International Journal of Internet Science
2014, 9 (1), 1–8 ISSN 1662-5544

International Journal of Internet Science

IJIS.NET

Mining "Big Data" using Big Data Services

Ulf-Dietrich Reips[1], Uwe Matzat[2]

[1]*University of Konstanz, Germany;*
[2]*Eindhoven University of Technology, The Netherlands*

Big Data on bare feet: Free tools for analysis

While many colleagues within science are fed up with the "big data fad", empirical analyses we conducted for the current editorial actually show an inconsistent picture: we use big data services to determine whether there really is an increase in writing about big data or even widespread use of the term. *Google Correlate* (http://www.google.com/trends/correlate/), the first free tool we are presenting here, doesn't list the term, showing that number of searches for it are below an absolute minimum that is even mastered by terms like "brobdingnagian" or "sockdolager" (which apparently correlates >.80 in search patterns with searches for "oriental princess" btw.).

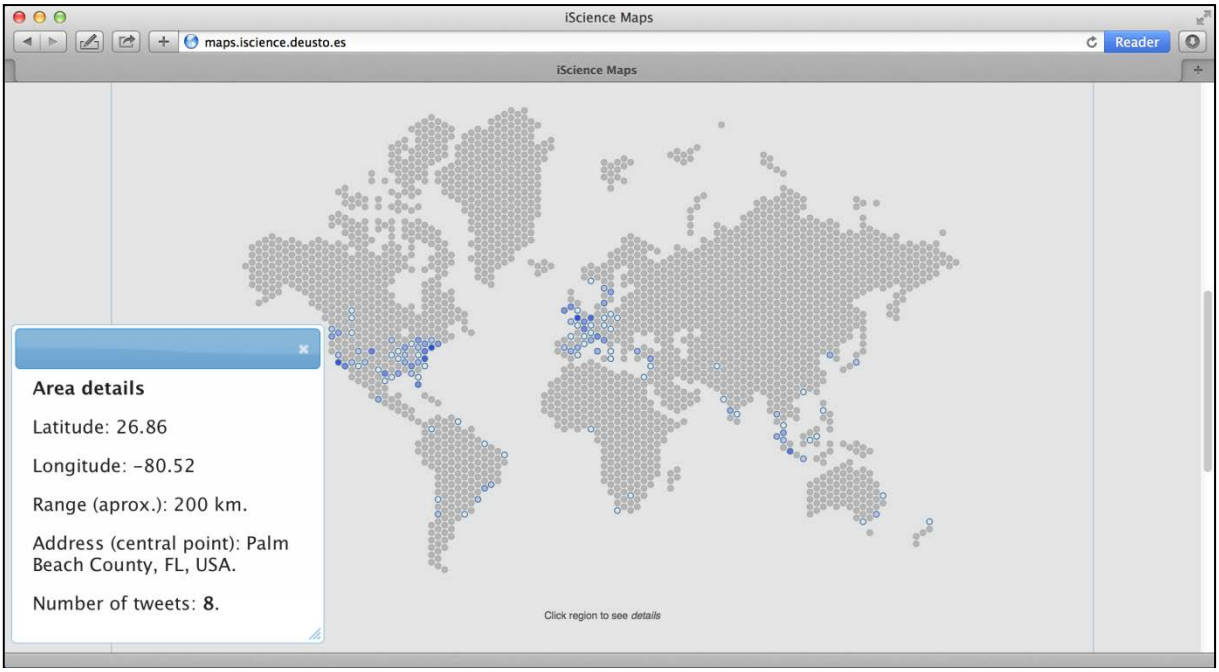

Figure 1. Searching for "big data" in tweets with iScience Maps Global search (http://tweetminer.eu).

Address correspondence to Ulf-Dietrich Reips, Psychological Methods, Assessment, and iScience, University of Konstanz, Box 31, 78457 Konstanz, Germany, reips@uni-konstanz.de. We acknowledge support by COST Action 1004 "WEBDATANET" (http://webdatanet.eu).

On the other hand, people do tweet about big data, and we use our tool *iScience Maps* (http://tweetminer.eu, Reips & Garaizar, 2011) to do a global search for "big data". Figure 1 shows a world map produced by the tool, showing the areas where most tweets originated from a random sample of about 5-10% of all tweets between 2010 and now. Each dot on the map can be clicked, showing further detail of the area, in the example case Palm Beach County in Florida. From there we can inspect each tweet for that geo-location on *Twitter*, see Figure 2, and the many results convince us there are big data in tweets.

Using *Google Books NGram Viewer* (http://books.google.com/ngrams) we can compare frequencies of appearance of terms in books. Choosing the US English digitized corpus of books–big data!–we search for appearances of our target term during the years between 1980 and 2008 (newest year possible in NGram Viewer). Indeed, albeit having peaked in 2001, "big data" seems to fare well compared to "pig data", see Figure 3. But doubts remain. Welcome to the wonderful new world of pig data mining!

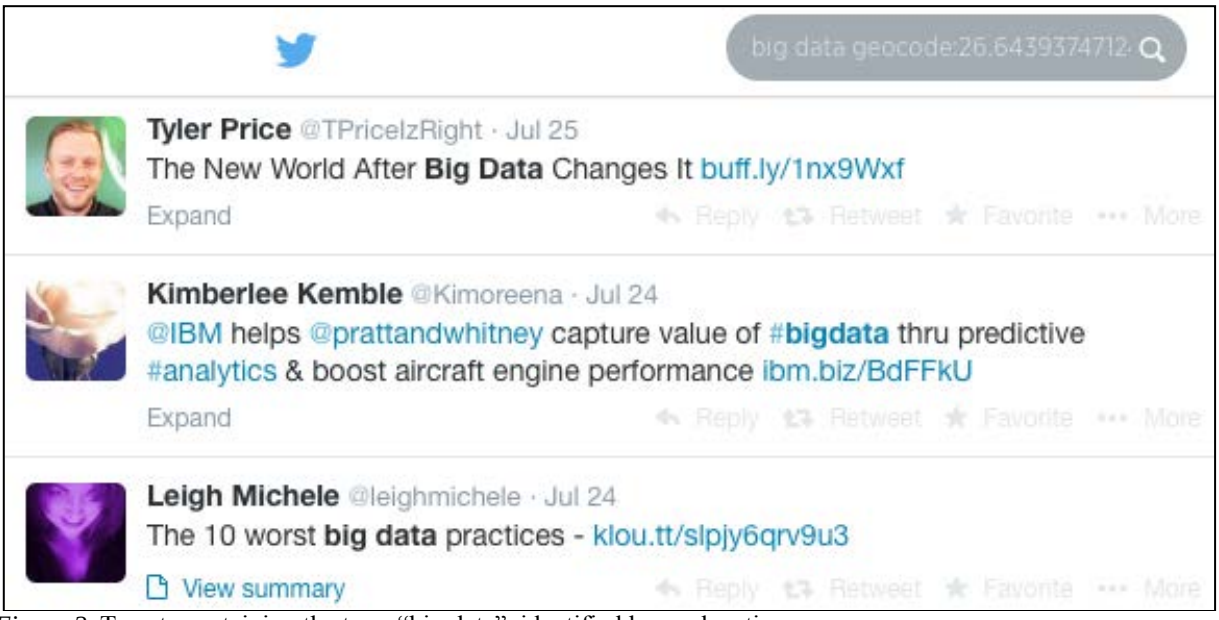

Figure 2. Tweets containing the term "big data", identified by geolocation.

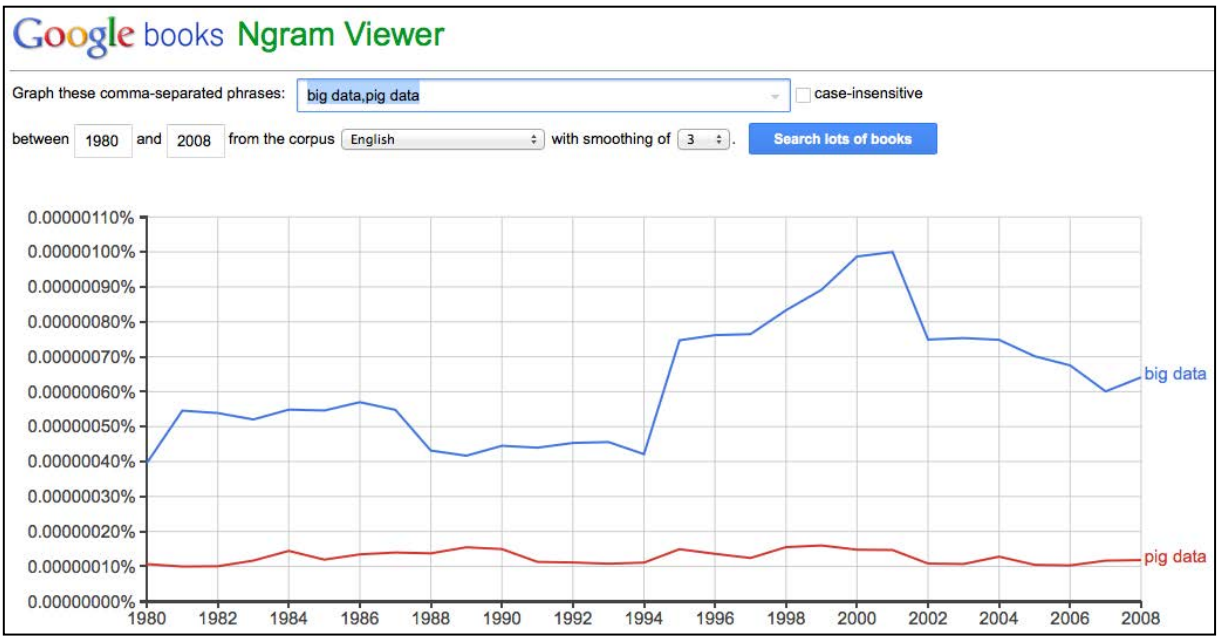

Figure 3. Identifying the frequency and timeline of appearance of the terms "big data" and "pig data" in books via Google Ngram.

Updated figures for the International Journal of Internet Science

The International Journal of Internet Science continues to be under evaluation for inclusion with the ISI Web of Science database. You may have seen from our editorials in issues 3(1), 6(1), 7(1), and 8(1) that calculations result in high journal impact. We will show below that the journal continues to be an increasingly important source for researchers.

Figure 4. Web page for the International Journal of Internet Science in Google Scholar, available at http://scholar.google.com/citations?hl=en&user=OCYy1o4AAAAJ

Our journal page on Google Scholar Citations (see Figure 4) at http://scholar.google.com/citations?hl=en&user=OCYy1o4AAAAJ allows one to track citations to the journal and its articles, link as an author, and enter one's e-mail address for automatic notifications. When we calculated the impact in late December 2011 Google Scholar showed 198 citations (Reips, 2011), in June 2013 there were 441 (Reips & Matzat, 2013), now there are 703. The three most frequently cited articles are by Göritz (2006), Freelon (2010), and Mesch and Talmud

(2006), with 168, 104, and 51 citations, respectively. The most popular recent item, our 2012 editorial on Big Data (Snijders, Matzat & Reips, 2012), has already received 31 citations.

IJIS' 2013 impact: > 6.1 ($ISI_{IJIS\ not\ included}$ > 1.3; $ISI_{IJIS\ included}$ > 2.0)

The ISI journal impact is calculated as the average number of citations in a year given to those articles in a journal that were published during the two preceding years (Impact factor, 2011). The most current journal impact that can be calculated is thus the one from 2013. We therefore need to count all citations in 2013 to citable items (articles, editorials, book reviews etc.) that appeared in the International Journal of Internet Science during the years 2012 and 2011. These citations are listed in the following paragraph.

To estimate impact factors for the International Journal of Internet Science we processed the 703 references Google Scholar returns. For the seven articles that have appeared in either 2011 or 2012 Google Scholar finds 25 citing references in 2013. The most frequently cited article, Newman, Dutton and Blank (2012), was cited seven times: by Achté (2013), Bolsover (2013), Hermida (2013), Mathurine (2013), Reips and Matzat (2013), Vis (2013), and Zhang (2013). The article by Stiglbauer, Gnambs and Gamsjäger (2011) was cited in Jadin, Gnambs and Batinic (2013), Mazarakis (2013), Reips and Matzat (2013), and Stiglbauer and Gnambs (2013). Alqudsi-Ghabra, Al-Bannai and Al-Bahrani (2011) was referenced three times, by Baabdullah, Dwivedi, and Williams (2013), Reips and Matzat (2013), and Zhang (2013). Attrill (2012) was cited in Casilli (2013), Desloge (2013), and Ko and Liu (2013). Kaye and Bryce (2012) was referenced in Collins and Cox (2013), Collins and Freeman (2013), and Costa and Schmitz (2013). Wolfe, Fisher, Reyna and Hu (2012) was cited in Fisher, Wolfe and Reyna (2013), Hu, Morrison and Cai (2013), and Wolfe, Widmer and Reyna (2013). Opgenhaffen and d'Haenens (2011) was cited in Reips and Matzat (2013) and Stark (2013). Furthermore, our 2012 editorial (Snijders, Matzat & Reips, 2012) was cited 15 times in 2013, by Al-Khouri (2013), Ketchersid (2013), Fang, Xu, Zhu, Liu, Liu and Pei (2013), Li, Tian, Smarandache and Alex (2013), Park and Leydesdorff (2013), Rats and Ernestsons (2013), Ruppert (2013), Salah, Manovich, Salah and Chow (2013), Verheij (2013), Wada and Tsubaki (2013), Wang, Li and Li (2013), Weinzierl, Lepa, Böhringer and Damm (2013), Willis (2013), Zhang, Chitkushev, Brusic and Zhang (2013), and Sheng and Zhu (2013). Trailing this extraordinary citation rate for an editorial, the Reips (2011) editorial was cited a normal two times, by Marsden, David-Barrett and Pavan (2013) and Reips and Matzat (2013). The non-peer-reviewed supplement by Steinmetz et al. (2012) was cited by Guzi and De Pedraza (2013), increasing the total number of journal impact relevant citations to 43.

Thus, the **currently official 2013 journal impact calculation is 43 / 7 = 6.1**. If we calculate only with citations that appeared in ISI journals (Thomson only includes journals in its calculations that are already in its database), then we arrive at an **ISI impact of 9 / 7 = 1.3** for the International Journal of Internet Science. If we calculate as if the International Journal of Internet Science was already part of the ISI database, then its **ISI impact is 14 / 7 = 2.0**, because our 2013 editorial cites five citable items from the International Journal of Internet Science. Importantly, all figures calculated here are conservative estimates, because more citing articles published in 2013 are not yet known and couldn't be included in the present analysis. While the 2013 figure has not changed from the 2012 figure for Google Scholar, but is down from 2012 for the ISI figures (Google Scholar: > 6.1; $ISI_{IJIS\ not\ included}$ > 2.3; $ISI_{IJIS\ included}$ > 2.9), due to the remarkably successful Freelon (2010) article only contributing to the latter, we are happy to report that the International Journal of Internet Science's impact is still on a comparatively very high level and will likely substantially increase again next year because of high citation rates to citable items in 2012 and 2013. Already, there are 43 citations in 2014 (by July!) to citable items in 2012 and 2013. **If they double by the end of the year the 2014 journal impact for the International Journal of Internet Science will be >10.**

The IJIS compares well with other long-standing journals in the field, examples are *Behavior Research Methods* (ISI impact of 1.9), *Computers in Human Behavior* (2.1), *Cyberpsychology Behavior and Social Networking* (1.8), *International Journal of Human-Computer Studies* (1.4), *New Media & Society* (1.8), *Social Science Computer Review* (1.3) and *Information Society* (1.1, all values from the 2012 JCR Social Science Edition).

The present issue

In the first article, *Use of the Internet in Capital Enhancing Ways–Ethnic Differences in Israel and the Role of Language Proficiency*, Sabina Lissitsa (Ariel University, Israel) and Svetlana Chachashvili-Bolotin (Ruppin Academic Center, Israel) utilize large scale CBS data to analyse the digital divide and digital inequalities between immigrants and Jewish veterans in the Israeli society. With respect to differences in access to the Internet they find that important gaps in access disappear after controlling for Hebrew language proficiency. With respect to gaps in engagements in Internet activities that have the potential to increase the user's social capital, they demonstrate that immigrants, once they have access to the Internet, do not show less engagement. Furthermore, within the group of Internet users, only Western immigrants, but not other groups of immigrants show less engagement in human

capital-enhancing Internet activities when compared to Jewish veterans. The authors discuss the policy implications of their findings.

In the second article, Katharina Hanel and Martin Schulze (both Heinrich-Heine-University of Düsseldorf, Germany), in their article *Analyzing the Political Communication Patterns of Voting Advice Application users*, examine the use of a popular Voting Advice Application (VAA) by German Internet users. VAAs have recently become popular in Europe. They enable individual Internet users to compare their position on policy issues with those of political parties and candidates running for election. The authors study the communication habits of VAA users and distinguish five classes of political communication patterns in a latent class analysis. Their results show that the probability of VAA use is affected by the breadth of users' political communication. The German VAA reaches a diverse group of users with regard to political communication, socio-demographic characteristics, and political interests.

Michael Schreiner (University of Education, Heidelberg, Germany), Siegbert Reiss and Karl Schweizer (both Goethe University Frankfurt, Frankfurt a. M., Germany) report on *Method Effects on Assessing Equivalence of Online and Offline Administration of a Cognitive Measure: The Exchange Test*. In assessing online-offline equivalence articles in the literature often report analyses based on one measurement only or aggregate data from repeated measurements. The present study shows by counterbalanced test-retest-design that there is the possibility of method effects that can only be detected in a cross-mode test-retest design and with appropriate analyses. In a counterbalanced test-retest-design the web-enabled version of the Exchange Test, a cognitive measure of working memory capacity that requires the reordering of four symbols in single exchanges until the sequence of these symbols corresponds to a pre-set sequence of the same symbols, was administered offline and online. Results indicated that both administration forms were parallel at the first measurement. Hence, online and offline administration seemed equivalent. However, data from the second measurement show they were not: Analyses revealed a main effect of repeated measurement and a systematic interaction of "repeated measurement" and "order of administration" (online-offline vs. offline-online). The authors argue convincingly that online-offline equivalence may often only seemingly exist at a superficial level.

Finally, in a non-peer-reviewed supplement, members of the Webdatanet COST network with *Innovation and quality in web-based data collection* present an update on past and upcoming activities within the network. We encourage you to consider joining the network's activities, there will be several meetings, training schools, and conferences in locations all over Europe.

Acknowledgements and Contact

The current issue of the International Journal of Internet Science is a result of the devotion, time, and effort of many individuals and institutions that support and help us. Our editorial assistant, Ruoyun Lin (Eindhoven University of Technology, now University of Tübingen), has handled the office work very reliably and professionally. We also thank our outgoing editorial assistant, Dr. Frederik Funke, for providing advice when it was needed.

Grateful acknowledgement goes to the University of Konstanz, the Eindhoven University of Technology, and Webdatanet (http://webdatanet.eu) for their institutional support of and partnership with the journal. And of course, we very much would like to thank the members of IIJS's Editorial Board and Panel and its many ad hoc reviewers for their valuable contributions to the quality of this journal.

Our editorials in issues 3(1), 6(1), 7(1), 8(1), and the calculations above show very high journal impact that continues to grow. The *rejection rate* for articles submitted to the International Journal of Internet Science is at **84.1%**, meaning that roughly one out of six manuscripts passes the stages of desktop rejection and repeated rounds of multiple peer review.

We encourage our readers to stay in contact with us. Our journal page on Google Scholar Citations at http://scholar.google.com/citations?hl=en&user=OCYy1o4AAAAJ allows you to track citations to the journal and its articles and enter your e-mail address for automatic notifications. Authors can even link themselves in this role. Facebook users may want to subscribe to and "like" our Facebook page at https://www.facebook.com/pages/International-Journal-of-Internet-Science/251579034934885. To get notfied of new developments at the International Journal of Internet Science, please subscribe at http://www.ijis.net/subscribe.html.

References[1]

Achté, A. (n.d.). *Is it essential for speech radio programmes to utilise social media in order to stay relevant to the audience?* Working Paper. Retrieved from http://www.hssaatio.fi/images/stories/Anne_final_paper_PRINT_EDIT.pdf

Al-Khouri, A. M. (2013). Identity management in the retail industry: The ladder to move to the next level in the internet economy. *Journal of Finance and Investment Analysis*. Online First publication.

Baabdullah, A., Dwivedi, Y., & Williams, M. (2013). *IS/IT Adoption research in the Saudi Arabian context: analysing pst and outlining future research directions*. Paper presented at the European, Mediterranean & Middle Eastern Conference on Information Systems. Windsor, United Kingdom. Retrieved from http://www.iseing.org/emcis/emcis2013/EMCISWebsite/Accepted Papers/07 Management and Organisational Issues in Information systems/emcis2013_submission_52.pdf

Bolsover, G. (2013). *News in China's New Information Environment: Dissemination Patterns, Opinion Leaders and News Commentary on Weibo*. *Opinion Leaders and News Commentary on Weibo.* Working Paper. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2257794

Casilli, A. (2013). Contre l'hypothèse de la «fin de la vie privée»: La négociation de la privacy dans les médias sociaux. [Against the hypothesis of the 'end of privacy': Negotiation of privacy in social media]. *Revue Française Des Sciences de L'information et de La Communication 3*. Retrieved from http://rfsic.revues.org/630?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+nectar-du-bouillon+(Le+nectar+du+Bouillon+des+bibliobs%C3%A9d%C3%A9s)

*Collins, E., & Cox, A. (2013). Switch on to games: Can digital games aid post-work recovery? *International Journal of Human-Computer Studies*, *72*(8-9), 654–662.

*Collins, E., & Freeman, J. (2013). Do problematic and non-problematic video game players differ in extraversion, trait empathy, social capital and prosocial tendencies? *Computers in Human Behavior*, *29*(5), 1933–1940.

Costa, C., & Schmitz, R. (2013). As modernas tecnologias de informação e comunicação eo espaço público: Explorando as fronteiras de uma nova relação. [Modern information and communication technologies and public space: Exploring the frontiers of a new relationship]. *Revista de Geografia e Ordenamento do Territorio, 3*, 197–229.

Desloge, M. (2013). *Do clinicians perceive a connection between their personal and professional habits of self-disclosure?: A study exploring self-disclosure on social networking sites*. Master Thesis. Smith College School for Social Work. Retrieved from https://dspace.smith.edu/handle/11020/24181

Fang, S., Xu, L., Zhu, Y., Liu, Y., Liu, Z., & Pei, H. (n.d.). An integrated information system for snowmelt flood early-warning based on internet of things. *Information Systems Frontiers*, 1–15. doi: 10.1007/s10796-013-9466-1

*Fisher, C., Wolfe, C., & Reyna, V. (2013). A signal detection analysis of gist-based discrimination of genetic breast cancer risk. *Behavior Research Methods*, *45*, 613–622.

Freelon, D. G. (2010). ReCal: Intercoder reliability calculation as a Web service. *International Journal of Internet Science, 5*, 20–33.

Göritz, A. S. (2006). Incentives in Web studies: Methodological issues and a review. *International Journal of Internet Science, 1*(1), 58–70.

Guzi, M., & De Pedraza, P. (2013). *A Web survey analysis of the subjective well-being of spanish workers* (No. 7618). IZA Discussion Paper.

[1] ISI-listed articles citing articles in the IJIS are marked with an asterisk.

Hermida, A. (2013). # JOURNALISM: Reconfiguring journalism research about Twitter, one tweet at a time. *Digital Journalism, 1*(1), 295–313.

Hu, X., Morrison, D., & Cai, Z. (2013). On the use of learner micromodels as partial solutions to complex problems in a multiagent, conversation-based intelligent tutoring system. In Sottilare, R., Graesser, A.. Hu, X., & Holden, H. (Eds.), *Design recommendations for adaptive intelligent tutoring systems learner modeling (Volume I)*. Orlando, Florida: U.S. Army Research Laboratory.

Impact factor (2011, December 22). In *Wikipedia, the free encyclopedia*. Retrieved December 22, 2011, from http://en.wikipedia.org/wiki/Impact_factor

*Jadin, T., Gnambs, T., & Batinic, B. (2013). Personality traits and knowledge sharing in online communities. *Computers in Human Behavior,* 29, 210–216.

*Ketchersid, T. (2013). Big Data in nephrology: Friend or foe. *Blood Purification*, *36*, 160–164. doi: 10.1159/000356751

Ko, H., & Liu, M. (n.d.). *Understanding the factors afffecting self-disclosure on facebook.* Retrieved from http://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&sciodt=0%2C5&cites=12206049894776289207&scipsc=&as_ylo=2013&as_yhi=2013#1

Li, X., Tian, Y., Smarandache, F., & Alex, R. (2013). An extension collaborative innovation model in the context of Big Data. *International Journal of Information Technology & Decision Making*, *0*(0), 1–23. doi: 10.1142/S0219622014500266

Marsden, C., David-Barrett, T., & Pavan, E. (n.d.). *FP7-288021. internet-science.eu*. Retrieved from http://www.internet-science.eu/sites/internet-science.eu/files/biblio/D6.1_Final.pdf

Mathurine, J. (2013). Walking the (social media) line: Regulation, ethics, accountability. *Rhodes Journalism Review*, *33*(1), 79–82.

Mazarakis, A. (2013). *Feedback und Anreize für die Nutzung von Web 2.0 Diensten* [Feedback and incentives for using Web 2.0 services]. München: Grin. Retrieved from http://content.grin.com/document/v211930.pdf

Mesch, G., & Talmud, I. (2006). Online friendship formation, communication channels, and social closeness. *International Journal of Internet Science, 1*, 29-44.

Newman, N., Dutton, W. H., & Blank, G. (2012). Social media in the changing ecology of news: The Fourth and Fifth estate in Britain. *International Journal of Internet Science*, *7*, 6–22.

Opgenhaffen, M. & d'Haenens, L. (2011). The impact of online news features on learning from news: A knowledge experiment. *International Journal of Internet Science*, *6*, 8–28.

*Park, H., & Leydesdorff, L. (2013). Decomposing social and semantic networks in emerging "big data" research. *Journal of Informetrics, 7, 756-765.*

Rats, J., & Ernestsons, G. (2013). Clustering and ranked search for enterprise content management. *Journal of E-Entrepreneurship and Innovation*, *4*, 20–31. doi: 10.4018/ijeei.2013100102

Reips, U.-D. (2011). Journal impact revisited. *International Journal of Internet Science, 6*(1), 1–7.

Reips, U.-D., & Garaizar, P. (2011). Mining Twitter: Microblogging as a source for psychological wisdom of the crowds. *Behavior Research Methods, 43*, 635-642. doi:10.3758/s13428-011-0116-6

Reips, U.-D., & Matzat, U. (2013). Article impact means journal impact. *International Journal of Internet Science, 8*, 1–9.

Ruppert, E. (2013). Rethinking empirical social sciences. *Dialogues in Human Geography*, *3*(3), 268–273.

*Salah, A., & Manovich, L. (2013). Combining cultural analytics and networks analysis: Studying a social network site with user-generated content. *Journal of Broadcasting & Electronic Media*, *57*, 409–426.

Snijders, C., Matzat, U., & Reips, U.-D. (2012). 'Big data': Big gaps of knowledge in the field of Internet science. *International Journal of Internet Science, 7*, 1–5.

Stark, B. (2013). The many faces of interactivity in convergent media environments: Assessing uses and effects of interactivity from a user and management perspective. In Diehl, M. K. S. (Eds.), *Media and Convergence Management* (pp. 299–315). Berlin, Heidelberg: Springer.

*Stiglbauer, B., & Gnambs, T. (2013). The upward spiral of adolescents' positive school experiences and happiness: Investigating reciprocal effects over time. *Journal of School Psychology*, *51*, 231-242.

Verheij, B. (2013). *The process of Big Data solution adoption*. Working Paper. *Delft University of Technology*. Retrieved from http://www.tbm.tudelft.nl/fileadmin/Faculteit/TBM/Over_de_Faculteit/Afdelingen/Afdeling_Infrastructure_Systems_and_Services/Sectie_Informatie_en_Communicatie_Technologie/medewerkers/jan_van_den_berg/news/doc/bverheij-big-data-adoption-process-final.pdf

Vis, F. (2013). Twitter as a reporting tool for breaking news: Journalists tweeting the 2011 UK riots. *Digital Journalism*, *1*(1), 27–47.

Wada, K., & Tsubaki, H. (2013). Parallel computation of modified stahel-donoho estimators for multivariate outlier detection. In *Computing and Big Data–2013 International Conference*. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6821008

Wang, D., Li, X., & Li, Y. (2013). China's "smart tourism destination" initiative: A taste of the service-dominant logic. *Journal of Destination Marketing & Management, 2*, 59–61.

Weinzierl, S., Lepa, S., Böhringer, G., & Damm, T. (2013). Musikempfehlungsdienstleistungen im Internet. Master Thesis. TU Berlin. Retrieved from http://www.ak.tu-berlin.de/fileadmin/a0135/Masterarbeiten/Masterarbeit_Boehringer.pdf

Willis, M. (2013). *Interpreting "Big Data": Self-quantifiers, data distance, and rock star analysts*. Dissertation Proposal. Boston College. Retrieved from http://www.sarkisian.net/sc781/willis_proposal.pdf

*Wolfe, C., Widmer, C., & Reyna, V. (2013). The development and analysis of tutorial dialogues in AutoTutor Lite. *Behavior Research Methods, 45*, 623–636.

Zhang, Z. (2013). *Rock the Journalism—The function of Weibo in foreign media's news practice in China*. Working Paper. *Available at SSRN 2258101*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2258101

Zhang, P., Chitkushev, L., Brusic, V., & Zhang, G. L. (2013, September). Biomarkers in Immunology: from Concepts to Applications. In *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics* (pp. 826). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2512358

Sheng, D., & Zhu, Z. (2013). 基于大数据的教育技术研究新范式 [A new paradigm of educational technology research: Big data)]. *电化教育研究 [Education Research], 34*(10), 5–13.