"Big-data" and "Open-data": Two key words to support future science

*Ken T. Murata[1]

1.National Institute of Information and Communications Technology

A pioneering computer scientist Jim Gray proosed "the Fourth Paradigm: Data-Intensive Scientific Discovery" in 2008. Increasingly, scientific breakthroughs will be powered by advanced computing capabilities that help researchers manipulate and explore massive datasets. The speed at which any given scientific discipline advances will depend on how well its researchers collaborate with one another, and with technologists, in areas of eScience such as databases, workflow management, visualization, and cloud computing technologies.

The data intensive science, which is often called "data driven science", is a methodology to start science at data, not to data; Data is a starting point. This idea implies that all of the necessary information are included in the data. In these days, this kind of concepts have been presented and discussed in books, papers, web and other media. This means that the "Opinion for All" is correct unexpectedly as James Surowiecki mentions. This is a concept recently denoted as "collective intelligence".

To the author, "big data", "data oriented science", "data driven science", "collective intelligence" and "information commons" in science seem to be directed to a certain objective. What is required to order to move in this direction is "open data"; all of the scientific data should be open and accessible to any researchers. Data scientists who have any technologies for big data would be motivated to analyze these open data, even if they are not specialist of the domain fields. Nobody knows what will be derived, or nothing new will be exracted. To examine this, anyway, data must be open.