone and not everybody uses social media so that any research with new data whose origin includes smart phone technology or social media may still suffer from selection bias. To be clear: Selection bias is not automatically ruled out by coverage, for instance when there is only access to the data of users who made a particular choice. Furthermore, low coverage may not rule out representativeness, this is the basis of traditional surveys, and one has to examine how representative Internet data are.

A natural concern by economists is how to adequately record and measure all these transactions that take place via the Internet. Some raised the issue that while the Internet dramatically increased the welfare of consumers who receive information and entertainment for free since 2000, this is not well reflected in the GDP numbers. Facebook for example, produces advertising supported media as does Google, YouTube, etc. While this is not a problem yet,[6] it can be a challenge in the future as new technology, gadgets and services emerge through the Internet.

Two of the biggest challenges we face, as technology makes all this possible are data privacy issues and data custody and ownership issues. As a society we will need to remain vigilant in both protecting privacy as well as in reforming the normative, ethical and legal framework within which to discuss it. We need to strengthen the institutional structures to keep markets contestable to avoid a misuse of the monopoly of data by a handful of firms, which is vital for society. It is also problematic that the data are unavailable to the common good because they are locked up in proprietary data silos.

[5] Source: Internet World Stats, http://www.internetworldstats.com/top25.htm.
[6] In the US, the current System of National Accounts 2008 that measures GDP includes the value added of producing advertising supported media. However, the output is treated as an intermediate input to other industries rather than as a final household consumption (Soloveichik, 2015).

Finally, a question is related to the government's use of data about its citizens and the limits of this use.[7]

3. A first guide to relevant literature

Internet data can successfully be applied to a wide range of human resource issues including forecasting (e.g. of unemployment, consumption goods, tourism, festival winners and the like), nowcasting (obtaining relevant information much earlier than through traditional data-collection techniques), detecting health issues and well-being (e.g. flu, malaise and ill-being during economic crises), documenting the matching process in various parts of the individual life (e.g., jobs, partnership, shopping, preferences), and measuring complex processes where traditional data have known deficits (e.g. international migration, collective bargaining agreements in developing countries).

In the sequel, we provide information on some literature that has used Internet data in the context of human resources within the social sciences. The early contributions have applied Google activity data; among them, we find Constant and Zimmermann (2008), Ginsberg et al. (2009), Askitas and Zimmermann (2009) and Choi and Varian (2009). An exception is the paper by Ettredge et al. (2005) that used data from the WordTracker's Top 500 Keyword Report.

3.1 Analyzing and predicting unemployment

Predicting the present has been a particular challenge during the Great Recession, where short-term information on unemployment was badly needed but unavailable. The seminal paper by Askitas and Zimmermann (2009) was the first to demonstrate strong correlations between particular Google keyword searches and monthly German unemployment rates. The authors then used the observed structure to predict unemployment behavior under the complex and changing circumstances of the up-rise of the crisis.

This type of exercise has been replicated and extended for various other countries. It turned out that research demonstrates that there is additional information content of Google or similar Internet activity data over alternative time-series models or other business cycle indicators. Such studies have been conducted for the US (Choi and Varian, 2009 and 2012, claims for unemployment benefits), France (Fondeur and Karamé, 2013, unemployment rates), Italy (D'Amuri, 2009, unemployment rate), Spain (Vicente et al., 2015, unemployment levels), the UK (McLaren and Shanbhogue, 2011, unemployment rates), Ukraine (Oleksandr, 2010, unemployment levels, but no acceptable performance), Israel (Suhoy, 2009, unemployment rates), Norway (Anvik and Gjelstad,

[7] Going after tax evaders, the Greek tax authorities in 2010 used Google maps to locate houses with swimming pools that should have been taxed. They managed to increase the number of taxable swimming pools in the Athens suburbs from the originally believed 324 to 16,974 swimming pools. This is a prime example of both data privacy invasion of the individuals and of benefits for the collective. (http://www.spiegel.de/international/europe/finding-swimming-pools-with-google-earth-greek-government-hauls-in-billions-in-back-taxes-a-709703.html)

2010, unemployment rates) and China (Su, 2014, unemployment internet search indicators from Baidu and Google, significant correlation with Purchasing Managers' Employment Indices), among others.

Before Google activity data became available, Ettredge et al. (2005) were able to utilize Internet search engine keyword usage data recorded in the WordTracker's Top 500 Keyword Report published weekly by Rivergold Associates Ltd covering the Web's largest meta-search engines. Providing an unbiased view of searches, they exploited six terms they thought would be mostly used by job seekers to predict US unemployment rates; namely job search, jobs, monster.com, resume, employment, and job listings. The authors concluded that their findings would provide an encouraging correlation. Note that when we did our analysis (Askitas and Zimmermann, 2009), we were not aware of the Ettredge et al. (2005) study.

3.2 Other forecasting, nowcasting and proxying

A variety of research has studied the predictive properties of Internet data, in particular Google activity data, for current and future macroeconomic variables beyond unemployment and the labor market. Investigating business cycles is one well-researched area: Askitas and Zimmermann (2013) demonstrate that the German business cycle can be nowcasted by highway Toll data. Chen et al. (2015) find that Google search volume data help improve the timeliness of business cycle turning point identification during the 2007-2008 US recession.

Another example is aggregate consumer behavior: Choi and Varian (2012) use Google activity data for the US for automobile sales, travel destination planning and consumer confidence. Carrière-Swallow and Labbé (2013) show that Google search queries of automobile purchases in Chile improve the fit and efficiency of nowcasting automobile sales and are better at identifying turning points, although Internet use has been still low in Chile. In a study of the US housing market for 2006-2011, Askitas and Zimmermann (2011) evaluate search intensity data for "hardship letter" from Google Insights to detect ensuing mortgage delinquencies. Other studies are on food stamps data in the US (Fantazzini, 2014), private consumption (Kholodilin et al., 2010 for Germany; Vosen and Schmidt, 2011, for the US) and hotel demand from web traffic data (Yang et al., 2014). Vosen and Schmidt (2011) show that in almost all of their forecasting experiments a Google search activity indicator outperforms well-known survey-based indicators.

Saiz and Simonsohn (2013) suggest to systematically use Internet data to proxy unobservable variables and demonstrate the usefulness of this technique for a selection of occurrence frequencies of crucial social phenomena in the US.

3.3 Health issues

The early innovative study by Ginsberg et al. (2009) used Google activity data to study the influenza epidemic process, applying complex computational methods. While this study has received large attention, it was also found that the structure identified between the searches, the defined keywords and the individual behavior was under change (Lazer et al., 2014). This guides us to the understanding that models have to be adjusted over time.

Another issue of recent concern is the relationship between seasonality, business cycles and mental health. Yang et al. (2010) using US Google data present for the first time on a global level evidence that the incidence of depression varies seasonally. Tefft (2011) studies the relationship between unemployment indicators and Google searches for depression and anxiety in the US around the time of the Great Recession. His empirical evidence implies that unemployment and continued unemployment insurance claims are positively correlated with searches for depression, while initial unemployment insurance claims are negatively related with searches for depression and anxiety.

3.4 Matching on the labor market

As the Internet becomes more and more the place where demand and supply meet, it may soon overshadow the markets' role as a source of information. This matching role is even more important for the labor markets; experts expect the internet to become the dominant platform of exchange (Kuhn, 2014). Online Job Boards, of rising importance in the real world, are being heavily used by researchers (see Askitas and Zimmermann, 2009) and they improve matching in the Job Market (Kuhn and Mansour, 2014). Internet Job Boards not only reduce search frictions, but they are even useful in documenting and studying artificial frictions such as discrimination (Kuhn and Shen, 2012; Maurer-Fazio, 2012). Kureková et al. (2014) provide an insightful study of the methodological challenges that result from the new online job vacancy data and voluntary web-based surveys provided by those platforms.

3.5 Demographic issues

Our understanding of migratory processes is still limited, one of the reasons being that the available data are insufficient since international migrants are not well covered in traditional national surveys. A number of researchers in social sciences have explored the use of Internet data to close gaps. An example is Reips and Buffardi (2012) who are examining social networking Web pages to study migrant biculturalism. Billari et al. (2013) used Google search queries like birth and pregnancy to predict fertility measures with some success.

Hitsch et al. (2010) and Belou (2015) are prominent research examples on the rising role of the Internet on marriage markets. Employing data on user attributes and interactions from an online dating site, Hitsch et al. (2010) estimate mate preferences and are able to predict stable matches. Bellou (2015) shows that in the US, marriage rates in age groups that are more likely to act online have increased due to the diffusion of the Internet, which has even substituted other matching methods (such as via family and friends). In fact, one of the driving factors of reduced divorce rates is better (online) matching.

3.6 Political processes

Internet data can be considered as a universe of answers to questions that were not yet posed. In an early contribution to the literature, Constant and Zimmermann (2008) used Google search engine query data to measure economic and political activities by documenting the evolution of particular keyword searches (financial crisis, credit crunch, recession, unemployment rate) and offered a good impression of which topics

influenced the US presidential elections. Stephens-Davidowitz (2014) employed Google activity data for the US involving racially charged language. Contrary to previous attempts with standard survey data he found solid evidence with these data that racial animus had cost Barack Obama votes as Presidential candidate in 2008. Also Reilly et al. (2012) used Google data for state politics research in the context of the 2008 Presidential election; here by successfully relating the searches with actual engagement in the ballot measures.

Ripberger (2011) compared for the US the public attention of various political issues (health care, global warming, and terrorism) measured by Google activity data with issue coverage in the New York Times and found high correlation and validity.

4. This special issue

In this issue of the *International Journal of Manpower* we put together a set of six carefully selected articles contributing to the rising use of Internet data from different sources and for a variety of purposes.

The paper by Emilio Zagheni and Ingmar Weber on "Demographic Research with Non-Representative Internet Data" (Zagheni and Weber, 2015) addresses the two most critical methodological issues in the use of internet data: non-representativeness and selection bias. It proposes a framework to collect web data and discusses possible estimation methods, while it also makes clear that there are still some large challenges to address. The paper also surveys relevant demographic literature, in particular in the area of migration, where useful data about the mobility process are typically scarce in the traditional data sources.

Two papers study well-being from different data sources. Nikolaos Askitas and Klaus F. Zimmermann are examining "Health and Well-Being in the Great Recession" (Askitas and Zimmermann, 2015) using Google activity data to trace and document the impact of the 2008 Financial and Economic Crisis on well-being. They are able to confirm previous knowledge from the economics of health, well-being and the business cycle. Martin Guzi and Pablo de Pedraza in their article "A Web Survey Analysis of Subjective Well-being" (Guzi and de Pedraza, 2015) employ data from the voluntary web-survey WageIndicator project. They confirm that job characteristics affect job satisfaction and identify spillovers, since satisfaction in one domain affects other domains.

Margaret Maurer-Fazio and Lei Lei study the Chinese Internet job board labor market in their paper “As Rare as a Panda”: How Facial Attractiveness, Gender, and Occupation Affect Interview Callbacks at Chinese Firms" (Maurer-Fazio and Lei, 2015). They examine in a resume audit (correspondence) study, how discrimination derived from gender and facial attractiveness varies across occupation, location, and firms’ ownership type and size. They find that women are generally preferred to men and unattractive job candidates have a disadvantage.

In their paper "Comparing Collective Bargaining Agreements for Developing Countries", Janna Besamusca and Kea Tijdens employ for the first time the web-based WageIndicator Collective Bargaining Agreement Database for 11 developing countries (Besamusca and Tijdens, 2015). They find that few agreements specify wage levels, but

almost all collective agreements have clauses on wages. Their study also documents working hours, paid-leave arrangements and work-family arrangements.

The final paper by Concha Artola, Fernando Pinto and Pablo de Pedraza entitled "Can Internet Searches Forecast Tourism Inflows?" represents the large literature on using Internet data for forecasting purposes (Artola, Pinto and de Pedraza, 2015). Employing Google activity data, the authors demonstrate that traditional time-series forecasting models for tourism inflows into Spain can be improved using Google activity measures.

These contributions are an earnest attempt to evaluate the potentials of the new data sources, and an encouragement for their further use. Research has to go beyond the current borders, both in terms of methodological developments, as in terms of practical use of the data in concrete applications in all areas of the social sciences.

5. Conclusion

Internet data will record a large part of our life and will become, at least to some extent, part of the research in the social sciences. Already today, there is a rising and successful use of proxying, nowcasting or forecasting reality, explaining behavior and modeling market processes. It seems that there is a large potential to push the research frontier ahead. A problem is still the coverage of the data and the limited techniques we have available to deal with the application challenges. Ahead of us is the collection of "Big Data"; therefore, we may not be as concerned with selection bias in the future. However, this expectation is based on the free and complete access to the data. The often proprietary nature of such data and privacy concerns may delay such access. Nevertheless, we think that with the rise of the Internet of Things we will soon see a further revolution in applied research in the social sciences.

References

Anvik, C & Gjelstad, K 2010, 'Just Google it. Forecasting Norwegian unemployment figures with web queries', *Working Paper 11*, Center for Research in Economics and Management, Oslo.

Artola, C, Pinto, F & de Pedraza, P 2015, 'Can Internet Searches Forecast Tourism Inflows?', in this issue.

Arthur, WB 2011, `The second economy`, McKinsey Quarterly: October. http://www.mckinsey.com/insights/strategy/the_second_economy

Askitas, N 2014, **'**Social Media: eine technologische und ökonomische Perspektive', in *Social Media im Unternehmen – Ruhm oder Ruin*, eds C Rogge & R Karabasz, Springer Vieweg, Wiesbaden, pp. 155-166. **DOI**: 10.1007/978-3-658-03087-2_14.

Askitas, N & Zimmermann, KF 2009, 'Google Econometrics and Unemployment Forecasting', *Applied Economics Quarterly*, vol. 55, no. 2, pp. 107-120. **DOI**: 10.3790/aeq.55.2.107.

Askitas, N & Zimmernann, KF 2011, **'**Detecting Mortgage Delinquencies', *IZA DP 5895*, IZA, Bonn.

Askitas, N & Zimmermann, KF 2013, 'Nowcasting Business Cycles Using Toll Data', *Journal of Forecasting*, vol. 32, no. 4, pp. 299-306. **DOI**: 10.1002/for.1262.

Askitas, N & Zimmermann, KF 2015, 'Health and Well-Being in the Great Recession', in this issue.

Bellou, A 2015, 'The Impact of Internet Diffusion on Marriage Rates: Evidence from the Broadband Market', *Journal of Population Economics*, vol. 28, no. 2, pp. 265-297. Available from: <http://link.springer.com/article/10.1007/s00148-014-0527-7>.

Besamusca, J & Tijdens, K 2015, 'Comparing Collective Bargaining Agreements for Developing Countries', in this issue.

Billari, F, D'Amuri, F, & Marcucci, J 2013, 'Forecasting births using google'. Paper presented at the *Annual Meeting of the Population Association of America*, PAA, New Orleans, LA.

Carrière-Swallow, Y & Labbé, F 2013, 'Nowcasting with Google Trends in an emerging market', *Journal of Forecasting*, vol. 32, no.4, pp. 289-298.

Chen, T, So, EPK, Wu, L & Yan, IKM 2015, 'The 2007-2008US Recession: What Did the Real-Timer Google Trends Data Tell The United States?, *Contemporary Economic Policy*, vol. 33, no. 2, pp. 395-403. DOI: 10.1111/coep.12074.

Choi, H, & Varian, H 2009, 'Predicting initial claims for unemployment benefits', *Google Inc*.

Choi, H & Varian, H 2012, 'Predicting the present with google trends'. *Economic Record*, vol. 88 (s1), pp. 2-9.

Constant, A & Zimmermann, KF 2008, ´Im Angesicht der Krise: US-Präsidentschaftswahlen in transnationaler Sicht`, *DIW Wochenbericht*, vol. 44, 688-701.

D'Amuri, F 2009, 'Predicting unemployment in short samples with internet job search query data', *MPRA Paper 18403*, Bank of Italy.

Das, M, Ester, P & Kaczmirek, L 2011, *Social and behavioral Research and the Internet, Advances in Applied Methods and Research Strategies*, Routledge, New York et al.

Ettredge, M, Gerdes, J & Karuga, G 2005, 'Using Web-based Search Data to Predict Macroeconomic Statistics', *Communications of the ACM*, vol. 48, no. 11, pp. 87–92.

Fantazzini, D 2014, 'Nowcasting and Forecasting the Monthly Food Stamps Data in the US using Online Search Data'. *PloS one*, vol. 9, no.11, e111894.

Fondeur, Y & Karamé, F 2013, 'Can Google data help predict French youth unemployment?'. *Economic Modelling*, vol. 30, pp. 117-125.

Ginsberg, J, Mohebbi, MH, Patel, RS, Brammer, L, Smolinski, MS, Brilliant, L 2009: 'Detecting Influenza Epidemics using Search Engine Query Data', *Nature*, vol. 457, 1012 – 1014.

Greenstein, S & Zhu, F. 2012, Ís Wikipedia biased?, *American Economic Review*, 102, vol. 3, pp. 343-348.

Guzi, M & de Pedraza, P 2015, 'A Web Survey Analysis of Subjective Well-being', in this issue.

Hitsch, GJ, Hortaçsu, A & Ariely, D 2010, 'Matching and Sorting in Online Dating', *American Economic Review*, vol. 100, no. 1, pp. 130-63. DOI: 10.1257/aer.100.1.130.

Kholodilin, KA, Podstawski, M & Siliverstovs, B 2010, '*Do Google searches help in nowcasting private consumption? A real-time evidence for the US'*, DIW DP 997, DIW, Berlin.

Kuhn, P 2014, 'The internet as a labor market matchmaker', *IZA World of Labor*, no. 18. Available from: <http://wol.iza.org/articles/internet-as-a-labor-market-matchmaker>.

Kuhn, P & Mansour, H 2014, 'Is Internet Job Search Still Ineffective?', *The Economic Journal*, vol. 124, no. 581, pp. 1213-1233. **DOI**: 10.1111/ecoj.12119.

Kuhn, P and Shen, K 2012, 'Gender Discrimination in Job Ads: Evidence from China', T*he Quarterly Journal of Economic, vol. 128, no.1, pp. 287-336.* DOI: 10.1093/qje/qjs046.

Kurekova, L, Beblavy, M & Thum, E 2014, 'Using internet data to analyse the labour market: a methodological enquiry'. IZA *DP 8555*, IZA, Bonn.

Lazer, D. M., Kennedy, R., King, G., and Vespignani, A. (2014). The parable of google flu: Traps in big data analysis. Science.

Maurer-Fazio, M 2012, ´Ethnic discrimination in China's internet job board labor market`, *IZA Journal of Migration*, 1:12

Maurer-Fazio, M & Lei, L 2015, 'As Rare as a Panda": How Facial Attractiveness, Gender, and Occupation Affect Interview Callbacks at Chinese Firms', in this issue.

McLaren, N & Shanbhogue, R 2011, 'Using internet search data as economic Indicators', *Bank of England Quarterly Bulletin*. Available from: June http:// www.bankofengland.co.uk/publications/quarterlybulletin/qb110206.pdf.

Oleksandr, B 2010, *Can Google's search engine be used to forecast unemployment in Ukraine*. MA thesis, Kyiv School of Economics, Ucrania.

Reilly, S, Richey, S & Taylor, JB 2012, 'Using Google Search Data for State Politics Research An Empirical Validity Test Using Roll-Off Data', *State Politics & Policy Quarterly*, vol. 12, no. 2, pp.146-159.

Reips, UD, & Buffardi, LE 2012, 'Studying migrants with the help of the Internet: methods from psychology', *Journal of Ethnic and Migration Studies*, vol. 38, no. 9, pp. 1405-1424.

Ripberger, J T 2011, 'Capturing curiosity: Using internet search trends to measure public attentiveness', *Policy Studies Journal*, vol. 39, no. 2, pp. 239-259.

Saiz, A, Simonsohn, U 2013, 'Proxying for unobservable variables with internet document-frequency', *Journal of the European Economic Association*, vol. 11, no. 1, pp. 137-165.

Soloveichik, R 2015, ´Valuing 'Free' Entertainment in GDP: An Experimental Approach.´ Presented at the 2015 AEA meeting in Boston.

Stephens-Davidowitz, SI 2014, 'The cost of racial animus on a black candidate: Evidence using Google search data', *Journal of Public Economics*, Elsevier, vol. 118, issue C, pp. 26-40.

Su, Z 2014, 'Chinese Online Unemployment-Related Searches and Macroeconomic Indicators', *Frontiers of Economics in China*, vol. 9, no. 4, pp. 573–605.

Suhoy, T 2009, 'Query indices and a 2008 downturn: Israeli data', *Technical Report. Bank of Israel*. Available from: http://www.bankisrael.gov.il/deptdata/mehkar/papers/dp0906e.pdf.

Tefft, N 2011, 'Insights on unemployment, unemployment insurance, and mental health'. *Journal of Health Economics*, vol. 30, no. 2, pp. 258-264.

Vicente, MR, Lòpez-Menéndez, AJ & Pèrez, R 2015, 'Forecasting unemployment with internet search data: Does it help to improve predictions when job destruction is skyrocketing?', *Technological Forecasting and Social Change*, vol. 92, pp. 132-139. **DOI**: 10.1016/j.techfore.2014.12.005.

Vosen, S & Schmidt, T 2011, 'Forecasting private consumption: survey-based indicators vs. Google trends', *Journal of Forecasting*, vol. 30, no. 6, pp. 565-578.

Yang, AC, Huang, NE, Peng, CK & Tsai, SJ 2010, 'Do seasons have an influence on the incidence of depression? The use of an internet search engine query data as a proxy of human affect', *PloS one*, vol. 5, no. 10, e13728.

Yang, Y, Pan, B & Song, H 2014, 'Predicting Hotel Demand Using Destination Marketing Organization's Web Traffic Data'. *Journal of Travel Research*, vol. 53, no.4, pp. 433-447.

Zagheni, E & Weber, I 2015, 'Demographic Research with Non-Representative Internet Data', in this issue.