
Scientific “Agendum” of Data Science

Shuichi Iwata

The University of Tokyo, Kashiwa, Japan
iwata@k.u-tokyo.ac.jp

Data on nuclear cross section have integrated metals, ceramics, waters and other materials to form various nuclear reactors, which is a basic for nuclear engineering. Data on intrinsic properties of substances are used to link extrinsic properties of substances and structure-sensitive engineering properties of defects as fundamental constants have been defined in a network of elaborate experiments and models in science. These understandings have driven me to data activities since 1970 when I at first became to know CODATA by name, and it is a good time now for me review the committed works in these 37 years and write down my personal gendumh reflecting the re-view so as to have more collaborators for the next milestone.

As scientific agendum of CODATA, we have added an aspect gdata and society explicitly through such opportunities of WSIS(World Summit on the Information Society)s and lessons on tragedies, namely, suffering from natural disasters and epidemics. Not only by global warming problems pointed out clearly by IPCC (Intergovernmental Panel on Climate Change) but also by other global issues as known well partly by UN MDGs(the eight Millennium Development Goals), we have become to know the necessity of linking scientific, technical, economic, social and political agenda with proper missions and guidelines for the society. Here human-centered reorganizations of domain-differentiated sciences from natural sciences to social sciences, are requested to be carried out, where and when we need common data for the proper and holistic reorganization to reach right decisions and consensus of society. If we do not share common data on global warming effects, we cannot write down our remedies against inconvenient truths and establish flexible and steady roadmaps for the sustainable society. It is necessary for us experts to write down proper remedies together by linking such associated scientific domains as politics, laws, ethics, economics, environmental sciences, ecologies, civil engineering, manufacturing, waste management and so on and complimenting missing links there for better solutions. And as a consequence of such efforts we may come to create a new scientific domain gdata-driven sustainable scienceh to design and manage the society properly. It is really requiring us continuous big efforts with challenging spirits to start everything from facts and data, which is our raison dfetre and concerns our identity as data scientists.

However, the above agendum essentially has already been associated to our core activities. Data and knowledge corresponding missing links of domain specific sciences have been daily works of CODATA. For example,

- fundamental constants have been compiled reflecting advances of precise measurements and basic sciences, and also higher coherences of scientific models, which has resulted in creating new sciences like nano-sciences, spintronics and other specialized scientific domains, and also several key standards for the information society, namely, radio wave standards, current, voltage and so on,
- spectra data and diffraction data have guided us to get microstructural information on substances and materials, and also on life. Together with models and interatomic potentials derived from first principles calculation, new scientific and technical fields have been exploited rapidly, and application areas are spreading widely over drug design, defect theory, fracture dynamics, materials design, process design, earth sciences, and even in bioscience to see the origin of life and medical diagnostics.
- many exemplars in biosciences to get insight through using common data: arabidopsis thaliana data for evolving botany; RNA data to link DNA and protein, and consequently associated with disease and health; data on Tradescantia ohiensis and/or nude mice for irradiation effects, and by taking advantage of recent advances in genomics, proteomics biophysics and biochemistry, an breakthrough is going on to overcome limitations of epidemiological survey based on statistics.

Data-oriented statistical approaches are combined with scientific models and practical monitoring, and traditional established safety/risk/reliability standards are changing into proactive and dynamic adaptive standards. Safety/risk issues in medical services, nuclear reactors, aircrafts, company managements, energy resource security and so on can be dealt with in a similar way. Openness and transparency of many disciplines and scientific domains promoted by e-science projects and so called global information commons are prerequisites for a revolution of sciences by the 7 billion people. Devices for the revolution might be something creative of emerging wisdoms and welling up emotions in the internet, which may be more than such knowledge management approaches as ontology, metadata, object-oriented approach, semantic-web, common sense reasoning and so on.

Through evaluation of fundamental constants we are integrating quantum worlds, atomistic worlds and macroscopic worlds quantitatively, where scientific disciplines and domains are networked with a certain consistency. Through preparation of geometric data of parts with properties we can design an artifacts and assemble the available parts into the artifacts of integrity and cost-effective performances, where domain differentiated engineering disciplines are integrated to establish manufacturing industries. Design and maintenance of landscapes, cities, countries, regional environments and global climates are used to be carried out in a similar way, namely, sharing data by stakeholders and coordinating different views and opinions by the shared common data. Quality of data matters every time and everywhere. Gold in, gold out. Reasonable estimations of uncertainties on data may produce better results

and outcomes. The more the problem to be solved is uncertain, the more we should become flexible. Evidence-based deterministic approaches do not work effectively, and adaptive and heuristic approaches work better coupled with in situ data capture, evaluation, and quick decision and timely actions. Holistic creativity as a group is a key for a success of the group, where practical maintenance of data quality for proper decision is important. Time constants of data life cycle are becoming shorter, and diversities of stakeholders and complexities of data are increasing. New disciplines are to be continuously created by taking advantage of available data and devices so as to prepare solutions on time. Without proper managements of continuously produced big data, and without productivity of new disciplines based on data, we cannot solve important problems of the world. Data science may play an important role there.

Shuichi Iwata

Communications and Discoveries from Multidisciplinary
Data

Iwata, S.; Ohsawa, Y.; Tsumoto, S.; Zhong, N.; Shi, Y.;
Magnani, L. (Eds.)

2008, XIV, 340 p. 121 illus., Hardcover

ISBN: 978-3-540-78732-7