Visions of Data Technology and e-Statistics with their roles in Industry and Government

Sung H. Park*

(目 次)

Abstract

The term Data technology(DT) is defined and its differences with respect to IT(Information Technology) are discussed. Major characteristics of DT are explained along with its role in industrial development. Finally, the concept of *e-Statistics* is proposed and illustrated in relationship with DT.

Ⅰ. Introduction

It is well known that modern technology for the 21st century is regarded as comprising the 6Ts: IT(information technology), BT(bio-technology), NT(nano-technology), ET(environment technology), ST(space technology), and CT(culture technology). We believe that one more T should be added to these 6Ts, which is DT, data technology.

DT(data technology) may be defined as follows. DT is a scientific methodology which deals with

* Department of Statistics, College of Natural Sciences, Seoul National University.

systematic collection, storage and retrieval of required data,
refinement and object-oriented analysis of data,
conversion of data into information,
statistical and computational modeling based on data,
formation of inferences and relevant knowledge,
diagnosis of the present states and prediction of future events by use of statistical and probabilistic modeling techniques.

We believe that DT is an essential element for organizational management, and in general for global competitiveness. The importance of DT will grow rapidly in this knowledge-based information society. The term DT was first introduced by Park(2001), and later elaborated by Park(2001) and Park and Suh(2002) in a different angle.

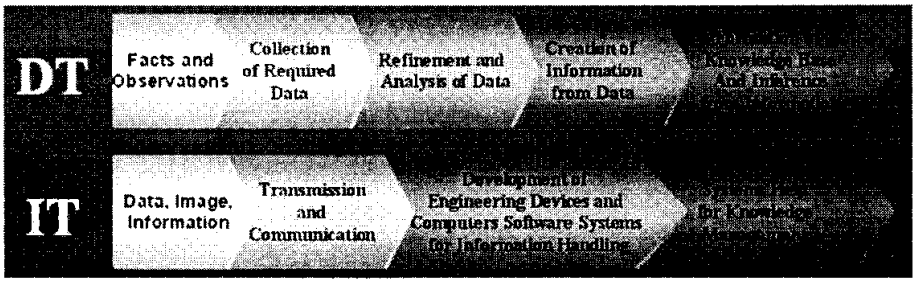
Ⅱ. Differences between IT and DT

Many people believe that DT is a subset of IT. We do not think that this is true. IT may be defined as follows, which is somewhat different from DT. IT is an engineering methodology which deals with

presentation and control of information and knowledge generated by DT,
efficient information transmission and communication,
manufacturing of electronic systems/devices for information transmission and communication,
manufacturing of computer-based networking instruments,
systems management dealing with data, information and networks.

The differences between DT and IT can be seen in the information flow diagram as shown in 〈Figure 1〉.

〈Figure 1〉 Information flow of DT and IT



DT is mainly concerned with data collection, statistical analysis of data, generation of information, and creation of necessary knowledge from information. However, IT is mainly concerned with data/information/image transmission and communication, and development of engineering devices and computers for information handling. Also IT is concerned with engineering tools for knowledge management. Generally speaking, DT is the infra-structure of IT. Without DT, the growth of IT is limited. DT is software-oriented, but IT is hardware-oriented and systems-oriented. Without IT, DT cannot be well visualized. IT is the vehicle for DT development.

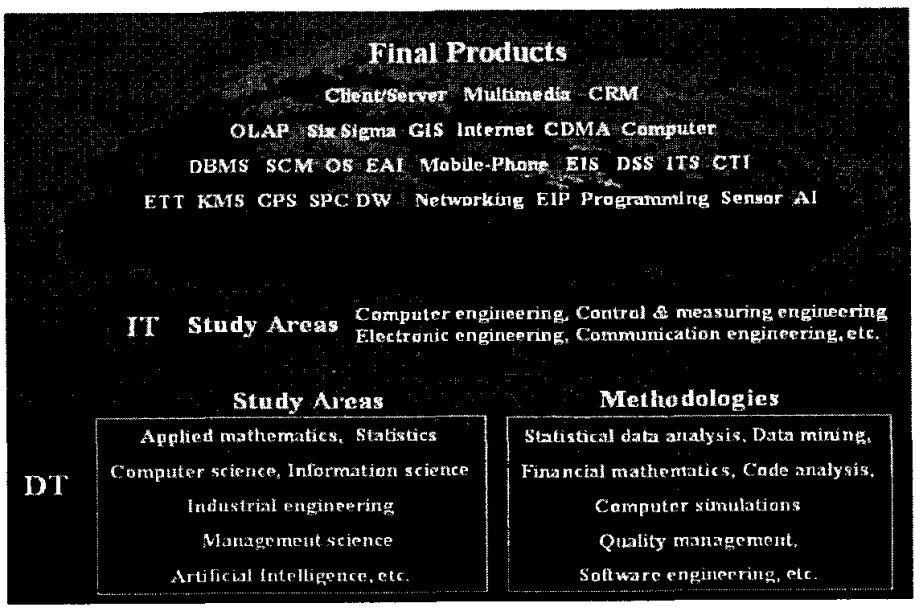
〈Table 1〉 shows the differences between DT and IT in terms of characteristics, major products, and major fields of study.

〈Table 1〉 Comparison of DT and IT

Contents	DT	IT
Major characteristics	Software-oriented, Scientific approach for data analysis modeling and future prediction	Systems-oriented, Engineering approach for transmission & communication information and knowledge
Major Products	Softwares such as DBMS, CRM, SPC, ERP, Statistics, Data-mining, Simulation, Cryptography	Electronic devices and auxiliary devices for communication, measurement and control
Major fields of study	Mathematics, Statistics, Information science, Computer science, Decision sciences	Computer engineering, Electronic engineering, Communication engineering, Control & systems engineering

To explain the differences between DT and IT clearly, we have sketched 〈Figure 2〉. Notice that DT is the root of a big tree, IT is the trunk of the tree, and there are many visible fruits as the final products of DT and IT. Without DT, IT cannot survive. Without DT and IT, we cannot harvest the final products.

〈Figure 2〉 The relationship and difference between DT and IT

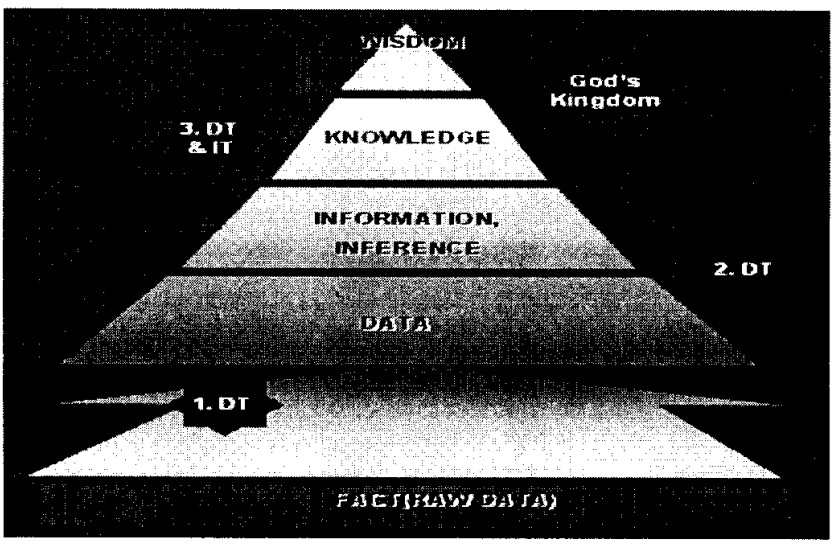


Ⅲ. Knowledge triangle

It is said that the 21st century is the knowledge-based information society. We can think about the knowledge triangle as shown in 〈Figure 3〉 in which DT and IT play important roles.

In each step, the following activities are usually implemented.

〈Figure 3〉 Knowledge triangle



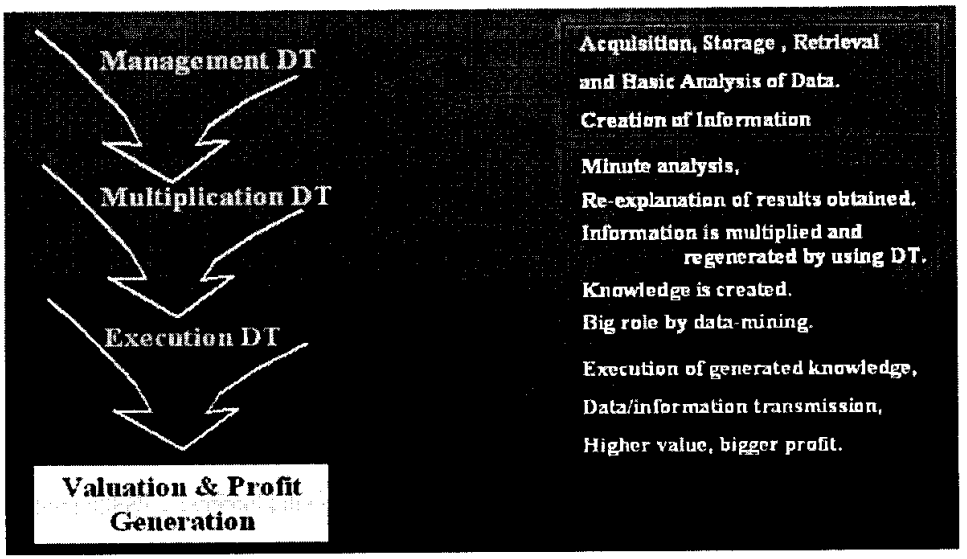
〈Table 2〉 Major activities in each step of knowledge triangle

Step	Basic roles
Step 1	Measurement, Data refinement, Sampling design, Design of experiments, Meta data management, Gage R&R test
Step 2	Data analysis and statistical modeling, Data mining, Application scope refinement, Diagnosis and control, Prediction modeling
Step 3	Knowledge generation, Decision support Review, Valuation, Enhanced modeling
Step 4	Wisdom generation from information clustering

Ⅳ. Scope of DT

The scope of DT can be divided into three categories: management, multiplication, and execution. Management DT comes first, and then multiplication DT comes next, and finally execution DT provides valuation and profit generation for the organization concerned. The sequential scope can be shown as in 〈Figure 4〉.

〈Figure 4〉 Scope of DT



V. Loss due to insufficient DT

A weak DT can result in big losses to a company, to a society, and to a nation. Some examples of national loss due to insufficient DT are as follows.

5.1 Economic crisis in 1997:

Korea faced an economic crisis in 1997, and the International Monetary Fund assisted Korea at that time. The major reason was that important economic data, so-called Foreign Exchange Stock(FES), had not been well managed. If collection of FES, trend analysis of FES, and prediction of FES had been well conducted, there would have been no economic crisis.

5.2 Inherent political dispute in politics:

Politics is perhaps the most underdeveloped area in Asia including Korea. Unproductive political disputes hamper development of all other areas such as industry, education, culture, etc. If popular opinion surveys are properly

conducted, and political parties appropriately consider the opinion of the majority of people, politics can become more mature, and can assist the development of all other areas.

5.3 Cost of poor quality:

The cost of poor quality(COPQ) is the total cost incurred by poor quality and poor management. It is believed that the COPQ of most companies in the world is about 20-40% of total sales value(see Juran, 1988). Conway(1992) even claims that in most organizations 40% of the total effort, both human and mechanical, is wasted, and the waste can be eliminated or significantly reduced. The COPQ can be divided into visible COPQs(so-called quality cost) and hidden COPQs. The visible COPQs consist of prevention cost, appraisal cost, and failure cost. The hidden COPQs are the costs which are not directly captured on the balance sheet, such as lost opportunity cost, lost goodwill cost, and frequent design change cost.

If DT is well utilized for the data analysis of COPQs, the COPQs can be reduced to about 10-20% of total sales value.

Ⅵ. e-Statistics and DT

Recently, the field of e-Statistics has emerged, and it can be defined as follows. e-Statistics is a scientific methodology which deals with

electronic data collection, storage, and retrieval,
electronic data refinement and statistical analysis,
electronic statistical model building,
electronic simulation and statistical prediction,
automated generation of knowledge base and inference.

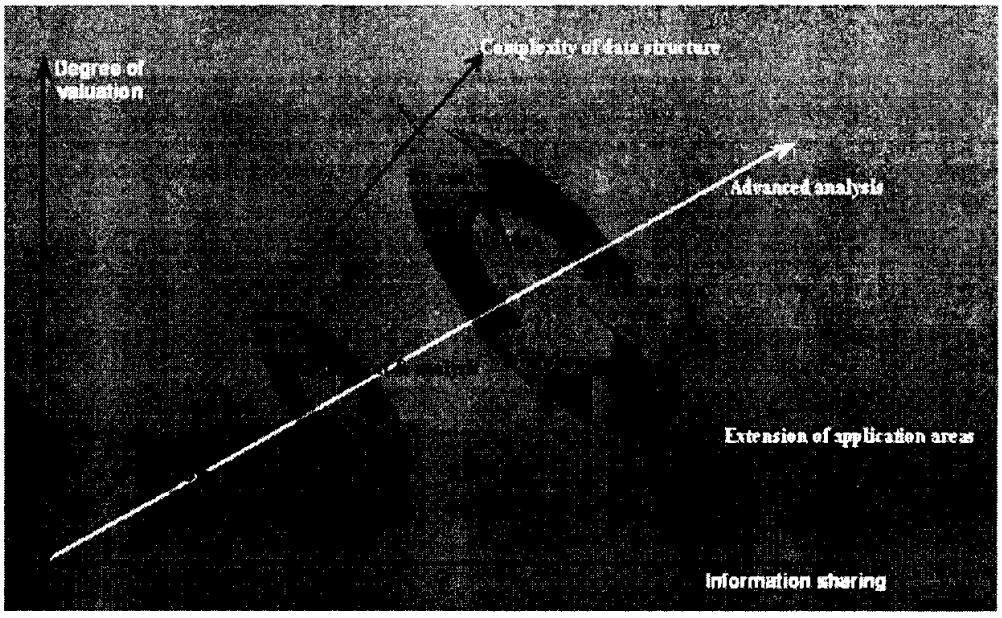
From this definition, we can judge that e-Statistics is a subset of DT. However, e-Statistics is the most important ingredient of DT. With sound

e-Statistics, DT and IT can realize their full potential.

We believe that e-Statistics plays a most important role in the software industry. This 21st century is the so-called knowledge-based information society. Strong e-Statistics and DT will pave the way for strong IT and eventually strong global competitiveness.

〈Figure 5〉 shows the valuation flow in e-Statistics and DT. The major X-Y axes are the information sharing and the degree of valuation. There are also three more dimensions: Z1=extension of application area, Z2=advanced analysis, and Z3=complexity of data structure. As Z1, Z2, and Z3 become larger, the information sharing and the degree of valuation also become larger. Along the Z1, Z2, and Z3 axes, three steps of valuation are progressed. The first step is data storage, the second step is data refinement and analysis, and the third step is data valuation. In the third step, a continuous valuation flow(re-analysis, information, and application) constitutes a cycle.

〈Figure 5〉 Valuation flow in e-Statistics and DT



Ⅶ. Visions of e-Statistics and DT

The vision of e-Statistics and DT can be summarized as follows.

7.1 Revolution in business decisions

- systems approach for all decision support functions,
- appraisal of projects and business strategies by automated analyses from objective data and models,
- optimization and objective decision criteria,
- formation of global trade and marketing strategies.

7.2 Reform in government affairs

- systems approach for all government policies,
- reduction of arbitrary policy changes,
- mandated feasibility requirements for all major legislation and major budget requests,

7.3 Enhanced R&D efficiency for all sectors

- effective statistical modeling for verification of research hypotheses,
- design and analysis of scientific and industrial experiments,
- automated or semi-automated formulation of inferences,

7.4 Promotion of information-rich and scientific lifestyles

- optimal decisions for individuals by *public domain programs and shells* (financial investment, tax strategies, activities planning and other personalized decision support tools),
- social reforms through scientific lifestyles.

Finally, we would like to make the comments that DT and e-Statistics are exciting new tools and new paradigms for industry, government, society, and individuals in the information age.

REFERENCES

1. Conway, William E. *The Quality Secret: The Right Way to Manage*, Conway Quality, Nashua, N.H., 1992.
2. Juran, J.M. *Juran on Planning for Quality*, Free Press, New York, N.Y., 1988.
3. Park, Sung H. Paradigm shift of statistics and development of data technology in knowledge based society, Journal of Information and Operations Management, Vol. 11, page 53-59, 2002.
4. Park, Sung H. Economics of Data Technology, An article appeared in the *Korea Economic Daily*, December 3, 2001.
5. Park, Sung H. and Suh, Moon W. Data Technology and e-Statistics with their applications in industry, Proceedings of the 4th Asian Regional Section of the International Association of Statistical Computing, pp. 201-204, Dec. 5-7, 2002, Busan, Korea.