The Challenges of Data Quality

Kate B. Showers
Centre for World Environmental History, University of Sussex

There is no question that using quite different sources of information enhances understandings of environmental processes and events. A critical analysis of a wide range of data types is possible when different sources are used to interrogate each other. The result is more nuanced and sophisticated interpretations of complex events than could be arrived at from a single data source. The Indian Ocean World network's intention to engage in such a process on a very broad scale offers enormous opportunity, while the potentials of digitalization promise great power. However, the proposed identification of new data sources, facilitation of novel data combinations, and increased technological access to data sets – disconnecting them from their original contexts - poses great challenges.

One of my current research foci is the quality of numbers in national, regional and global databases currently used to describe soil and water properties and availability, as well as the consequences of their use (and mis-use) in models and policy formulation. The goal of this work is to draw attention to the dangers of formulating policy based on inadequate, incomplete or misleading, but apparently authoritative, databases. I am developing a research project that will 1) examine the sources of databases used in the description of land and water availability quality and 2) trace their past application to policy and projects. The result will be a mapping of the generation and flow of numbers; their inter-relationships; and the intellectual networks they support which will, in turn, be linked to narratives, policy and the consequences of action. Not only will this provide a further critical perspective for narrative analysis of historical landscape conditions, but should also aid in the refinement of current and future approaches to perceived problems. The entire exercise should serve as a vehicle to call attention to the United Nations' statisticians repeated calls for increased and improved data collection, as well as to draw attention to the responsible management and use of datasets and databases. The latter could contribute to discussions of 1) appropriate architectures for digitalization technologies for historical datasets; and 2) how best to format historical qualitative and quantitative data so that they can be meaningfully connected to current measurement analytical efforts.

The research will have two major focal points: 1) the data bases themselves and 2) the consequences of their application. Databases will be identified, the process of their construction elaborated, and any associated caveats about number quality or appropriate use will be collected. Interviews with database compilers and users (modelers, planners, policy-makers) will provide deeper levels of analysis. The consequences of both qualitative and quantitative database application in policy and practice will be assessed from literature surveys. Finally, interviews with statisticians, modelers and policy makers will be conducted to learn more about the limits and possibilities of data currently available, the potential for their strengthening or extension through the use of historical

materials, and the limits and possibilities of all data sources for accurate prediction of future events.

Collaboration would strengthen this project in several ways. Comparative analysis at inter-continental and Indian Ocean basin scales - and in past centuries - would compliment my own experiences with, and knowledge of, regions on the African continent in the 20th and 19th centuries. Collaboration with botanists, meteorologists and hydrologists would expand the project's focus beyond databases concerned with landscape dynamics, soil processes and soil-water-plant relationships. Archivists could help identify historical databases to consider, the origins of values in data sets, and past data combination exercises. Conversation with data managers would facilitate consideration of new ways to examine contemporary and historical data quality - especially those from beyond the possibility of confirmation by oral histories collected from direct observers. I hope that my investigations of data quality would assist researchers making highly subject-matter specific studies, those interested in the more distant past, as well as designers of digitalization procedures interested in problems of quality control for past, present and future applications.

I am currently a collaborator on an inter-disciplinary research proposal (economist, engineer, physicist, sociologist, environmental historian) to study the EU's renewable energy technologies and policies with respect to the proposed European Supergrid. My area of research will be the environmental history, consequences and policy implications of concentrated solar power (CSP) generation and desalination plants in the Sahara desert. In the recent past, I was the leader of a fully collaborative and interdisciplinary (biologist, development studies, history, environmental history, soil science/agronomy, anthropology) international research team in Lesotho, southern Africa studying the arrival and spread of the ideas of 1) tree planting, 2) afforestation. and 3) forest management, as well as of exotic trees, in a subhumid temperate grassland ecosystem. Both data collection and analysis required interdisciplinary consultation during which each researcher learned skills from the others. Data collection ranged from botanical surveys to interviews and archival searches, and analysis required both quantitative and qualitative tools.

In the more distant past I confronted the problems of data quality associated with soil classification, soil erosion and soil conservation literature. It was normal for measurements and estimates to be combined as well as for carefully defined observations to be used as proxies or as the basis for projection to larger, and largely undefined areas. Authority assigned to data bases came from different types of representation (from tables and maps through GIS images) and form the collecting institutions. In my research, conflicting field research results (soil analysis, survey data and oral histories) were compared and contrasted with quantitative and numerical descriptions found in technical and archival materials to achieve understandings of landscape dynamics as well as the reasons for the generation and perpetuation of unreliable data (Showers 2005, 1996, 1989, 1982; Showers and Malahleha 1992). More recently, I traced the origin of the EU Renewable Energy mandate and Biofuels policy from models to policy formulation, and surveyed the consequences of implementation for African landscapes (Showers 2012). These research experiences impressed upon me the need for wider attention to the

'troublesome detail' of the meaning and accuracy of data, their presentation and use in models, and of their power in guiding policy.

I look forward to the possibility of collaborative opportunities in new geographical and intellectual regions through the Indian Ocean World network.

References

Showers, Kate B. 2012 Land Use from Below: Biofuels, Urbanization and Sustainable Soil Management in Europe and Africa *IN* Verena Winiwarter and Martin H. Gerzabek (eds) *The challenge of sustaining soils: Natural and social ramifications of biomass production in a changing world*. Vienna: Austrian Academy of Sciences.

Showers, Kate B. 2005 *Imperial Gullies: Soil erosion and conservation in Lesotho.* Ohio University Press, Athens.

Showers, Kate B.1996 Soil erosion in the Kingdom of Lesotho and the development of historical environmental impact assessment, *Ecological Applications*, 6(2):653-664.

Showers, Kate B. and Gwendolyn M. Malahleha. 1992. Oral evidence in historical environmental impact assessment: soil conservation in Lesotho in the 1930s-1940s, *Journal of Southern African Studies* 18(2):276-96.

Showers, Kate B. 1989 Soil erosion in the Kingdom of Lesotho: origins and colonial response, 1830-1955. *Journal of Southern African Studies* 15(2):263-286.

Showers, Kate B. 1982. Assessment of the land use potential of ha Makhopo, Lesotho, Southern Africa: a holistic approach to agricultural evaluation. Unpublished PhD dissertation, Cornell University.