Digital Typhoon: A Data-Centric Approach to Events on the Earth

Asanobu KITAMOTO
National Institute of Informatics
Tokyo, Japan

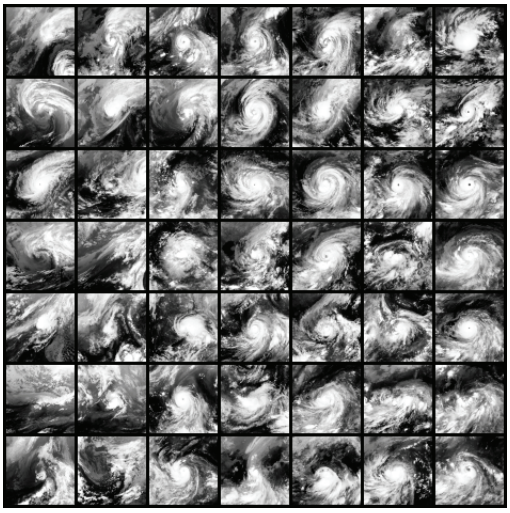

Figure 1 Typhoon image collection.

Typhoons are one of the most significant atmospheric events on the earth. They give large impacts to human society, hence early warning, accurate forecast, and the sharing of information, among others, is regarded as important research challenges for mitigating typhoon impacts. Although these challenges have been tackled by many researchers, our aim is to give a new perspective on these issues from informatics point of view. We mainly focus on two issues. Firstly, we combine real-time data with past data in the databases so that users can compare the present typhoon with past typhoons. Secondly, we expand data sources from authorities to grassroots so that local and diverse information can be aggregated and shared among people in many places.

A concept behind these approaches is “data-centric” science, in comparison to “computation-centric” science, which is typical in the current researches of meteorology. Data-centric science is claimed to be the fourth generation in the history of science, namely starting from empirical science, theoretical science, computational science, to data-centric science. In the context of meteorology, it begins with the observation of the sky, and theories of the atmosphere were established later. After the invention of computers, research shifted to computational meteorology, where the simulation of the atmosphere became the most important tool of research.

Now, a data-centric meteorology, which we call “meteoinformatics,” started to play an important role, because the value of research is emerging in the analysis of huge amount of data produced from observations and simulations. These analyses should be done through the integration of many kinds of data, such as observation data, reanalysis data, simulation data, social data, and many other data. Relationships between data can be defined theoretically, statistically or logically, and similarity or significance of data is a useful measure in some cases to characterize the important aspects of data.

Digital Typhoon (http://www.digital-typhoon.org/) is our project on typhoons based on a data-centric approach [1]. It is a database of typhoons, but it is significant in the amount of data archived and the heterogeneity of data integrated. Table 1 summarizes the amount and variety of archived data, while Figure 1 shows the typhoon image collection in Table 1. The table shows that the data ranges from image data to text data, and to numerical data. *Digital Typhoon* archives not only objective data such as simulation data, but also subjective data such as participatory reports. It

integrates those heterogeneous data and computes relationships between many kinds of data. The advantage of dealing with typhoons is that they can be identified as events with unique names (in fact, giving unique names is operationally done by Japan Meteorological Agency), which are used throughout the system. We then formalize basic searching operations and allow users to combine them sequentially, such as visualizing tracks of typhoons whose news articles contain the word X. Other unique searching operations focus on similarity between data, such as searching past typhoons that have similar cloud patterns with the present typhoon. These data-centric operations are useful when rigorous theories are not known and/or people are familiar with analogy-based inferences.

In a data-centric approach, the visualization of data is also an important challenge. *Typhoon Front* (http://front.eye.tc/) is an example of information visualization designed for participatory media. The purpose of this interface is to visualize user generated content on local typhoon information as a time-series animation so that people can easily see when, where and what local people are saying about typhoons. Because many events on the earth are essentially local, people on the spot can tell the situation in most detail. Participatory media is often blamed for its unreliability of information, but we believe that this will play an important role in the detection of changes or events on the earth.

Our research achievements have been developed and opened on the Web with advanced searching and visualization functionality introduced above. Our websites have been attracting many people since the opening in summer 2003 -- the total number of page views exceeded 65 million, with 850 thousand page views in a day at maximum (this is when a typhoon made landfall in Japan). Our future work includes the integration of databases with simulators to construct an integrated prediction system that can learn from the past and the present events.

Table 1. The amount and heterogeneity of data in *Digital Typhoon* (as of June 2008).

Data Type	Year Since	Number of Data (Records, Scenes)
Meteorological satellite imagery	1978-	About 200,000 scenes
Typhoon best track data	1951- (1906-)	1520 typhoons (955 Australian cyclones)
Typhoon image collection	1978-	About 150,000 images
AMeDAS observations	1976-	About 350,000,000 records
News articles (text data)	2003-	About 10,000 articles
Participatory reports (text + image + movie data)	2004-	About 1,000 reports
Ground-based camera images	2004-	About 2,000,000 scenes
Grid Point Value (GPV) data	2002-	About 5,000 scenes (GSM) + 11,000 scenes (MSM)

[1] Asanobu KITAMOTO, "Digital Typhoon: Near Real-Time Aggregation, Recombination and Delivery of Typhoon-Related Information", Proceedings of the 4th International Symposium on Digital Earth, (CD-ROM), 2005.