



**RIVERS STATE UNIVERSITY**  
**P.M.B 5080, NKPOLU-OROWORUKWO,**  
**PORT-HARCOURT**

**D**ATA PROCESSING  
&  
COMPUTER ORGANIZATION

- **DEPARTMENT:** MECHANICAL ENGINEERING
- **LEVEL:** 200
- **COURSE:** COMPUTING AND SOFTWARE ENGINEERING
- **LECTURER:** DR. MISSION FRANKLIN
- **DATE:** 10<sup>th</sup> FEBRUARY, 2025
- **GROUP:** G

**CONTRIBUTORS**

• MICHAEL SAVIOUR	<i>DE.2023-3892</i>	• MARVELLOUS IHEMJIRIKA CHINEMEREM	<i>DE.2023-2878</i>
• MICHAEL IHENAYI ALOZIE		• MONDAY EXCELLENT SOMTOCHUKWU	<i>DE.2023-3798</i>
<i>DE.2023-3725</i>		• NJOBUANWU GODFREY ORONDA	<i>DE.2023-3866</i>
• MBAMA ADVISER	<i>DE.2023-3795</i>	• NNENANYA EMMANUEL UGOCHUKWU	<i>DE.2023-3701</i>
• MARTINS BETTY	<i>DE.2023-3874</i>	• NWANEZI EMMANUEL CHUKWUKA	<i>DE.2023-3746</i>
• MARGARET KPALAP	<i>DE.2023-3877</i>	• NWAOBUM JOSEPH CHINAZAEKPERE	<i>DE.2023-3846</i>
• MAXWELL-LEH NUALE	<i>DE.2023-3837</i>	• NDUBUISI KELVIN ONYEDIKACHI	<i>DE.2023-3841</i>
• NOBLE EMMANUEL	<i>DE.2023-3833</i>	• NLERUM PROSPER PRAYER	<i>DE.2023-3755</i>
• MICHELL MANJO	<i>DE.2023-3710</i>	• MICHAEL PRECIOUS IZUCHUKWU	<i>DE.2023-3879</i>

## **AIM**

This paper describes the processes involved in data processing, highlighting relevant tools used. Additionally, it explains how computer organization can enhance effective data processing.

## **RELEVANT SECTIONS**

- **DATA PROCESSING**
- **STAGES OF PROCESSING DATA**
  - **DATA COLLECTION AND VALIDATION**
  - **DATA RECORDING**
  - **SORTING**
  - **CLASSIFICATION**
  - **STORAGE**
  - **ANALYSIS**
  - **REPORTING**
- **DATA PROCESSING CYCLE**
- **TOOLS FOR PROCESSING DATA**
- **COMPUTER ORGANIZATION**
- **THE RELATIONSHIP BETWEEN DATA PROCESSING AND COMPUTER ORGANISATION**

## INTRODUCTION

Data is a property of facts and statistics collated in a measure in which can be extracted, modified, analyzed and interpreted. It quantifies and qualifies physical and abstract properties, acquired through observations and experiments, into distinct and predictable representations. It can be visualized as symbols, figures, numbers, alphabets, graphs or recordings of the interactive characteristics of the physical world, such as sound, light, etc.

In order to fully understand the purpose of processing and organizing data, we need a model that can describe the manipulation of data at a low level. An example of such model is the computer. In the computer, series of 1s and 0s form the fundamentals of its data structure. These are used to construct instructions that could be executed to guarantee the working of certain operations. An overview of these processes begins with the supply of alternating electronic and magnetic signals from specific designed circuits. These circuits can recognize and directly execute a limited set of instructions into which all of its programs must be converted to before they can be executed. These instructions are rarely much complicated than adding numbers, asserting the state of a number (1 or 0), as well as copying a piece of data from part of a computer's memory to another. Imperatively, they have to assume a specific order, pre-defined method of access and organization. Specifically, digital computers rely on proper and efficient processing of data. At its lowest level, the digital logic level, it manipulates objects commonly known as gates. Each gates has one or more digital inputs (signals) and compute as output some simple function of these inputs, such as AND or OR. A small number of gates can be combined to form a 1-bit memory. *A bit is the smallest unit of data operatable by the computer.* The 1-bit can be combined in groups, for example 16, 32, or 64 to form registers. Each registers can hold a single binary to some maximum. Gates can be combined to form the main computing engine itself. At the micro-architectural level, there are several collections of registers that form a local

memory as well as a circuit called the arithmetic logic unit (ALU). These registers are connected to the ALU to form a data path, over which data flow. The basic operation of the data path consist of selecting one or more registers, having the ALU, operate on them and storing the result back in some registers.

Other processes including hardware and software intrinsic are subsequently followed to ensure a proper, systematic functionality of the computer. Further description of these processes would require a comprehensive explanation which may rather introduce more complexity to this paper. However, what should be noted is that the design and implementation of the computer is done in a highly processed and organized manner. This places a foundation on how data should be processed within the computer.

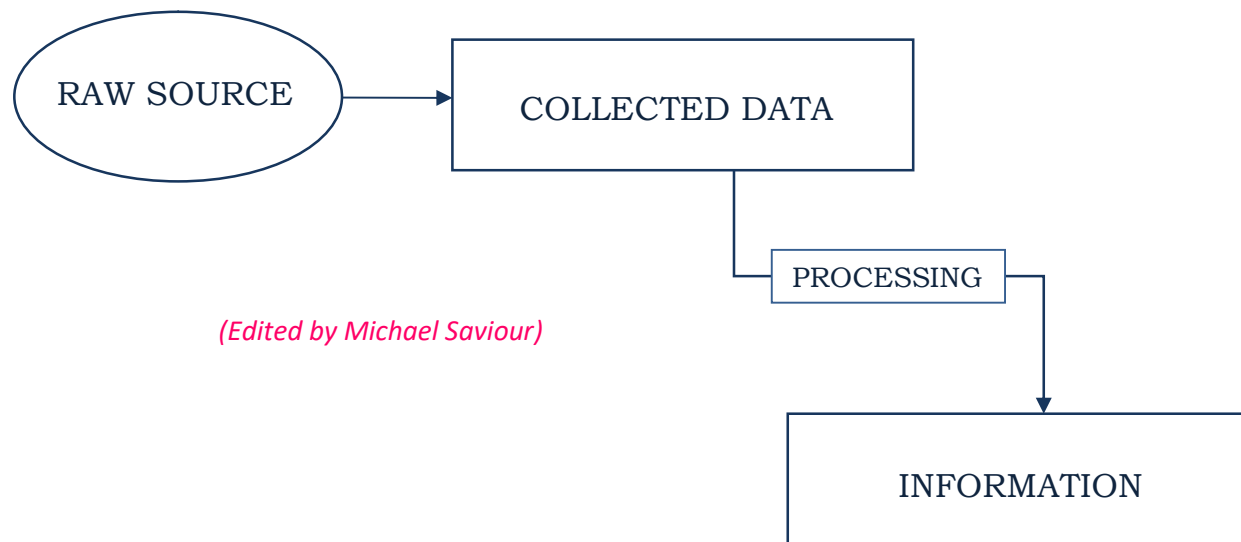
Besides the computer, there is plethora of areas which depends on efficient processing of data in other to establish effective works in some of its activities. For example, in the medical field, researches are done to identify a common disease and to conclude the symptoms that may be observed when an individual is contracted with such disease. The results and conclusions from these researches are presented as collection of data, which is further processed through *heavy* analysis, testing and reproduction in other to ascertain possible biological causes, and proceed with a procedural diagnostics and treatment for the disease.

Essentially, data is considered as the inner layer of communication, therefore the efficiency in its processing is a requirement in its *area* of concern (Edited by Michael Saviour)

## 1.0

## DATA PROCESSING

As stated earlier in our introduction, data is a distinct piece of information that can be modified to express diverse physical or abstract properties. It can be measured, recorded, reported as well as visualized in several ways. It is important to understand that data is a raw fact, meaning it is neither edited nor structured. In this state, it lacks cohesion and may contain a lot of garbage and errors. This unstructured fact has to be collected, extracted, corrected and modified in other to be useful. *Data processing* is the activity of converting these disorganized, error-prone facts into a structured and useful form that is easy to read and visualize. This form is commonly referred to as *information*. It is commonly performed in stages, and until a complete processing, the data in each stage may still be considered raw.



Processed data is the external layer of communication. Its use, ensures data-backed decision-making and provide unique insights that can benefit organizations in many ways, for example, business intelligence analysts can extract useful and accurate information about the condition of their business from raw data such as audience interest, sales figures, marketing campaign performance, and overall productivity; expert researchers or scholars could conclude or determine the genuity or possible outcomes of their study cases through the acquisition of related information on such subject.

## **1.1**

## **STAGES OF PROCESSING DATA**

Until lately, we have postponed the discussion of how data can be transformed from an ill-organized form to a structured and filtered form. In this section, we shall consider the various stages/activities that raw data undergoes in order to be considered processed. Because of the unstructured nature of Data, a wide range of processing approaches could be applied. We shall however, limit the scope of this paper to the basic activities that will guarantee an effective processing of data for general purposes.

### **1.1.1 DATA COLLECTION AND VALIDATION**

Data collection or data gathering is the process of gathering and measuring information on targeted variables in an established system, which then enables one to answer relevant questions and evaluate outcomes. It is a methodical process of gathering and evaluating accurate information from sources.

The collection of data is the first reasonable procedure in data processing. Required data has to be collated in bits before any kind of processing must take place. It is worth mentioning that data may exist as different types, and each of these types may require specific method of extraction and collation. Some of these types include discrete and continuous data, ordinal, nominal, ratio and interval data. Specific information about a person or a group of people may first be represented in any of these types, for example, in a typical survey form, age may be represented as numbers; name, address and occupation may be represented as strings of characters. Data may be obtained from various sources, some of which include interviews, surveys, online articles, books or documentations, recording devices, downloadable materials, sensors in the environment, etc.

Data collection can be mainly classified into primary and secondary aspects. The primary aspect involves the collection of original data directly from the source. This usually involves the use of specific techniques such as observations, experiments, surveys and interviews. The secondary aspect involves the use of pre-existing data originating from public sources, accessible databases, depreciated researches or institutional records in order to further a related research, test new hypothesis or determine certain factors related to an event.

During the collection process, the extracted data may be faulty and erroneous. It might also lack consistency in structure and format, especially when it's been obtained from different sources. Therefore, there is a need for verification of the data in order to eliminate useless and erroneous input. This ensures accurate, high quality and consistent outcomes.

### **1.1.2 DATA RECORDING**

After the collection of data, it must as well be recorded for further access and processing. This entails that the data is represented in a form that can be recognized and processed by humans or machines. According to an article by Science-direct, *Data recording is the production of measured variable information either automatically at set intervals or on demand.* This is required because further execution of the successive stages of data processing from this point, depends on data being available in a feasible form, which means that it must be practically available and modifiable during each processing stage. An easy and achievable way of ensuring this is by preserving and documenting the data concurrently with its collection.

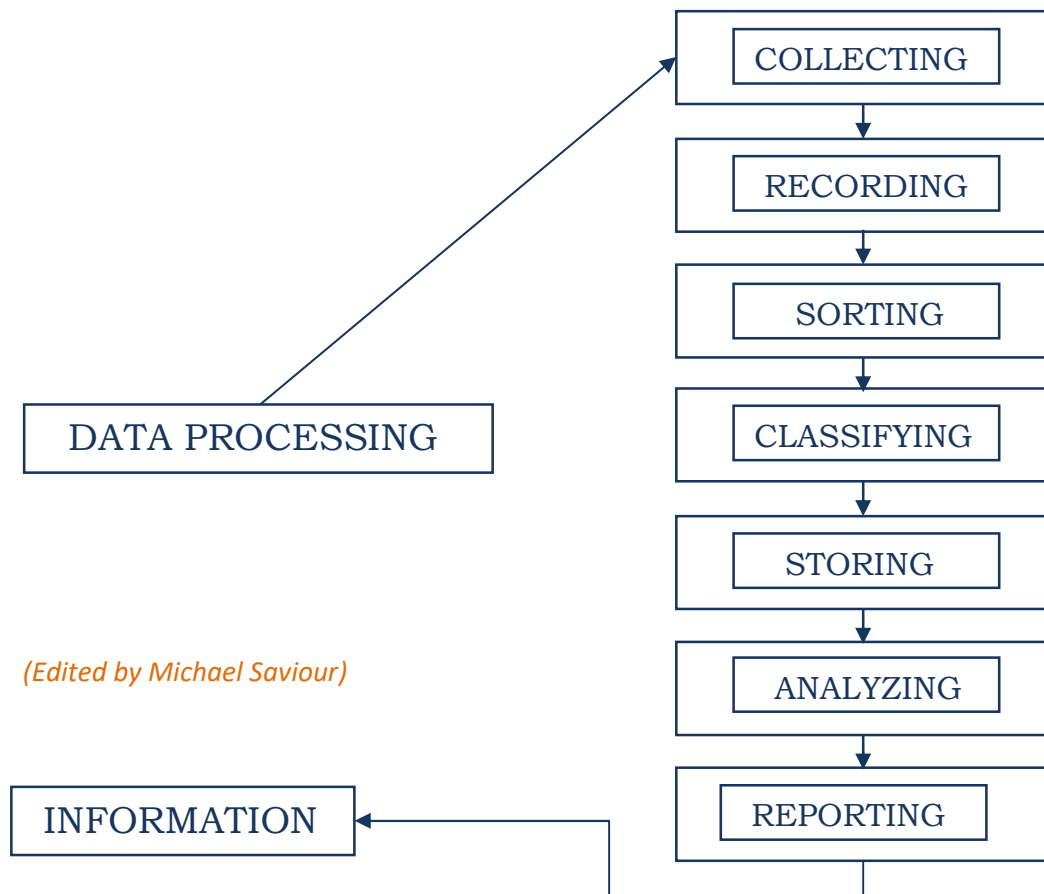
There are several methods through which recording of data can be accomplished. It can be achieved manually with the use of pen and papers, typewriters; mechanically, using devices like typewriters, mechanical printers or other mechanical processing systems; also through the use of specialized electronic

devices with the capabilities of performing efficient data accounting and organization, real-time signaling, batching, etc.

### **1.1.3 DATA SORTING**

Sorting is the method of arranging data in a specific order. It is done by ordering or categorizing items sequentially based on some certain relationship/properties that exist amongst them. Such properties could either be type, size, name, time of access/modification, group, and so on. The possibility of sorting is guaranteed only when the data is of similar nature, and its efficiency depends strictly on the method of sorting used. Sorting makes transformed data comprehensible and relatable; hence, it is the first level of organization. The computer is one of the devices that assist in the proficient sorting/organization of data. It may store data in structures such as arrays, linked list, maps, trees and graphs, and sorts them using highly efficient algorithms such as *insert sort*, *quick sort*, *bubble sort* and newer, faster sorting algorithms like the *radix sort*. These algorithms are well optimized; require minimal system resources whilst exploiting the full capacity of the processor. Sorting algorithms may be combined to handle special cases, for example, in sorting an array of data, the quick sort may be combined with an insertion sort for cases where the data is almost in a sorted order. This reduces the quadratic time complexity of sorting an almost sorted array of data using quick-sort to an efficient linear logarithmic function. Other methods of sorting discovered by humans all serve the purpose of putting data in a specific pattern.





#### 1.1.4 DATA CLASSIFICATION.

One of the contributors of this paper, *Marvelous*, clarified that there is a distinction between sorting and classification. He stated that *sorting involves arranging objects in order of their increasing or decreasing similarities while classification is a property of organization based on classes or groups of pre-defined traits*. That is, certain characteristics are attributed to entities in groups or sets rather than doing so singularly to the items of that constitute such entities. Since data can assume the representable state of any object/entity, it is safe to apply this logic to the definition of data classification. Therefore, *data classification is the process of organizing raw data into meaningful categories based on shared characteristics or attributes*. Data classification permits the addressing of data by certain characteristics which is also shared by other related data. It allows bulk structuring and organization of data, explicitly

implying that large data could be quickly condensed to the specific requirements, which are easily extractable and analyzed. This can be done on the basis of chronology, quantity, quality, physiology or geography. The classification of data could be structural, non-structural or semi-structural: structural in sense that the data is organized with a standard basis or format in place; non-structural when the data is classified with no pre-defined standard or format and semi-structural when the data is partially organized using a flexible scheme or format.

### **1.1.5 DATA STORAGE AND RETRIVAL**

We laid emphasis on the importance of preserving data when we discussed about data recording. Preserving data requires a methodical approach involving several techniques, tools and devices. *Data storage is the process of recording transformed data in a storage medium.* Data can be stored as handwriting, recording (phonological), in databases or in devices such as optical-drives, magnetic tapes, memory boards and so on. Some services provide an on-site or remote (clouds) data storage medium. These mediums may be different in types but all serve this same purpose, which is to preserve the integrity of the data stored in it. It is also necessary that the stored data can to easily located and retrieved at any point, making them available for use by other processing activities.

*(Edited by Michael Saviour)*

### **1.1.6 DATA ANALYSIS AND SUMMARIZATION**

According to Wikipedia, *data analysis is the process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.* This definition explicitly connotes a procedure that actively and strategetically exploits data in order to extract and transform it into meaningful information that is useful for decision-making by users or organizations. In the course of analysis, the data is divided

into separate experimental units, which then undergoes thorough examination. The process of data analysis is quite broad and complex, however this complexity revolves around the use of descriptive, predictive, diagnostic and prescriptive techniques to evaluate inputs and produce necessary outputs of importance.

Data analysis uses heavy computing to modulate data; therefore, various advance mathematical models and statistical theories are applied to resolve this process. Several tools are often used, mostly mathematical and computing softwares such as ELKI (data mining framework in Java with data mining oriented visualization functions), SciPy (Python library for scientific research), DevInfo, Panda, ROOT (C++ data analysis framework developed by CERN), etc; algorithms such as linear and logistic regression, decision trees, K-means clustering, Naïve Bayes and many others.

Data may further be summarized in order to maintain concise obtainable information.

### 1.1.7 DATA REPORTING

The purpose of processing data is so that it can be interpreted and communicated to people/things of interest. It involves the act of transferring, interpreting and communicating processed data (information) to a specific targeted audience.

*(Edited by Michael Saviour)*

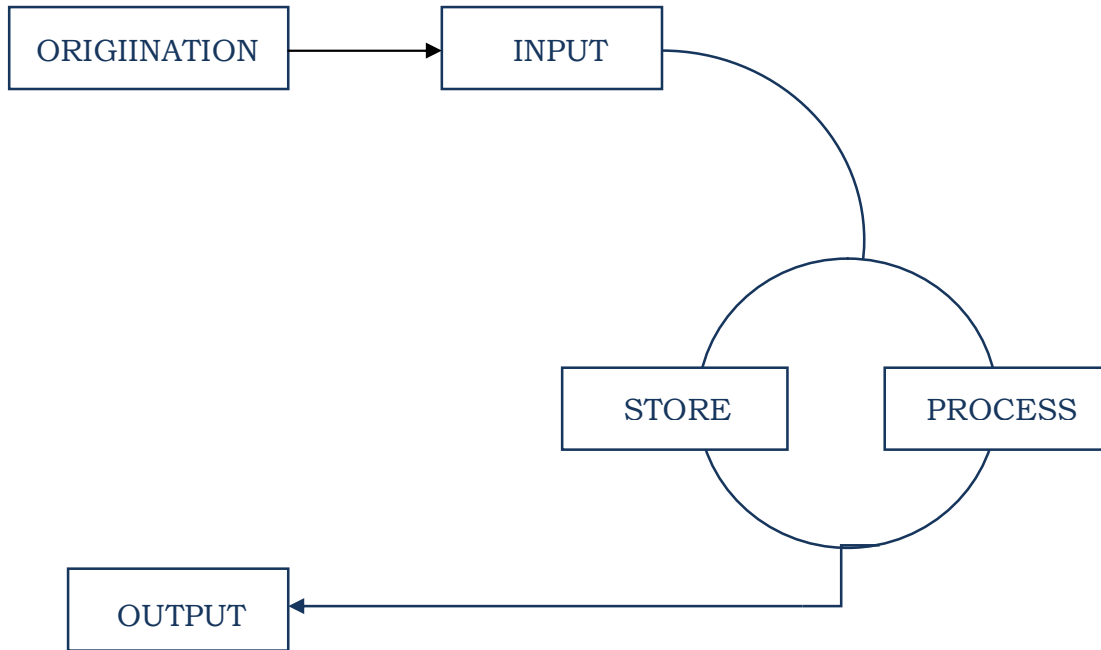
- **Section contributors:** Mbama Adviser, Martins Betty, Margaret Kpalap, Maxwell-leh Nuale, Nlerum Prosper Prayer, Noble Emmanuel, Marvellous Ihemjirika Chinemerem.
- **Signatures:** \_\_\_\_\_, \_\_\_\_\_, \_\_\_\_\_, \_\_\_\_\_, \_\_\_\_\_  
\_\_\_\_\_, \_\_\_\_\_

## **2.0**

### **DATA PROCESSING CYCLE**

Data processing is a dynamic process. This implies that the processing of data into information is a non-stop process. As long as available streams of data are provided as inputs, there is an expectation of acquiring generated outputs. This exemplifies the principle of GIGO (garbage in, garbage out). The more data that comes into the system, the more data that is processed, and goes out of the system, vice-versa. For example, a company may maintain a record of data of its staffs. On certain occasions the company may decide to withdraw or employ some staffs, or some may decide to resign or retire on their own accord. In all cases, there is either loss or addition of data. The information of the withdrawn staffs is expected to be removed from the company's record database and that of the newly employed must be added to the record. This requires frequent update and validation of the record to ensure that data integrity is maintained. These collection/removal, validation, update and maintenance of staff's information, is simply data processing. Regardless of the infinite amount of changes required to maintain an updated record, the steps in processing still remains unchanged. It only cycles through the process for different inputs.

Therefore, the repetitive nature of data processing is best described in terms of a processing cycle; sequence of procedures or steps required to perform task repetitively. A processing cycle involves origination (first intake of data), input, processing, output (desired end product) and storage. This is best visualized in the figure below. Observe that the procedural cycle involves processing and storing. The origination and input supplies the system with raw data which it uses to work. The end product is a well-processed output.



### 3.0

## DATA PROCESSING TOOLS

Several tools and technique can be applied to facilitate the generation of information through data processing. In the section, we will briefly consider some of these tools and their featured application to the processing of data.

### 3.1 APACHE SPARK

An original documentation of Apache Spark from their website, [spark.apache.org](http://spark.apache.org), describes the software as *a multi-language engine for executing data engineering, data science, and machine learning on single node machines or clusters*.

It unifies the processing of data in batches and real-time streaming interfacing with different programming languages such as Python, SQL, Java and R. Its key features include batch/streaming data, SQL analytics, exploratory data analysis, machine learning and many others.

### 3.2 STRUCTURED QUERY LANGUAGE (SQL)

This is a domain-specific language used to manage data, especially in relational database management system (RDBMS). It is the standard language as approved by the American National Standards Institute (ANSI), and the International Organization for standardization (ISO). It can retrieve, insert, update, delete or create records in a database. It also allows the creation of new record views, as well as setting permissions on specific record/field of data in a database.

### 3.3 MICROSOFT EXCEL

Microsoft Excel is a versatile spreadsheet software developed by Microsoft. It is used for data entry and management, charts and graphs, and project management.

Other tools that are worth mentioning are Python (a versatile scripting language for data analysis, manipulation, and machine learning), Avast (a security software that ensures safety and data integrity within a system), Talend, SAS, and Rivery.

- **Section contributors:** *Mejeh Tochukwu Ikechukwu, Monday Excellent Somtochukwu, Njobuanwu Godfrey Oronda , Nnenanya Emmanuel Ugochukwu, Nwanezi Emmanuel Chukwuka, Nwaobam Joseph Chinazaekpere*
- **Signatures:** \_\_\_\_\_, \_\_\_\_\_, \_\_\_\_\_, \_\_\_\_\_, \_\_\_\_\_  
\_\_\_\_\_, \_\_\_\_\_

## **4.0**

## **COMPUTER ORGANIZATION**

A computer is an electronic machine that stores and processes data. It is a programmable device that can automatically carry out sequences of arithmetic or logical operations. These operations are known as programs.

Organization is the collection of objects/entities, in a regular, structural and defined manner.