

## **Data Efficiency: A White Paper on a Distinction of Data**

**Data Efficiency** refers to efficiency in the many *processes* that can be applied to data such as storage, access, filtering, sharing, etc., and whether or not the processes lead to the desired outcome within resource constraints.

A management definition of Data Efficiency would be the measure of how data storage and usage across an enterprise or within a department or within a project impacts the organization's costs and revenues.

On the broadest level, for management purposes:

$$\text{DE} = \frac{\text{Expected benefits from applying IT to a given task}}{\text{Cost of application of IT}}$$

On the technical side, in the development of hardware, software or systems, Data Efficiency can refer to many things such as:

- packing bits on a limited physical medium (a DVD)<sup>1</sup>
- chip area usage on a silicon wafer<sup>2</sup>
- the use of data in programming so as to take less time and computation resources<sup>3</sup>
- the time and resources required to collect or input raw data<sup>4</sup>.

The first two above depend on the physics of the actual media, and for the third example, some code is more efficient than other code because of the way that data sets are captured and accessed and used within the code. For the last example, the standard practices of an organization are either manual or automated, for the aggregation and transcription of data.<sup>4</sup>

### **Data Attributes**

Besides the efficiency of processes that act on data, data by itself can have some attributes that can affect the efficiency of the processes used to arrive at the desired outcome. These attributes are:

Availability  
Reliability  
Usability.

If Data is available then it has been defined, observed, collected and recorded in such a way that it can be accessed.

If Data is reliable then it consistently represents accurately the events and conditions that it was intended to.

If Data is usable then it is suitable for the various processes that can be brought together to create the needed information to reach the desired outcome.

If Data is available, reliable, and useable then processes can be created to turn the data into information and the information into desired results.

Data processes cannot be efficient in relation to the desired results, unless the data is available, reliable and useable.

## Process

Data efficiency encompasses terms such as *Data Storage Efficiency*, *Data Retrieval Efficiency*, *Data Management Efficiency*...etc.

Data processing is simply defined as the process that converts data into information or knowledge<sup>5</sup>. In today's world, data processing is linked to the handling of information specifically by computers in accordance with strictly defined systems of procedure. The efficiency of these systems and how effectively they pass information from one system to another for further processing can be broken down into several components, the sum of which is the total efficiency of each system.

The components of each process are data storage, data retrieval, and data management. Each individual system handles data in this manner in one form or another and the measurement of the efficiency of these processes can be defined as Data Efficiency.

## Optimization

Desired results are achieved with high efficiency depending on data attributes and process optimization. These are the two sides of Data Efficiency. For example, in the business world, BI (Business Intelligence) is the practice of gathering and using information about projects, operations, companies or industries and employing a broad range of applications and technologies for analyzing the data to help make business decisions.<sup>6</sup>

As the lower-level technical components of storing, retrieving and managing the data throughout the entire BI process are optimized, the entire managed BI process gradually becomes more efficient.

## Metric

"Data Efficiency" (DE) can be quantified. This metric is calculated to be the ratio of the value to the user (in dollars, for example) of the *output information* ... as compared to the process cost.

$$DE = \frac{\text{Value of Output Information (\$)}}{\text{Process Cost (\$)}}$$

*For example:* Any given I.T. application might consist of 7 processes. These processes in their entirety comprise a system. The system can be thought of as inside a "black box" with only one input and only one output interface. The 7 processes are, in order:

- (1) Input
- (2) Storage
- (3) Retrieval
- (4) Filtering
- (5) Applying logical operations
- (6) Tabulating results
- (7) Transmitting the output.

The entire process might cost \$5 million to complete, but the results are estimated to be worth \$7 million to the client. DE is calculated to be 7/5 or 140%.

For an inefficient system, DE would measure less than 100%. The process of generating the output would cost more than the value of the output itself.

## IRR

When evaluating costs and benefits of IT applications across an enterprise, the rules of corporate finance and Internal Rates of Return (IRR) commonly apply.<sup>7</sup> “Data Efficiency” can be thought of as a form of IRR calculation. If the Data Efficiency of a project is at 140%, and if the IRR hurdle rate is 120%, then the project’s expected return is 20% over the IRR hurdle rate, and the project should be undertaken.

$$\text{DE} = \frac{\text{Expected Benefits ... \$ 7 M}}{\text{Expected Costs... \$ 5 M}} = 140\%$$

IRR hurdle rate is 120%, DE is 140%, so, project should be undertaken.

If a manual information processing operation is shown to have a low Data Efficiency, what this is saying is that ongoing manual information processing costs are higher than the additional revenues they are expected to bring in. So in this case, the IRR falls below the hurdle rate, and the process should be discontinued or upgraded.

As an example:

$$\text{DE} = \frac{\text{Expected Benefits (additional revenues)... \$ 1 M}}{\text{Expected Costs... \$ 5 M}} = 20\%$$

If the IRR hurdle rate is 120%, DE is only 20%, so the manual process should be upgraded.

Care should be taken on defining parameters in this analysis. What exactly is being measured and analyzed? Are the comparisons being made with equivalent data?

## Summary

Data Efficiency applies to the technical side of hardware, software, computer networking and data communications, but it also applies to managing information technology inside an organization with defined goals. It can be used for evaluating information technologies and their effectiveness across an enterprise, within a department, or as part of a project.

Data Efficiency is a valuable construct. It actually forces a mindset or awareness in discussion and decision making. Using this term implies that the same rigorous principles from computer science that underlie the creation of hardware, software and systems can also be carried into management decision-making processes. This is especially valuable for new applications of I.T. such as improving older processes, or establishing leadership in a mature industry or opening up new markets.

End

- 
- <sup>1</sup> Takashi Ishida, "High density phase-change type optical disk having a data efficiency of more than 80 percent" US Patent for Matsushita Electric Industrial Co. Ltd.. Issued [October 12, 2004](#).  
[<http://birdnest.com/index.php?id=20>]
  - <sup>2</sup> Frederick A. Ware and Craig Hampel. "Improving Power and Data Efficiency with Threaded Memory Modules" ICCD Proceedings 2006 Oct 4, 2006 Austin, TX.  
[[http://www.iccdconference.org/proceedings/2006/paper\\_133.pdf](http://www.iccdconference.org/proceedings/2006/paper_133.pdf)]
  - <sup>3</sup> Bob Steigerwald, Rajshree Chabukswar, Karthik Krishnan, Jun De Vega: "Creating Energy-Efficient Software" Intel's online journal: *Software Community* Published On: Thursday, September 13, 2007.  
[<http://softwarecommunity.intel.com/articles/eng/1462.htm#dataefficiency>.]
  - <sup>4</sup> This is referring to BirdNest Services' main market focus, the water/wastewater management industry.
  - <sup>5</sup> Wikipedia definition of Data Processing [[http://en.wikipedia.org/wiki/Data\\_processing](http://en.wikipedia.org/wiki/Data_processing)]
  - <sup>6</sup> Webster's New Millennium Dictionary of English  
[<http://dictionary.reference.com/browse/Business%20intelligence>]
  - <sup>7</sup> NetMBA Business Knowledge Center NetMBA.com. Website operated by Internet Center for Management and Business Administration [<http://www.netmba.com/finance/capital/budgeting/>]