Levels of Trace Data for Social and Behavioural Science Research

Kevin Crowston

Syracuse University School of Information Studies

To appear in Matei, S., Goggins, S. & Jullien, N. (Eds.),

Big Data Factories: Collaborative Approaches. Springer.

Abstract

The explosion of data available from online systems such as social media is creating a wealth of trace data, that is, data that record evidence of human activity. The volume of data available offers great potential to advance social and behavioural science research. However, the data are of a very different kind than more conventional social and behavioural science data, posing challenges to use. This paper adopts a data framework from Earth Observation science and applies it to trace data to identify possible issues in analyzing trace data. Application of the framework also reveals issues for sharing and reusing data.

Introduction

The social and behavioural sciences are said to be on the verge of a data-driven revolution. There is great interest in the scientific inferences that can be drawn from digitally-captured records of human activity, such as in an online community, user-generated content systems, search engine searches, cellular phones or digital badges (Lazer et al., 2009; Manovich, 2012), what Howison, Wiggins, and Crowston (2011) call trace data. As Agarwal, Gupta, and Kraut (2008) stated: "Most transactions and conversations in these online groups leave a digital trace ... this research data makes visible social processes that are much more difficult to study in conventional organizational settings." For example, researchers have noted that social media data show great potential to address long-standing research questions about human behaviour (Edwards, Housley, Williams, Sloan, & Williams, 2013). Chang, Kauffman, and Kwon (2014) go as far as to suggest that the rise of big data is leading to a "paradigm shift in scientific research methods", what Watts (2007) called a "21st century science".

However, these claims about the transformative capacity of big data for the social and behavioural sciences need to be viewed with caution. Records of online behaviour certainly amount to terabytes of data, but these data are of a very different sort than social and behavioural scientists would obtain from more traditional research approaches such as surveys or experiments and so require different research approaches. The most closely related commonly used data are events data in international relations (e.g., McClelland, 1967) and consideration of the issues in using these data provides some insights.

The goal of this paper is to discuss differences between trace data and traditional social and behavioural science data and the implications of these differences for using trace data for social and behavioural science research. The main contribution of the paper is a more precise vocabulary for talking about the processes of using trace data and the products of these processes that clarifies different levels of processing. The framework also highlights issues involved in sharing and reusing trace data.

Framework: From trace to variable

Howison et al. (2011) identify three differences between long-used sorts of social and behavioural research data and trace data: trace data are event-based, longitudinal and most importantly, found, rather than created to support research. These features are found in other settings, e.g., political scientists have built databases of events data (e.g., the World Events Interaction Survey, WEIS (McClelland, 1967)) and longitudinal data are common across many fields.

The difference that is key for our argument is the final point. Data from scientific sources such as surveys or experimental measurements are most often purposefully collected to measure constructs of theoretical interest. Rigourous quantitative research employs carefully-refined instruments with known psychometric properties to ensure that the instrument reliably measures what it should. (Poorly-designed research might be sloppier, but is hard to argue as a model for future research.) In contrast, social media and other trace data are records of human activity without inherent theoretical import. As Howison et al. (2011) say, "Wikipedia was not designed to test theories about knowledge production, nor are corporate email systems designed to collect research data". Rather, these data need to be interpreted to be useful for social and behavioural scientists.

In some ways, the interpretive flexibility of trace data is an advantage. They reflect actual behaviour rather than opinion, belief or attitude, and can be used for different kinds of studies, unlike data from most surveys or experiments that measure specific constructs. The implication though is that trace data require considerable additional processing to be useful for research. Unfortunately, the term "data" is overloaded and does not distinguish between different kinds of data, processed or not, leading to potential confusion and unwarranted optimism about the utility of found data. A framework is needed to sort out the different kinds of data. The main contribution of this paper is to develop such a framework.

Levels of data in the earth sciences

This situation—having multiple kinds of data with different levels of scientific interpretation—is by no means unique to the social and behavioural sciences or to trace data. It is thus instructive to examine how the distinctions among data with different kinds of processing are addressed in other disciplines. The earth sciences provide a particularly helpful framework, as the kinds of data created by processing satellite observations have been given different labels with clear definitions in this research community.

The NASA Earth Observation program distinguishes data at 6 levels of processing, as shown in Table 1 (from Parkinson, Ward, and King (2006)). Data at each level is derived from the data at the lower level through defined data-processing steps. For example, consider a satellite collecting data about the earth using a sensor that receives some signal from the earth (e.g., light or radar reflections) that can be interpreted as evidence for a geophysical variable (e.g., temperature or sea wave heights). To move from Level 1 to Level 2 data in the framework, for example, data from the sensor are interpreted to reveal geophysical variables, e.g., certain wavelengths of light indicate particular kinds of vegetation; particular scattering of radar indicate wave heights. In the earth sciences, level 0 and 1 data are generally not useful for research, other than for studies of the properties of the satellite and its sensors. Instead, earth scientists want Level 2 or 3 data, data about a geophysical process, plotted on a map. That is, rather than a time series of voltages from a sensor, scientists want a map showing what vegetation is where (for example).

Level	Definition
Level 0	Reconstructed, unprocessed instrument/payload data at full resolution; any and all communications artifacts, e.g., synchronization frames, communications headers, duplicate data removed.
Level 1A	Reconstructed, unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters, e.g., platform ephemeris, computed and appended but not applied to the Level 0 data.
Level 1B	Level 1A data that have been processed to sensor units.
Level 2	Derived geophysical variables at the same resolution and location as the Level 1 source data.
Level 3	Variables mapped on uniform space-time grids, usually with some completeness and consistency.
Level 4	Model output or results from analyses of lower level data, e.g., variables derived from multiple measurements

Table 1. Levels of Earth Observation Data
(from raw data as collected to processed and synthesized data (Parkinson et al., 2006))

Example: From tweet to variable

We can apply the Earth Observation data framework to the case of trace data. We use as an example data from the social media platform, Twitter. By analogy to Table 1, we define different levels for Twitter data, as shown in Table 2. Level 0 are the raw tweets, e.g., collected from a Twitter API. Level 1 adds metadata about the tweets as they were collected (e.g., time, date, sender). Level 2 interprets the tweet content as indicating some social and behavioural science variable of interest (e.g., political discourse, topic or sentiment). Level 3 aggregates evidence from multiple tweets to develop data about the unit of analysis of interest for the study: an individual, a political figure, a topic, etc. Note that our interpretation of this level for trace data differs somewhat from the definition of level 3 in the original Earth Observation framework, which refers to mapping data to a uniform space-time grid. Here we generalize that concept to mapping data to other conceptual spaces. Finally, Level 4 is created by linking data from the tweet corpus to data from other datasets or to a model.

The same distinctions can be made for other kinds of trace data. For example, a study about leadership in an open source project (Crowston, Wiggins, & Howison, 2010) might draw on developer emails (Level 0), annotated with information about the sender (e.g., the role in the project, Level 1), coded for evidence of leadership behaviours (Level 2), aggregated to suggest which members of the project exhibit signs of being project leaders (Level 3) and linked to other data about contributions or project outcomes (Level 4).

As with satellite data, for social and behavioural science research, Level 0 or 1 social media data are unlikely to be of much interest for research: raw Tweets or email messages by themselves and "as is" are not that useful for research. However, it is at this level that we see the explosion in available data. To test theory, social science researchers need data at Level 3, which corresponds to the kind of data a researcher would get from a survey. Unfortunately, such data are much less readily available. An implication for development of data archives is that it would likely be more useful to focus these on higher levels of data.

Level	Definition
Level 0	Raw tweets.
Level 1	Raw tweets annotated with ancillary information, e.g., sender information.
Level 2	Derived social and behavioural science variables at the same resolution as Level 1 (i.e., coded tweets).
Level 3	Derived social and behavioural science variables at unit of analysis of interest (e.g., data about individuals).
Level 4	Model output or results from analysis that merges multiple sources of data.

Table 2. Levels of Twitter Data

Level	Definition
Level 0	Raw email messages.
Level 1	Raw email messages annotated with ancillary information, e.g., sender information.
Level 2	Derived social and behavioural science variables at the same resolution as Level 1 (i.e., coded email messages).
Level 3	Derived social and behavioural science variables at unit of analysis of interest (e.g., data about individuals).
Level 4	Model output or results from analysis that merges multiple sources of data.

Table 3. Levels of Open Source Development Data

Discussion: Moving up the levels

The issue then is how to process data to move from level 0 to level 3 or 4. For geospatial data, scientists have developed data-processing algorithms based on their knowledge of the physical properties of the satellites and sensors, and the geophysical properties of the systems being observed: e.g., known performance of instruments converting radiation to a sensor signal, mathematical models for translating between satellite position and orientation to the observed location on the ground, or models of what different vegetation look like to support inference from an observed intensity of light at a particular wavelength to geophysical data about ground cover. Even with this level of theoretical development and knowledge of the geophysical processes, automated algorithms are not always sufficient by themselves. For example, for best precision, images might have to be adjusted by hand by manually matching known benchmarks on the base map. Predicted geophysical variables (e.g., vegetation) might need to be ground truthed to verify the reliability of the interpretation.

Interpretation of data is also a common analysis approach in social research. Qualitative researchers frequently employ the technique of content analysis (Krippendorff, 2004) to code textual documents for theoretical constructs of interest. In the framework above, content analysis is a technique to move data from Level 0 or 1 to Level 2. The political science databases described above take newspaper or wire series press reports as Level 0 data and code them against an event coding scheme that identifies actors and actions of theoretical interest (Veen, 2008). For example, WEIS's (McClelland, 1967) coding scheme codes events reported by the New York Times into 61 categories of action. Researchers employing observational techniques develop coding schemes that identify which observed behaviours are of interest, essentially skipping Levels 0 and 1 and collecting data directly at Level 2. Considering social media again, tweets might be interpreted as indicating support for or opposition to a political candidate.

Unfortunately, moving up levels of social media and other social and behavioural trace data is less routinized and predictable than for Earth observation data and even for international relations. Some of these problems are inherent in the nature of the social and behavioural sciences. The processes by which the social and behavioural constructs of interest (e.g., leadership) get reflected in recorded behaviours (e.g., emails) are much less regular than the corresponding geophysical processes (e.g., vegetation reflecting light). But there are also differences that reflect the rigour and reproducibility of the data processing in research practice. At present, social and behavioural researchers typically derive variables from observed data in their own idiosyncratic ways. As with satellite data, processing may require manual intervention and validation, making the process hard to replicate or even to completely describe. And unfortunately, provenance of data is often not well recorded, so how these steps were carried out may be unclear to those reading the research. For example, Liang and Fu (2015) found that they could not reproduce the results of 6 out of 10 studies of Twitter they examined using a random sample of Tweets, which they attributed to "variations of data collection, analytic strategies employed, and inconsistent measurements".

We next discuss the specific issues involved in each step of the chain from event to Level 4 data to further explore the issues involved in using trace data for social and behavioural research.

Collecting Level 0 data

Level 0 is the lowest level in the framework, but it is worth noting that even Level 0 data has had some processing. As noted in Table 1 above, satellite data is processed to remove communications artefacts. For trace data or social media data, there is a comparable process of removing artefacts from the data collection that needs to be documented (e.g., removing spam emails from an email archive before analysis). However, additional problems can arise. Howison et al. (2011) point out that collecting trace data from an information system raises a number of validity issues. They focus on validity issues for social network analysis, but a number of their issues are more general. Two relate in particular to the collection of trace data from an information system, that is, the creation of what we are labelling Level 0 data: "system and practice issues" and "reliability issues".

The first issue refers to the need to understand actual system use in order to be able to interpret the data created. An example given by Howison et al. (2011) is a group-support system that requires individuals be team "members" to access team documents, leading to many people being listed as members mostly to enable document access. The point is that the system definition of a team member in this case is different than the offline definition, posing challenges for interpreting the system data about membership.

The second issue refers to the need to assure that the data have been collected reliably by the system itself. As system database are maintained for the operation of the system rather than for scientific purposes, decisions about data collection are usually made for operational reasons rather than to preserve the scientific integrity of the data (e.g., system databases might be periodically purged of old data for performance reasons). However, those decisions and their consequences are unlikely to be visible to an external researcher. In political science, similar concerns are raised about biases in news sources' selection of events to report and indeed, whether certain relevant events are reported at all (McClelland, 1983). boyd and Crawford (2012) note that most Twitter APIs yield a subset of tweets, but it is not clear how that subset is selected, making the generalizability of the sample questionable.

Data processing from Level 0 to 1

To move from Level 0 to Level 1, data are annotated with additional information about the observations that were made. The issues here for trace data parallel those for collecting Level 0 data, namely ensuring the completeness and reliability of the data collected. As an example, email messages include a time stamp (a kind of metadata for the email observation), but may omit the time zone, making the interpretation of the timing of messages problematic (Howison et al., 2011).

Data processing from Level 1 to 2

For Earth observation satellite data, data at Level 2 are the results of interpreting satellite sensor data as geophysical variables. Such an interpretation is inherently theoretically based. For example, to interpret light reflected from the Earth as evidence of vegetation requires a good model (possibly empirical rather than strictly theoretical) of how different kinds of vegetation reflect light under varying conditions.

In the world of trace data, traces need to be interpreted to serve as evidence for social and behavioural concepts of interest. For example, political science events databases arise from human or machine coding of events reported in news stories. As Venn (2008) notes, each "event scheme is informed by theoretical assumptions about the international system and the interaction of political actors". However, Venn notes that researchers often wan to analyze variables such as the level of cooperation or conflict between two countries, which requires further interpreting the events as evidence for these constructs.

Returning to social media data, to create Level 2 Twitter data, raw tweets can be content analyzed for any number of social and behavioural science concepts, e.g., for what topic a tweet addresses or what speech act the tweet represents (Hemphill & Roback, 2014). Again, such interpretation relies on a theory about the concept in question and how it affects or is reflected in the observed behaviour. As Manovich (2012) notes, online behaviour is "not a transparent window into peoples' imaginations, intentions, motifs, opinions, and ideas" and thus needs to be carefully and thoughtfully interpreted.

In some cases, interpretation is sufficiently well understood and mechanical that it can be done automatically. For example, natural language processing techniques have been developed to determine the sentiment of a text, albeit with some imprecision. Recent political science events databases are generated by automatic coding of wire service articles (Veen, 2008). In other cases, human judgement might be needed, which can pose a significant bottleneck for processing as well as potentially adding individual human errors or biases to the judgements. These issues have led to the development of citizen science projects that have multiple human volunteers assess images or other data. In many cases though, this processing is more akin to processing to Level 1B in the original Earth observation framework. For example, annotating an image of a galaxy for its shape (as in the original Galaxy Zoo) or an image of an animal with the species (as in Snapshot Serengeti) provides useful information, but the data are still about the image with limited theoretical import. (This argument might also be made about political science events.)

Unfortunately, in many cases, analyses of trace data essentially skip this processing step: data instead remain at the level of the original phenomenon. For example, a social network can be constructed from email messages by interpreting replies to a message as creating links. While this process does yield a network, the theoretical import of such a network is unclear. At best, a reply suggests that the person replying read and was interested in the message, but many others likely also read the message without feeling a need to reply. Similarly, data from digital badges can identify how people move through space or who they have been close to, but without some theory about movement or propinquity, it is hard to interpret the data as evidence for research. Even when an interpretation is made, it may not be theoretically justified. In a study of published communications and social computing studies of hyperlinks, Twitter followers and retweets (3 kinds of trace data), Freelon (2014) found that “substantial proportions of articles from both disciplines failed to justify the social implications they imputed to trace data”, “more extensively in the latter” discipline (social computing).

Data processing from Level 2 to 3

Level 3 data require aggregating data from Level 2. To aggregate the data requires picking a unit of analysis and linking related observations. An obvious unit for aggregation for trace data of behaviours is the person involved in the recorded activity. For example, political sciences events are coded for the actor and recipient of an event to permit such aggregation. For social media data, one might link submitted tweets by the user ID. However, just as an event may not have an identified actor (e.g., an anonymous terrorist attack), in some settings users may have an option to work anonymously (Panciera, Priedhorsky, Erickson, & Terveen, 2010), which means that a user ID might not capture all work done by a person. In particular, it may omit work done while lurking in the early stages of involvement with a group (e.g., reading others’ posts), creating problems for studies of new members in particular. Data might also be aggregated to a population, e.g., to determine the average properties of particularly kinds of contributors. Interpreting such aggregated data requires more attention to the nature of the sample. As a specific example, boyd and Crawford (2012) note that “it is an error to assume ‘people’ and ‘Twitter users’ are synonymous: they are a very particular sub-set”.

Data processing from Level 3 to 4

Level 4 data are derived from the composition of different data sets. Unfortunately, such composition is difficult for trace data and for social and behavioural science data more generally. The problem is that to connect datasets, there needs to be a way to link the data. In database terms, there needs to be common field on which to join the data tables. More simply, the different data need to be about the same thing.

For geospatial data from a satellite, data are typically tied to a particular spot on the earth. There are difficulties in working out which spot a sensor has measured and aligning data collected in different patterns or at different resolutions, but once these issues are addressed, then collected data can be connected to other data about that spot, no matter how it was collected. The same principle also applies in astronomy: data about the same spot in the sky can be connected.

Alternately, data may be about a specific entity with a stable identity, allowing linking. For example, astronomical data can be thought of as being about particular celestial objects (stars or galaxies) that can be linked from dataset to dataset. Finally, data may be about an identifiable class of object. For example, ecological data might be about particular species and so of interest to others who study the same or similar species. Astronomical data can be about a particular type of star.

In the social sciences, data sets may sometimes be about identifiable entities, allowing linking of datasets. In particular, economic data are often about countries or companies, which makes it possible to link data about the same countries or companies (though even here there can be issues in making connections). This situation may also describe data collected about entire online communities: different perspectives on the various language Wikipedia projects can be compared.

For behavioural research though, data are likely to be about people. As with aggregating data to Level 3, it is possible to link data from a system for a particular user by using the user's system ID. However, such an ID likely has little meaning beyond the system. We might therefore be able to link a user across multiple Twitter databases, but not Twitter and anything else, meaning that we might not know anything more about users of a social media site than what they post. For the specific case of free/libre open source software developers, Crowston, Wei, Li, and Howison (2006) argued that developers are often attached to the user IDs and so attempt to use them on different sites, but it is not clear that this phenomenon generalizes. Without knowing the identity of the specific respondents, it is not possible to link individual responses to other data. At most, data can be cumulated with other data to increase the sample size, as in a meta-analysis, deepening the analysis but not broadening it.

Conclusion: Recommendations for future research

The framework presented here reinforces several recommendations that have already been made about social and behavioural research. First, there are clear implications for reporting research. Specifically, research using trace or social media data needs to provide more detail on the processing that took data from level to level. It would also be valuable to share techniques for moving between levels to promote reproducibility of research and to allow researchers to leverage each other's findings.

There are further implications for sharing data. Researchers sometimes face limitations on sharing Level 0 or Level 1 data. For example, the terms of service of some social media sites limit sharing such raw data. Data from proprietary services may simply be unavailable outside the organizations that run them (Lazer et al., 2009). It is worth noting that there are serious problems for the reproducibility of science if the datasets underlying studies can't be shared, meaning that other researchers are unable to check or reproduce findings. On the other hand, Level 2 or 3 data may not be so encumbered, and these are the levels that are likely to be of the most interest to other researchers. Coupled with a sufficiently detailed description of the data processing used to create the data, a Level 2 or 3 dataset may be sufficient, at least for checking results.

However, the discussion of creating Level 4 data suggests that even Level 2 or 3 data may be difficult for others to link to their own data. To useful, the data needs some ID on which to link the data. But if researchers know the identity of users, it is likely that they will not be able tell anyone else in order to maintain the privacy of participants (Daries et al., 2014). A possible direction for research is to apply the notion of a species as an entity for data collection. If researchers using trace data could agree on clearly defined classes of users of interest, then data might be shareable and reusable when aggregated at that level.

In summary, it is unarguable that the increased penetration of information technology across the spectrum of life activities is creating a vast trove of trace data and that such data can be of great interest to social and behavioural scientists. However, such trace data are different in kind from data more traditionally used in social and behavioural research. Applying a framework from Earth observation studies, we have shown how raw trace data must be processed to create data useful for advancing social and behavioural studies and identified the issues that arise. A particularly problematic issue is identifying what data are about in order to be able to link across datasets.

References

Agarwal, R., Gupta, A. K., & Kraut, R. (2008). Editorial overview: The interplay between digital and social networks. *Information Systems Research, 19*(3), 243–252. doi: 10.1287/isre.1080.0200.

boyd, d., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, *15*(5), 662–679. doi: 10.1080/1369118X.2012.678878.

Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems, 63*, 67–80. doi: 10.1016/j.dss.2013.08.008.

Crowston, K., Wei, K., Li, Q., & Howison, J. (2006). Core and periphery in Free/Libre and Open Source software team communications. In *Proceedings of Hawai'i International Conference on System System (HICSS-39)*, Kaua'i, Hawai'i.

Crowston, K., Wiggins, A., & Howison, J. (2010). Analyzing leadership dynamics in distributed group communication. In *Proceedings of Hawaii International Conference on System Sciences (HICSS-43)*, Lihue, HI. doi: 10.1109/HICSS.2010.62.

Daries, J. P., Reich, J., Waldo, J., Young, E. M., Whittinghill, J., Ho, A. D., Seaton, D. T., & Chuang, I. (2014). Privacy, anonymity, and big data in the social sciences. *Communications of the ACM, 57*(9), 56–63. doi: 10.1145/2643132.

Edwards, A., Housley, W., Williams, M., Sloan, L., & Williams, M. (2013). Digital social research, social media and the sociological imagination: Surrogacy, augmentation and re-orientation. *International Journal of Social Research Methodology, 16*(3), 245-260. doi: 10.1080/13645579.2013.774185.

Freelon, D. (2014). On the interpretation of digital trace data in communication and social computing research. *Journal of Broadcasting & Electronic Media, 58*(1), 59-75. doi: 10.1080/08838151.2013.875018.

Hemphill, L., & Roback, A. J. (2014). Tweet acts: How constituents lobby congress via Twitter. In *Proceedings of ACM conference on Computer Supported Cooperative Work & Social Computing*, Baltimore, MD, pp. 1200–1210.

Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis for the study of online communities. *Journal of the Association for Information Systems, 12*(12), 323–346.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Newbury Park, CA: Sage.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Life in the network: the coming age of computational social science. *Science, 323*(5915), 721–723. doi: 10.1126/science.1167742.

Liang, H., & Fu, K.-w. (2015). Testing propositions derived from Twitter studies: Generalization and replication in computational social science. *PLoS ONE, 10*(8), e0134270. doi: 10.1371/journal.pone.0134270.

Manovich, L. (2012). Trending: The promises and the challenges of big social data. In M. K. Gold (Ed.), *Debates in the Digital Humanities* (pp. 460-475). Minneapolis, MN: University of Minnesota Press.

McClelland, C. A. (1967). *Event Interaction analysis in the setting of quantitative international relations research*. Department of Political Science, University of Southern California. Los Angeles.

McClelland, C. A. (1983). Let the user beware. *International Studies Quarterly, 27*(2), 169-177. doi: 10.2307/2600544.

Panciera, K., Priedhorsky, R., Erickson, T., & Terveen, L. (2010, 04/2010). Lurking? Cyclopaths? A quantitative lifecyle analysis of user behavior in a geowiki. In *Proceedings of ACM Conference on Computer-Human Interaction (CHI)*, Atlanta, GA.

Parkinson, C. L., Ward, A., & King, M. D. (Eds.). (2006). *Earth Science Reference Handbook: A Guide to NASA's Earth Science Program and Earth Observing Satellite Missions*. Washington,

D.C.: National Aeronautics and Space Administration. Available from: http://eospso.gsfc.nasa.gov/sites/default/files/publications/2006ReferenceHandbook.pdf.
Veen, T. (2008). Event Data: A method for analysing political behaviour in the EU. In *Proceedings of Prepared for the Fourth Pan-European Conference on EU Politics, Riga, Latvia*. Available from: http://www.jhubc.it/ecpr-riga/virtualpaperroom/002.pdf.
Watts, D. J. (2007). A twenty-first century science. *Nature*, *445*(7127), 489-489. doi: 10.1038/445489a.