

The Data Corridor: You, The Past, and The Future

Most, if not all, of the natural sciences have caches of “historical” (aka “heritage”) data - observations or records made before the digital era. Access to those data can be critical when it comes to validating a model which represents how conditions have changed; anthropogenic influences were already quite strong when the digital era commenced, so any changes that are strictly natural will already be convolved with the side-effects of human activities. Natural changes are chaotic, and the only way to model them accurately is to include data from long enough ago. Some sciences have managed to recover and digitize substantial portions of their older data, but those are in a minority; many have only been able to recover a small amount, and some have no idea where even to start looking. Ironically, we now have the tools to manage full-scale data rescue, yet the political drive and the associated funding are not in place.

The need has now become both topical and urgent, and “Data Rescue” is where the necessary drive towards comprehensive rescue and management begins.

The RDA/CODATA Data Rescue Group plans to publish a book on “Data Rescue”. It will set the scene, describe and expand the rationale, define the benchmarks, and present selected Case Studies. Which Case Studies to include can be a topic for open discussion at this session. An added intention is that a (possibly annual) Newsletter (online) will be issued to update the Case Studies and report new ones.

The RDA/CODATA Data Rescue Group has also been charged with presenting ”Guidelines” for rescuing data. Activities involved in the recovery, digitizing, preserving of originals, plus the dissemination and archiving of the digital versions, all need to be included in those Guidelines. The Group will hold a Workshop on Data Rescue in Boulder in the Fall of 2016, where topics for the Guidelines will also be addressed. However, the Workshop will probably appeal more to data producers while the RDA community tends to include more data managers. This session will therefore seek input from aspects of data management regarding contributions to those Guidelines.

The need to recognize the scientific potential of heritage data has never been greater than in relation to present-day climate change, so its main theme is highly relevant to Plenary 6. It also resonates with the majority of interests and efforts represented within the RDA, such as “Geospatial”, “Big Data Analytics”. “Domain Repositories”, “Libraries for Research Data”, ”Publishing Data” or ”Preservation e-infrastructure” IGs (to date, those have explicitly stated their interest and support).

This session on themes of “Data Rescue” will (a) remind the community of the immense scientific benefits of accessing heritage data which are presently inaccessible (they may have analogue formats, are not reachable on-line, or have inadequate meta-data), (b) discuss what has caused Data Rescue to be so broadly unsupported in a techno-savvy world, and (c) collect ideas for consolidating individual efforts into a force that can no longer be ignored. It will attempt to achieve answers to (b) and (c) through a panel discussion to which the audience is also invited to contribute. A panel of 8 will include representatives of the various stages of the life-cycle of rescued data, but will primarily be selected from those with practical experience in discovering, recovering and rescuing heritage data, and on the new science which has thereby (and only thereby) been made possible. Likely topics (each given 10 minutes) are hand-written Polar or biodiversity data, historic ozone measurements from stellar spectra, epidemiology studies from historic medical records, and solar-system explorations in detail from abandoned (unread) tapes. but will depend on who will attend the Paris meeting. Other speakers may review

briefly the management and dissemination of those data in tandem with born-electronic ones, and discuss potential difficulties.

The desired outcome of the meeting will be a mandate to pursue the theme of Data Rescue on its broadest front by contacting and bringing together as many individual rescue projects as are known, to establish a regular forum for the activity, and to commit to creating the Guidelines mentioned above. In so doing, the session will make substantial strides towards setting both the Guidelines and the Book rolling.

Contact Person: Elizabeth Griffin

Co-chairs: David Gallaher; Lesley Wyborn, Suchith Anand, Peter Baumann, Luciene Delazari, Andrea Perego, Chris Pettit, Morris Riedel, Peter Baumann, Kwo-Sen Kuo, George Alter, Ruth E. Duerr, Robert J. Hanisch, Peter Doorn, Wolfram Horstmann, Kathleen Shearer, Michael Witt