ICT/Data Analysis in the Sciences

Samuel Ogbonna Enibe
Professor of Mechanical Engineering
formerly Dean, Faculty of Engineering (01-08-2014–31-07-2016)
University of Nigeria, Nsukka
Email: samuel.enibe@unn.edu.ng

[1]

Outline

Contents

[1] being text of paper delivered at the Workshop for Postgraduate Students in Science & Engineering organized by the School of Postgraduate Studies, University of Nigeria, Nsukka on 12 to 13 October 2017 as part of the Course: PGC 601 – Research Methodology and Application of ICT in Research

Abstract

In this contribution, data mining and analysis are presented as key aspects of postgraduate research methodology. Key steps are shown to involve clear problem definition, its subdivision into a set of simple problems, the solution of each of these problems and assembly of the solutions to achieve an integrated whole. Recommended steps in data analysis are presented. The use of information and communication technology (ICT) facilities for data gathering is highlighted. Emphasis is placed on the use of open source software where possible. For scientific and engineering analysis, the use of MATLAB/SCILAB is encouraged.

1 Introduction

2 Major steps in research

Research is a recursive process which has been successfuly applied to move civilization forward to greater heights. Some of the major steps in research described in Enibe (2008) are shown in figure 1.

Major steps in research They are itemized as follows:

1. Problem definition
2. Subdivision into n simple problems
3. Solution of each of the n simple problems
4. Assembly of the n solutions
5. Satisfaction tests
6. Report

Each of these is now described briefly
Problem Identification/definition

- The first step in the proper solution of any significant problem is its proper identification.
- This involves proper *separation* of the problem from all other problems and its *description* in some detail in the relevant technical terms.
- The relationship of the problem to be solved with its environment is *represented* in terms of input/output or causes and effects.
- Any problem that is sufficiently *identified*, *described* and *represented* is likely to be effectively solved.

Subdivision into Simple Problems

Every problem can usually be subdivided into a set of simple sub-problems each of which can be solved quite independently. This approach has several advantages:

1. Many of the sub-problems may have existing solutions, which may be easily adapted, leading to a quicker solution of the main problem.

Subdivision into Simple Problems continued

2. the sub-problems may be assigned to different persons or groups of persons, working simultaneously with suitable coordination and cooperation, the lead time for the solution of the main problem may be significantly reduced.

3. the solution of each sub-problem may result in the development of a new process, technique or technology which may find application in several other fields.

Solution of each of the n simple problems

4. Each of the simple n sub-problems problems derived from the main problem is solved simultaneously using the methodology relevant to the specific field or subject area. Typically, this follows the procedure shown in figure 2..

    - A good solution of each of the n simple problems will usually involve a detailed study of the problem to understand its internal characteristics and how these characteristics affect its behaviour under external stimuli.
    - A good solution of a simple problem may often lead to a description of the characteristics using suitable technical expressions, such as mathematical relationships.
    - In such cases, prediction of the system behaviour under a wide range of values of the external stimuli may be possible.

Steps for the solution of a simple problem
Assembly of Solutions

5. Assembly of Solutions

    - The solutions to the set of sub-problems are gathered to form a solution to the main problem. Thus, we have

    $$\text{Solution} = \sum_{i=1}^{n} S(i)$$

    where $S(i)$ is the solution to subproblem i.
    - A thorough review is often recommended before application to the real problem.
    - With suitable feedback, the solution can be continuously improved for greater effectiveness.

3 Data mining

Data is any set of informative that can be obtained, quantified and represented in an acceptable form. The term "mining" or gathering/collection of relevant data is a major and often time-consuming activity in any research project.

3.1 Data Types

Project Data types Research projects may result in data of different types, such as numerical data, string data, Boolean data or stream data. These are further described as follows:

- Numerical data. These may be integers $(1, 2, 3, 4, \cdots,)$, floating point numbers (eg 1.2345), or complex numbers ($c = a + bi$), where a and b are real numbers, and $i = \sqrt{-1}$.
- String data. These are alphanumeric characters such as names (eg Sunday, Monday, Tuesday, ...,), identification marks, etc.
- Stream data
- Boolean data In many cases, data may be represented in only one of two ways, such as on/off, yes/no, true/false, alife/dead, good/bad, possible/impossible, white/black, 0/1, etc). A number system based on this concept is called the binary number system.

3.2 Data Sources

Research data for typical postgraduate projects may be obtained in many different ways, depending on the type of project. These include data from the literature, experimental data, questionnaire data or simulated data. Often, these data sources can be combined for as the need arises. These may be described further as follows:

3.2.1 Literature Data

Literature Data A careful and well-coordinated sreach of the literarure in a given project can provide very useful research data for further consideration. Care must be taken, however, to verify the authenticity of the data from each source. In particular, the methods employed by the original authors, their data accuracy and bounds for application must be carefully considered.

3.2.2 Experimental Data

Experimental Data

- Many research projects in science and engineering require extensive experimentation in the laboratory, workshop or field
- For effectiveness, the design of experiments using well established statistical tools are very strongly recommended.
- Many of these tools to ensure efficient data analysis are presented in Chukwu (2013)

3.2.3 Questionnaire Data

Questionnaire Data

- The use of questionnaire is common in the social sciences, and occasionally in the pure and applied sciences.
- These should also be well designed using the well established tools in the literarure.
- Questionnaire should be pretested thoroughly before final administration to the target audience.

3.2.4 Simulated Data

Simulated Data

- Simulated data may be generated from numerical experiments using a comptational procedure
- The numerical experiments should also be well designed

3.3 Experimental Design

In broad terms, an **experiment** is any activity specifically undertaken to obtain relevant data for a given study or project. According to Chukwu (2013), an experiment may be defined as a test or series of tests in which purposeful changes are made to the input variables of a process or system so that we may observe and identify the reasons for changes in the output responses.

In order to obtain sufficient and reliable data that could be analysed to produce useful results, the process of generating, gathering or "mining" the data must be carefully designed. Some of the well-established principles in this regard will be briefly highlighted.

3.3.1 Characteristics of a well-designed experiment

Chukwu (2013) identifies 5 important characteristics of a well-designed experiment. These are:

- Simplicity : The selection of treatments and the experimental arrangement should be as simple as possible, consistent with the objectives of the experiment
- Degree of precision: The probability would be high that the experimenter will be able to measure treatment differences with the degree of precision the experimenter desires. This implies an appropriate design and sufficient replication
- Absence of systematic error: The experiment must be planned to ensure that experimental units receiving one treatment, in no systematic way differ from those receiving another treatment so that an unbiased estimate of each treatment can be obtained.
- : Range of validity of conclusion: Conclusions should have as wide a range of validity as possible. An experiment replicated in time and space would increase the range of validity of the conclusions that could be drawn from it. A factorial set of treatment is another way for increasing the range of validity of an experiment. In a factorial experiment the effects of one factor are evaluated under varying levels of a second factor.
- Calculation of degree of uncertainty: In every experiment there is always some degree of uncertainty of conclusions. The experiment should be designed so that it is possible to calculate the probability of obtaining the observed results by chance alone

3.3.2 Recommended experimental procedure

Chukwu (2013) recommends an eleven step procedure fro experimentation in order to guarantee reliable data forfurther consideration. This can be expanded to the 14-step procedure illustrated in figure 3.

Recommended experimental procedure

Each of these steps is further described below:

1. **Definition of the problem**. The first step in problem –solving is to state the problem you are dealing with clearly and concisely. Once the problem is understood you should be able to formulate questions which, when answered will lead to solution.

2. **Statement of the objectives**. This may be in the form of question to be answered, the hypothesis to be tested, or the effects to be estimated. Objectives should be written out in precise terms. When this is done the experimenter is able to plan his experimental procedures more effectively. When there is more than one objective, prioritize them as this might have a bearing on the experimental design. In stating objective don't be vague or ambition.

3. **Selection of treatments**. The success of the experiment rests on the careful selection of treatments, whose evaluation will answer the questions posed.

4. **Selection of experimental material**: In selecting experimental material, consider the objective of the experiment and the population about which inferences are to be made The material used should be representative of the population on which you wish to test y our treatments.

5. **Selection of experimental design**. Here again a consideration of objectives is important, but a general rule would be to choose the simplest design that is likely to provide the precision you require.

6. **Selection of units of observation and the number of replication**. For field experiment in agriculture with plants, this means deciding on the size and shape of the plots. For experiments with animals, this means deciding on the number of animals to consider as a unit. In engineering processing plants a unit could be 2000 bolts etc. Here plot size and the number of replication should be chosen to produce the required precision of treatment estimates

7. **Consideration of data to be collected**. The data collected should properly evaluate treatment effects in line with the objectives of the experiment. In addition, consideration should be given to collection of data that will explain why the treatments perform as they do.

8. **Outlining statistical analysis and summarization of results**. Write out the sources of variation and associated degrees of freedom in the analysis of variance. Include the various F tests you may have planned. Consider how the results might be used and prepare possible summary tables or graphs that will show the effects you expect. Compare these expected results to the objectives of your experiment to see if the experiment will give the answers you are looking for.

9. **Conducting the experiment**. In the conduct of the experiment, use procedures that are free from personal biases or favouritisms. Make use of the experimental design in collecting data, so that differences among individuals or differences associated with order of collection can be removed from experimental errors. Organize the collection of your data to facilitate analysis and to avoid errors in recopying. If it is necessary to copy data, check the copied figures against the originals immediately

10. **Analyzing data and interpreting result**. All data should be analyzed as planned and the results interpreted in the light of the experimental conditions, hypothesis tested and the relation of the results to facts previously established. Understand that there is always a

possibility of making a wrong conclusion and so put into consideration the consequences of making an incorrect decision. Don?t jump to a conclusion, even though it is statistically significant if the conclusion appears out of line with previously established facts. In this case, investigate the matter further

11. Conduct trial experiments with sample data
12. Analyse results from trial experiments
13. If trial results are okay, proceed to next stage, otherwise go back to step 2
14. **Preparation of a complete**, readable and correct report of the research. If the null hypothesis is not rejected, it is a positive evidence that there may be no real difference among the treatments tested.

4 Data Analysis

derived from any given project must be processed to facilitate its interpretation and serve as a guide to decision making. This process is generally called data analysis.

4.1 Purpose of data analysis

Purpose of data analysis

- The main purpose of the analysis of any set of data is to obtain some understanding of the underlying features of the data and thus provide explanations, relationships, comparisons, generalizations and theories.
- This may result in some correlation, equations, recommended procedures or algorithms which capture the underlying processes responsible for the data. In turn, these could be used to predict some future outcomes given conditions similar to the ones associated with the data.

Purpose of data analysis continued

- The overall goal should be to use the data analysis to provide reasonable solutions to real existing or anticipated problems. They may also be applied to develop entirely new systems, processes or technologies.

4.2 Steps in data analysis

The process of data analysis could proceed in the seven steps suggested in figure 4.

As shown in the figure, data analysis is usually a recursive process

Design format for data storage

- **Paper storage** For small amounts of data, output could be written out manually on paper. If possible, copies could be made and stored in different locations for safety. The media should be protected from destructive environments (such as moisture, chemicals, etc), rodents and insects.

- **Electronic storage**. This form of storage is prefered if the amount of data is large or if a computerized data acquisition system is employed. Copies of the data should be stored in different locations or different online storage sites for safety.

Data cleaning

- Remove spurious data
- Remove outlanders
- Remove obvious typing errors

Store data sequentially

- Record data as and when received
- Assign unique mark or label to each record for ease of identification

Determine required output

- Determine output required from the data analysis
- **Tables** Output data should be arranged in tables for easy reference wherever possible. Each column should contain the same type of data, with the first row providing a brief but meaningful heading. A good example is shown in table 1.
- **Graphs** This could be line graphs, barcharts, piecharts among others.
- Equations, mathematical correlations, etc

Table 1: Recommended average days for each Month and values of daynumber n by Months

		For the Average Day of the Month		
Month	n for i^{th} day of Month	Date	n, Day of Year	δ, declination
January	i	17	17	-20.9
February	31+i	16	47	-13
March	59 +i	16	75	-2.4
April	90 + i	15	105	9.4
May	120 + i	15	135	18.8
June	151 + i	11	162	23.1
July	181 + i	17	198	21.2
August	212 + i	16	228	13.5
September	243 + i	15	258	2.2
October	273 + i	15	288	-9.6
November	304 + i	14	318	-18.9
December	334 + i	10	344	-23

Source: Duffie & Beckman, 1991

Graphs and their benefits Graphs are very useful in providing a visual representative of data. Some of the advantages of using graphs include:

- A graph condenses a large amount of data into a small space for visual examination. For example, a data set spanning several pages of tabulation can be compressed into a simple graph in a single page.
- A graph may suggest a data trend which could prove useful in correlating the data with mathematical expressions.

Graphs and their benefits continued

- For example, the data in column 5 of table 1 can be plotted as a function of the daynumber, n where $n = 1$ on January 1, and $n = 365$ on December 31. The value of n for the i^{th} day of each month is shown in column 2 of table 1. Such a plot is shown in figure 5.
- The plot suggests a sinusoidal relationship, and has been correlated with the equation 1

$$\delta = 23.45 \sin\left(360 \frac{284 + n}{365}\right) \tag{1}$$

Graphs and their benefits continued

- A graph may show spurious data that could be due to some experimental erros or that suggest that the assumptions in the analysis have been exceeded or are not valid.

Graphs and their benefits continued

Bar charts

- Line graphs are the graphical form of choice for scientific reports. However, sometimes barcharts may be used, especially when the number of data points is small, when the visual effect is more important, when the the x-axis data are discrete or strings, or when no particular mathematical correlation is desired.
- As an example, consider the income and expenditure data shown in table 2.

Bar charts continued

Table 2: Example income and expenditure data

Year	Income	Expenditure
2005	13	9
2006	21	18
2007	15	13
2008	17	14
2009	20	29
2010	32	36
2011	15	17

Bar charts continued

Pie Charts A pie chart may be used to display data using sectors of a circle, where the angle subtended at the centre is proportinal to a given item.

- For example, suppose a given data y contains three elements of values 2, 5 and 9.

- The sum of the elements is 16.
- Each element as a fraction of the whole is calculated as in column 2 of table 3 and multiplied by 360 to get the angle at the centre in column 4.
- The data in column 4 is now employed to draw the pie chart shown in figure 7.

Pie Charts continued

Table 3: Sample calculation for drawing a pie chart

i	Item, y(i)	Fraction of total, $f = y(i)/\sum_{i=1}^{n} y(i)$	Percent of total	$Angle = 360 \times f$
1	2	0.125	12.5	45.
2	5	0.3125	31.25	112.5
3	9	0.5625	56.25	202.5
Total	16	1.0	100	360

Pie Charts continued
Select or develop software for data analysis

- Select an existing software for data analysis
  - Many software already exist in most fields of learning
  - Select most commonly used ones in your field
  - Open source software
  - Commercial software
  - Selected software should be configured, if possible, to produce all the required output of the analysis upon application of the data
- Develop new software for data analysis if no suitable software exists

Compare output with expected results

- Determine a suitable criteria for comparisom of output with expected results.
- In many cases, some statistical technic could be used.
- If there is good agreement, proceed to the next stage.
- If there is substantial disagreement, review the data analysis steps, and/or the entire research process if necessary

4.3 Data analysis software

Types of data analysis software

- Many data analysis software exist.
- Spreadsheets

- General purpose software
- Special purpose software. Some of these may be specific to a particular discipline or even subdiscipline

Spreadsheets

- Spreadsheets are widely available and suitable for many routine tasks of data manipulation, such as sorting, arithmetic operations and even linear algebra
- The most common spreadsheet software is Microsoft Excel. Others include OpenOffice and its variants
- Spreadsheet software are very easy to learn and use, and are usually interactive and menu driven
- Their capabilities have grown in recent years
- Many software developers design their software for greater compatibility with one or more Spreadsheet software
- Further details are available in Liengme (2008)

General Purpose software

- A typical general purpose software is highly configurable and can be adapted to a wide variety of applications
- Hence, they require some form of *programming* knowledge for their use.
- Some form of learning is required for their productive use.
- Typical examples are shown in tables 4 and 5

Some high level programming languages

Table 4: High level programming languages

Language	Appeared In	Developed by
FORTRAN		
C	1972	Denis Ritchie & Bell Labs
C++		
C#		
JAVA	1995	James Gosling & Sun Microsystems
Visual Basic		

Some high level and matrix-based very High Level Programming languages

Table 5: Matrix-based very High level programming languages

Language	Appeared In	Develop
C	1972	Denis Ritchie and Be
Java	1995	James Gosling and S
LabView	1986	National Instruments
Maple	1980	Waterloo Maple Inc
Mathematica	1988	Wolfram Research
Matlab	Late 1970s Commercialized in 1984	Cleve Moler and Ma
Octave	1994	John W. Eaton
Python	1990	Guido van Rossum a Foundation
R	1996	R. Ihaka and R. Gent
Scilab	1994	INRIA and École nat chaussées (ENPC)

Source: Soni (2008)

4.3.1 Special purpose software

- Many commercial special purpose software exist for data analysis in specific disciplines.
- Of these, the SPSS (Statistical Package for Social Science Research) and Minitab are probably more widely used.
- They usually have menu-driven functions for common tasks required in data analysis (especially basic statistics) and good plotting and data visualization features.
- They have the advantage of being easy to learn and use, but have limited scope for programmability.
- As a result, they can only be used for pre-defined applications
- Examples in selected fields are now highlighted

4.3.2 Agriculture

- Baiyeri (2013) reviewed the list of software used in Agriculture and indicated that apart from SPSS, SAS (Statistical Analysis Software), GENSTAT and MSTAT-C have gained

much grounds.

- Each of these software can perform many of the statistical analysis required in agricultural research.

4.3.3 Physical Sciences

- According to Oyesanya (2013), the most commonly used software for data analysis in the Physical Sciences include Maple, Mathematica and Matlab. Others include REPSN and SINGULAR
- SCILAB has proved to be an open source alternative to MATLAB and is also widely used.

4.3.4 Biological & Pharmaceutical Research

- Osadebe (2013) considered commonly used software for pharmaceutical research
- Many of these are also applicable to biological sciences research

Data types in pharmacy research

According to Osadebe (2013), these include

1. Biological data such as protein, DNA, RNA sequences etc
2. Chemical data such as small molecules (drugs)
3. Experimental data such as absorbance and pH readings.
4. Miscellaneous data such as demographic, disease incidence

Visualization methods Osadebe (2013) indicates these as

- Generative topographic mapping(GTM)
- Hierarchial GTM
- Principal Component Analysis(PCA)
- Sammon's Mapping
- Self-organising Maps(SOMs)

Visualization tools for biological data Osadebe (2013) lists them as follows:

- VMD (`http://www.ks.uiuc.edu/Research/vmd/`)
- PyMol (www.pymol.org)
- Maestro (`https://www.schrodinger.com/products/14/12/`), etc

Visualization tools for chemical data Osadebe (2013) lists these as follows:

- Maestro (https://www.schrodinger.com/products/14/12/)
- Avogadro (avogadro.openmolecules.net)
- Chemcraft (www.chemcraftprog.com) etc
- Ligand explorer (`http://www.rcsb.org/pdb/explore/viewerLaunch.do?viewerType=LX&structureId=1Y7H&hetId=SCN`)

4.3.5 Engineering

Software for engineering analysis

- In the past, the major software for engineering analysis was FORTRAN.
- With the development of **C, C++** and their variants, they came into general use in addition to, or in place of, FORTRAN
- In recent years, **MATLAB** and its variants such as **SCILAB**, **R**, **OCTAVE**, **PYTHON**, among others have virtually replaced FORTRAN, c, C++, etc as the primary software for engineering data analysis.
- A more detailed discussion on this subject may be found in Enibe (2013), Soni (2008), among others.

4.4 Open source software (OSS)

OSS Philosophy

- A commercial software, such as MATLAB, MAPLE, SPSS, etc, ought to be purchased and paid for before legal use. Any installation of an unlicensed version of the software may be an infrringment of the copyright of the developers.
- In contrast, an open source software (OSS) may generally be available for free download, installation and use. The developers usually depend on donations, training and other activities to sustain their operations.

OSS Philosophy continued

According to Oranu (2013),

- Open source describes a practice in production and development in which the raw materials of an end product are made available to the users of the end product.
- In most cases, open source actually promotes access to the source material allowing input and modifications.
- It should be noted that the phrase "open source" is not exclusively associated with software.

OSS Philosophy continued

- In fact, the phrase actually predated computer and indeed other products have their own open source versions, for instance, beverages (eg OpenCoke an open source alternative to Coca Cola), electronics (eg openmoko – an open source mobile phone), health sciences, etc.

OSS Philosophy continued

Oranu (2013) traces the development of the OSS concept from the 1960's

- He explains that **Open Source Software** is therefore a software that has been released under a license which allows access to the source code, permits users to study, change, improve and at times also distribute the software.
- He lists over 30 of the most commonly used OSS, most of which are server infrastructure.

OSS benefits

Oranu (2013) outlines many benefits of open source software. These include

- Low or no cost of ownership (apart from download and related costs)
- Better learning outcomes
- Freedom from vendor hijack or dependence
- Opportunity to customize the software and contribute to knowledge development
- License rights

OSS benefits continued

- Security
- Transparency
- Interoperability and portability
- Stability/continuity

4.5 MATLAB/SCILAB for data analysis

MATLAB/SCILAB features

Introduction to MATLAB/SCILAB for data analysis

- From over 30 years of University level teaching and research as well as extensive use of a wide variety of software for data analysis, the present author encourages the use of MATLAB, SCILAB or similar software for data analysis, modeling and simulation, especially in science and engineering.
- Key features of the software are shown in figure 8
- A very good introduction to MATLAB and its use for data analysis is presented in Kiusalaas (2008)
- A simple but authoritative introduction to SCILAB by the software developers is presented in Baudin (2010)

They have a wide variety of applications, including

- heterogeneous simulations and data-intensive analysis of very complex systems and signals,
- comprehensive matrix and arrays manipulations in numerical analysis,
- finding roots of polynomials,
- two-and three-dimensional plotting and graphics for different coordinate systems,
- integration and differentiation,
- signal processing, control, identification, symbolic calculus,

- statistical analysis
- regression analysis

Other applications, include

- solution of linear and non-linear algebraic equations
- initial-value problems
- boundary value problems
- optimization, etc.

MATLAB/SCILAB benefits

Many benefits of using MATLAB/SCILAB are presented in Enibe (2013). They enable the users to

- solve a wide spectrum of analytical and numerical problems using matrix-based methods,
- attain excellent interfacing and interactive capabilities,
- compile with high-level programming languages,
- ensure robustness in data-intensive analysis and heterogeneous simulations,
- provide easy access to and straightforward implementation of state-of-the-art numerical algorithms,
- guarantee powerful graphical features, etc.

MATLAB/SCILAB benefits continued Due to high flexibility and versatility, the MATLAB/SCILAB environments have been significantly enhanced and developed during recent years. This provides users with advanced cutting-edge algorithms, enormous data-handling abilities, and powerful programming tools.

- Graphics capabilities include
  - barcharts
  - Pie charts
  - 2D plots
  - 3D plots and surfaces
- Over 110 toolboxes to handle most commonly required applications

5 ICT in Scientific Research

Information and communications technologies (ICT) afford us the opportunity for instant communication with a large number of people anywhere on earth where the facilities are available, and at negligible cost. In this paper, therefore, we first explore a brief history of some componenets of the ICT, especially the Internet, followed by a review of several means of Internet connection. We then explain various means of obtaining information from the Internet as well as some of the key beneficial uses of the facility.[2]

The current worldwide revolution in information and communication technology is not accidental, but was clearly ancticipated over 3000 years ago. A typical example is the Biblical text in Daniel 12:4, which reads "But thou, O Daniel, shut up the words and seal the book, even to the time of the end; many shall run to and fro, and knowledge shall increase".

King Solomon may have alluded to this in the Biblical text of Ecclesiatis 12:12 when he said, inter alia, "... of making many books there is no end". At no time in history has the technology for universal increase in knowledge and its dissemination in the form of books/multimedia been made available to mankind in potentially every part of the globe as it is today. With the development of information and communication technology within the past 3 to 5 decades, the rate of generation and dissemination of knowledge has become even much higher than many of us are aware of. Because of the impact of Internet technology on every facet of modern life, it is important that all of us know what it is, its origins, how to set it up and use it to our own benefit and those of our family members. Some dangers of its negative uses are also highlighted.

5.1 What is ICT

A key component of the information and communication technology is the Internet.

The Internet has been described by Wikipedia as

- *a global system of interconnected computer networks that use the standard Internet Protocol Suite (TCP/IP) to serve billions of users worldwide.*
- *It is a network of networks that consists of millions of private, public, academic, business, and government networks, of local to global scope, that are linked by a broad array of electronic, wireless and optical networking technologies.*
- *The Internet carries a vast range of information resources and services, such as the inter-linked hypertext documents of the World Wide Web (WWW) and the infrastructure to support electronic mail.*
- *Most traditional communications media including telephone, music, film, and television are reshaped or redefined by the Internet, giving birth to new services such as Voice over Internet Protocol (VoIP) and IPTV. Newspaper, book and other print publishing are adapting to Web site technology. The Internet has enabled or accelerated new forms of human interactions through instant messaging,*

[2]This section of the paper is largely derived from the text of lead paper presented at the maiden National Conference of the School of Sciences of the Federal College of Education (Technical), Umunze, Anambra State on 10^{th} September 2013. This was in turn adapted and presented at the Odenigbo Zonal camp seminer of the Scripture Union (Nigeria), Nsukka Area on 14/12/2013

- *Internet forums, social networking. and online shopping have boomed both for major retail outlets and small artisans and traders. Business-to-business and financial services on the Internet affect supply chains across entire industries.*

A good visualization of the Internet is shown in figure 9.

5.2 History of ICT

The Internet was the result of some visionary thinking by people in the early 1960s who saw great potential value in allowing computers to share information on research and development in scientific and military fields (Howe, 2010).

According to Wikipedia (2011),

- *The origins of the Internet reach back to research of the 1960s, commissioned by the United States government in collaboration with private commercial interests to build robust, fault-tolerant, and distributed computer networks.*
- *The funding of a new U.S. backbone by the National Science Foundation in the 1980s, as well as private funding for other commercial backbones, led to worldwide participation in the development of new networking technologies, and the merger of many networks.*
- *The commercialization of what was by the 1990s an international network resulted in its popularization and incorporation into virtually every aspect of modern human life. As of 2009, an estimated quarter of the Earth's population used the services of the Internet.*
- *The Internet has no centralized governance in either technological implementation or policies for access and usage; each constituent network sets its own standards. Only the overreaching definitions of the two principal name spaces in the Internet, the Internet Protocol address space and the Domain Name System, are directed by a maintainer organization, the Internet Corporation for Assigned Names and Numbers (ICANN).*
- *The technical underpinning and standardization of the core protocols (IPv4 and IPv6) is an activity of the Internet Engineering Task Force (IETF), a non-profit organization of loosely affiliated international participants that anyone may associate with by contributing technical expertise.*

Further details are available online at Wikipedia (2011), Howe (2010), among others.

5.3 Gaining access to the Internet

To gain access to the Internet, one requires at least the following:

1. An Internet-enabled device (Such as Computer, GSM handset, ipad, etc)
2. A software for accessing the Internet, usually called a web-browser
3. An Internet service, usually through an Internet Service Provider (ISP).

Each of these will now be discussed briefly:

5.3.1 Internet Enabled Device

A very wide variety of computing devices have facilities for connection to the Internet. These include:

1. **Desktop Computers**

   Most modern desktop computers come with the necessary hardware for Internet connection. When this is not the case, the hardware can readily be obtained and installed.

2. **Laptop Computer**

   Most, if not all laptop computers, notebooks, etc are Internet ready. It is not recommended to purchase any such device nowadays if it is not Internet-ready.

3. **Ipad Computers** The Ipad is a line of tablet computers designed, developed and marketed by Apple Inc. primarily as a platform for audio-visual media, including books, videos, etc. Most Ipads are Internet-ready.

4. **Mobile Phones**

   Many medium to high-prize mobile phones are preconfigured to access the Internet. These include the very expensive Black Berry models, the medium-prized Nokia E-series models to highly affordable Android phones with or without large screens.

5. **Other Mobile Devices**

   There are a large variety of other mobile devices which are preconfigured for Internet connection.

   These include

   - Datacards
   - Hand-held game consoles
   - Cellular routers
   - Point of sale (POS) devices

   Check the user manual of your device to find out if it is Internet-enabled.

5.3.2 Web-browser

A web-browser is a software for accessing files, documents and other resources available on the Internet. It enables users navigate from one web page to another via hyperlinks embedded in the documents. These documents may also contain any combination of computer data, including graphics, sounds, text, video, multimedia and interactive content including games, office applications and scientific demonstrations. The most-commonly used web-browsers today are

- Internet Explorer by Microsoft Corporation
- Mozilla Firefox web-browser
- Google Chrome web-browser
- Opera
- Safari by Apple Inc.

5.3.3 Internet Service Providers (ISP)

Internet Service Providers (ISPs) are Companies or government agencies whose key function is to provide Internet service, with or without the payment of a fee. The Internet Services may be provided through a variety of means, such as

- GSM phone network
- Dial-up telephone service
- Wireless network
- Wired network, which may be of copper cabling or fibre-glass cabling.

5.4 Uses of the Internet

Use of the Internet has penetrated virtually every area of human endeavour. Within the limited space and time available in this seminar, only a few of them will be highlighted.

5.4.1 Information Dissemination

The Internet is an effective and low-cost means of information dissemination worldwide. Only a few examples can be listed here.

1. **URLs**

   The uniform resource locator (URL) is a global means of uniquely identifying a particular document anywhere it is on any computer in the world. It allows people to symbolically identify services provided such that clients can locate and address the web services, file servers and other database that store documents and access them using the Hypertext Transfer Protocol (HTTP). In this way, it is possible to allow software systems to communicate in order to share and exchange data.

2. **Internet Search**

   This is a system whereby each document being made available on the Internet is thoroughly analyzed such that the most important key words in the document are identified. These keywords are then stored in a database called a search Engine. Once a keyword is submitted by user a user to the sear engine, he is automatically linked to all available documents for which that keyword has been associated.

   Currently, the most important search engines are Google and Yahoo.

   For Example, by typing any of the words *Matthew Henry* or M*atthew Henry Commentary*, on the Google or Yahoo search bar, the user may be automatically linked to website from where the complete volumes of the *Matthew Henrys* Commentary may be downloaded free of charge!

3. **Publishing** As observed in Wikipedia (2011), *the Web has enabled individuals and organizations to publish ideas and information to a potentially large audience online at greatly reduced expense and time delay. Publishing a web page, a blog, or building a website involves little initial cost and many cost-free services are available. Publishing and maintaining large, professional web sites with attractive, diverse and up-to-date information is still a difficult and expensive proposition, however.*

4. **Advertising**

   The Internet is a very effective and low-cost means of advertising worldwide. The sale of products and services directly on the Internet continues to grow. Many jobs, admission vacancies and scholarship opportunities are now advertised partly or exclusively online.

5. **News** Many people use the Internet to access news, weather and sports reports.

5.4.2 Communication

Among the many communication services available on the Internet, the three that will be highlighted are electronic mail, Internet telephone and short message services (SMS).

1. **Electronic Mail**.

   Wikipedia (2011) defines the Electronic mail, or email, as *an important communications service available on the Internet. The concept of sending electronic text messages between parties in a way analogous to mailing letters or memos predates the creation of the Internet. Pictures, documents and other files are sent as email attachments*. Emails can be copied to multiple email addresses.

2. **Internet Telephone**.

   Internet telephony is another common communications service made possible by the creation of the Internet. VoIP stands for Voice-over-Internet Protocol, referring to the protocol that underlies all Internet communication. The idea began in the early 1990s with walkie-talkie-like voice applications for personal computers. In recent years many VoIP systems have become as easy to use and as convenient as a normal telephone. The benefit is that, as the Internet carries the voice traffic, VoIP can be free or cost much less than a traditional telephone call, especially over long distances and especially for those with always-on Internet connections such as cable or ADSL. VoIP is maturing into a competitive alternative to traditional telephone service. Interoperability between different providers has improved and the ability to call or receive a call from a traditional telephone is available. Simple, inexpensive VoIP network adapters are available that eliminate the need for a personal computer.

   Voice quality can still vary from call to call but is often equal to and can even exceed that of traditional calls. Remaining problems for VoIP include emergency telephone number dialing and reliability. Currently, a few VoIP providers provide an emergency service, but it is not universally available. Traditional phones are line-powered and operate during a power failure; VoIP does not do so without a backup power source for the phone equipment and the Internet access devices.

   Internet telephone service is provided by **Skype**, among others.

3. **Short Message Service (SMS**

   The short message service (SMS) is a means of exchanging short text messages over a telephone network. Each message is typically limited to 160 characters. A message longer than 160 characters is usually broken into pages of no more than 160 characters, and charged accordingly. On a normal GSM network, an SMS may cost about N5.00.

   It is now possible to send SMS through the Internet, often at less than half of the cost through a GSM network.

To use such a service, typically one registers in the website of a company specialized for the service. The appropriate fees may be paid online through a bank. Thereafter, the amount is credited to the client and deducted for each SMS sent. Some companies may charge as low as N1500 for a set of 1000 short messages.

This is particularly convenient for sending the same message to many recipients, such as notice of meetings, invitations, etc.

5.4.3 Data Transfer

File sharing, as observed in Wikipedia (2011),

> is an example of transferring large amounts of data across the Internet. A computer file can be emailed to customers, colleagues and friends as an attachment. It can be uploaded to a website or FTP server for easy download by others. It can be put into a "shared location" or onto a file server for instant use by colleagues. The load of bulk downloads to many users can be eased by the use of "mirror" servers or peer-to-peer networks. In any of these cases, access to the file may be controlled by user authentication, the transit of the file over the Internet may be obscured by encryption, and money may change hands for access to the file. The price can be paid by the remote charging of funds from, for example, a credit card whose details are also passed – usually fully encrypted – across the Internet. The origin and authenticity of the file received may be checked by digital signatures or by MD5 or other message digests. These simple features of the Internet, over a worldwide basis, are changing the production, sale, and distribution of anything that can be reduced to a computer file for transmission. This includes all manner of print publications, software products, news, music, film, video, photography, graphics and the other arts. This in turn has caused seismic shifts in each of the existing industries that previously controlled the production and distribution of these products

Typical examples of material that can be obtained from the Internet in this way are

- Computer software, vedeos, music, photographs, among others
- Books in many different subject areas
- Complete Bibles in different versions and languages
- Bible helps, commentaries, books, sermons, etc
- Hymns, songs, music and similar material

5.4.4 Internet Broadcasting

Internet broadcasting, or transfer of *Streaming media, is the real-time delivery of digital media for the immediate consumption or enjoyment by end users. Many radio and television broadcasters provide Internet feeds of their live audio and video productions. They may also allow time-shift viewing or listening such as Preview, Classic Clips and Listen Again features. These providers have been joined by a range of pure Internet "broadcasters" who never had on-air licenses. This means that an Internet-connected device, such as a computer or something more specific, can be used to access on-line media in much the same way as was previously possible only*

with a television or radio receiver. The range of available types of content is much wider, from specialized technical webcasts to on-demand popular multimedia services. (Wikipedia, 2011).

Using this technology, many churches and Christian ministries are able to transmit their church Services, retreats, bible studies, special teachings, etc live to their members throughout the world at very low costs.

5.4.5 Social Networking

The ICT revolution has introduced an explosion into the age-long practice of maintaining academic, cultural, social, business or other links with partners and associates around the world. This has recently been systematized into social networks maintained principally via a network of computers. Principally, a given website is developed and maintained by the promoters, while individuals, groups and organizations register as members. A minimal set of rules are observed by the members to initiate or terminate their membership of the network, while the promoters endeavour to maintain confidentaility and privacy of the information held by them.

According to Wikipedia (2013), as at 7 September 2013, the largest social network is the **Facebook**. It was founded in 2004, and has over 1 billion registered users. This is followed by **Twitter**, with over 500 million registered members. Other major social networks together with their dates of establishment are shown in table 6. A more comprehensive list may be found in Wikipedia (2013).

Conclusion

Acknowledgements

References

1. Baiyeri K. P. (2013), *Commonly used software for data analysis and visualization in experimental agriculture*, Chapter 6 in Enibe S O (Ed) **MATLAB/SCILAB for Scientific & Engineering Systems Analysis**

2. Chukwu W. I. E. (2013) *Design of experiments for efficient data analysis/visualization*. Chapter 5 in Enibe S O (Ed) **MATLAB/SCILAB for Scientific & Engineering Systems Analysis**

3. Enibe S. O. (2013a) Introduction. Chapter 1 in Enibe S O (Ed) **MATLAB/SCILAB for Scientific & Engineering Systems Analysis**

4. Enibe S. O. (2013b) *Data/systems analysis with MATLAB/SCILAB*, Chapter 9 in Enibe S O (Ed) **MATLAB/SCILAB for Scientific & Engineering Systems Analysis**

5. Liengme B. V. (2009), **A guide to Microsoft Excel 2007 for Scientists and Engineers**, Academic Press, London WC1X 8RR, England

6. Oranu P. C. (2013) *Benefits of open source software in tertiary education*, Chapter 2 in Enibe S O (Ed) **MATLAB/SCILAB for Scientific & Engineering Systems Analysis**

7. Osadebe P. O. (2013) *Application of data analysis/data visualization in Pharmaceutical Research and Development*, Chapter 7 in Enibe S O (Ed) **MATLAB/SCILAB for Scientific & Engineering Systems Analysis**

8. Oyesanya M. O. (2013), *Available software for data analysis and visualization in the Physical Sciences*, Chapter 4 in in Enibe S O (Ed) **MATLAB/SCILAB for Scientific & Engineering Systems Analysis**

9. Soni, Amit (2008) *Ana analysis of Scientific Computing Environments: A Consumer's View.*, MSc Thesis, Massachussets Institute of Technology, Boston, Massachussets, USA

10. Wikipedia (2011) Internet Available online at `http://en.wikipedia.org/wiki/Internet` (Retrieved 22 March 2011)

11. Wikipedia (2013) List of social networking websites. Available online at `http://en.wikipedia.org/wiki/Social_network_service`. Retrived on 7 September 2013

Thank You Thank you very much for listening. God bless you richly.

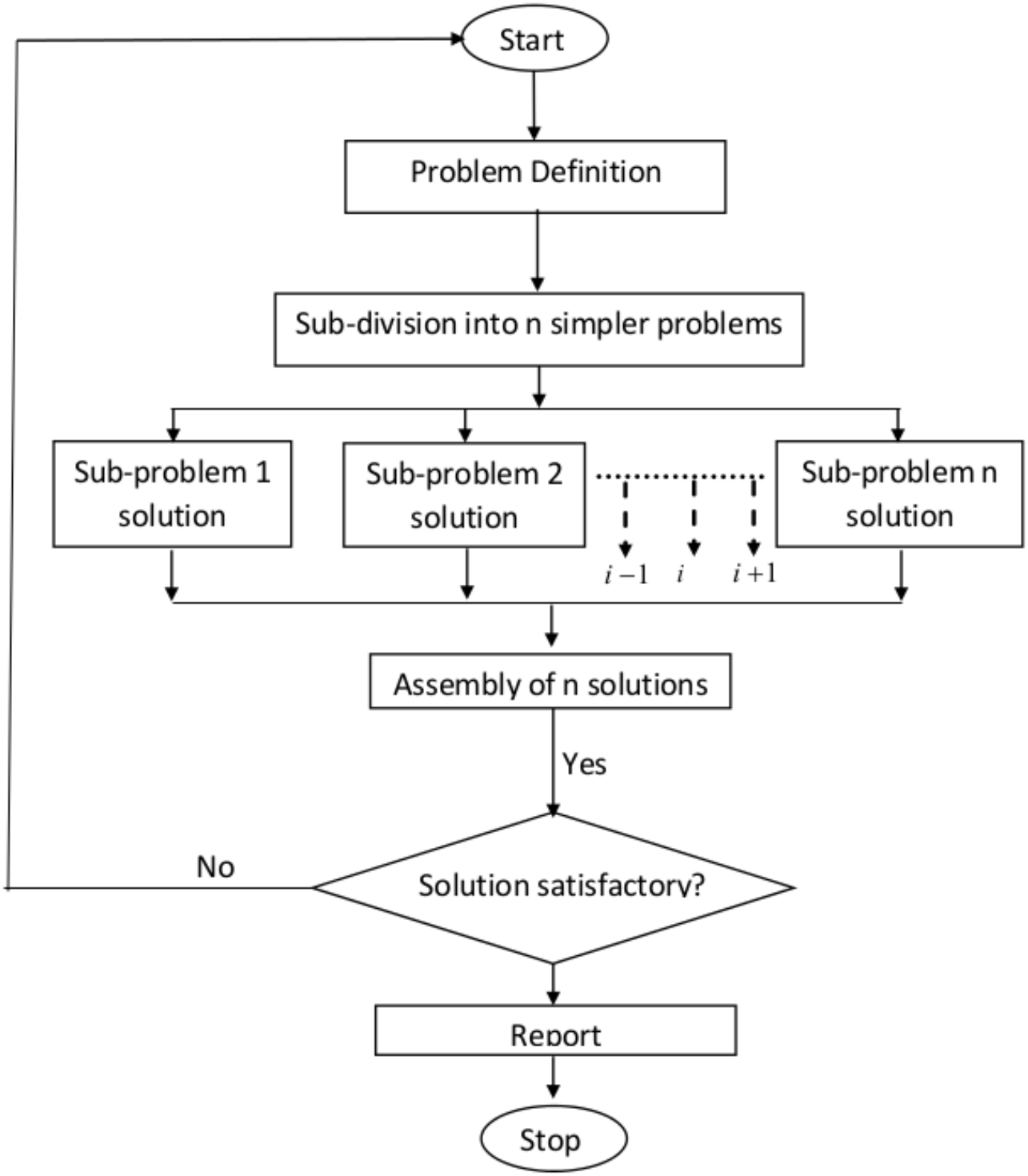

Figure 1: Chart of major steps in research

Figure 2: Steps for the solution of a simple problem

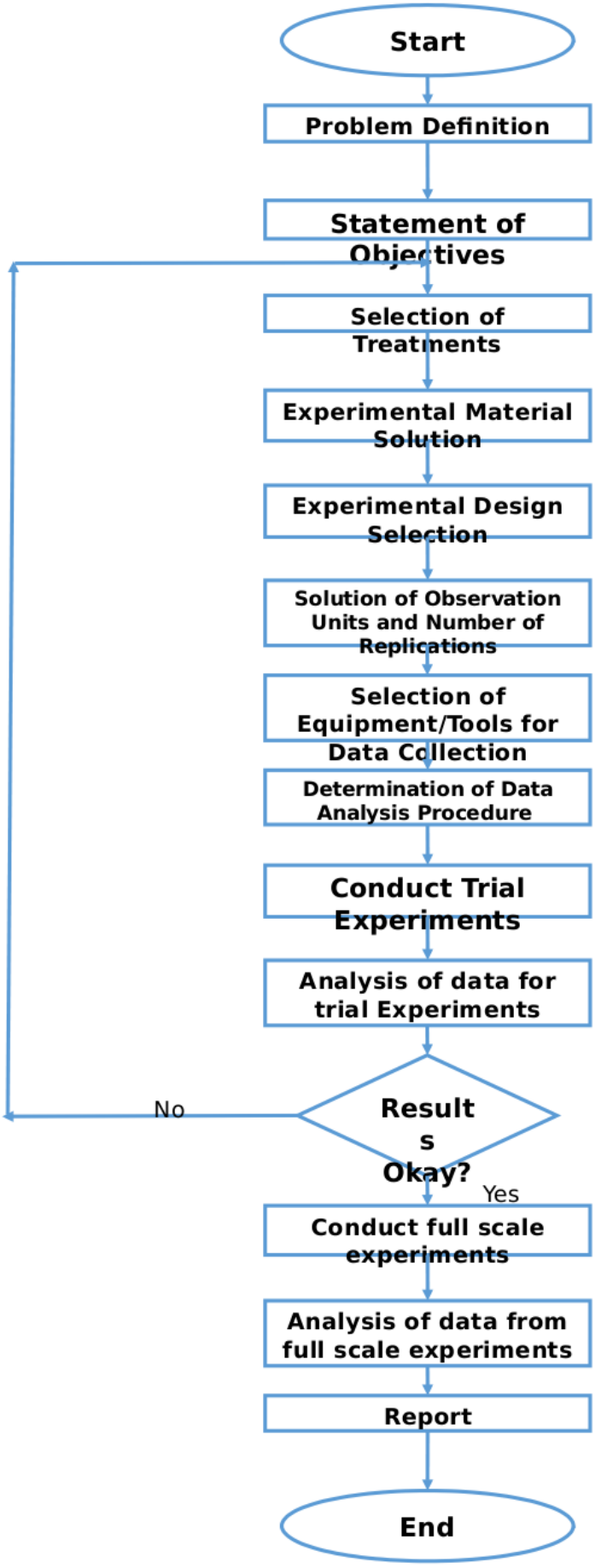

Figure 3: Recommended experimental procedure

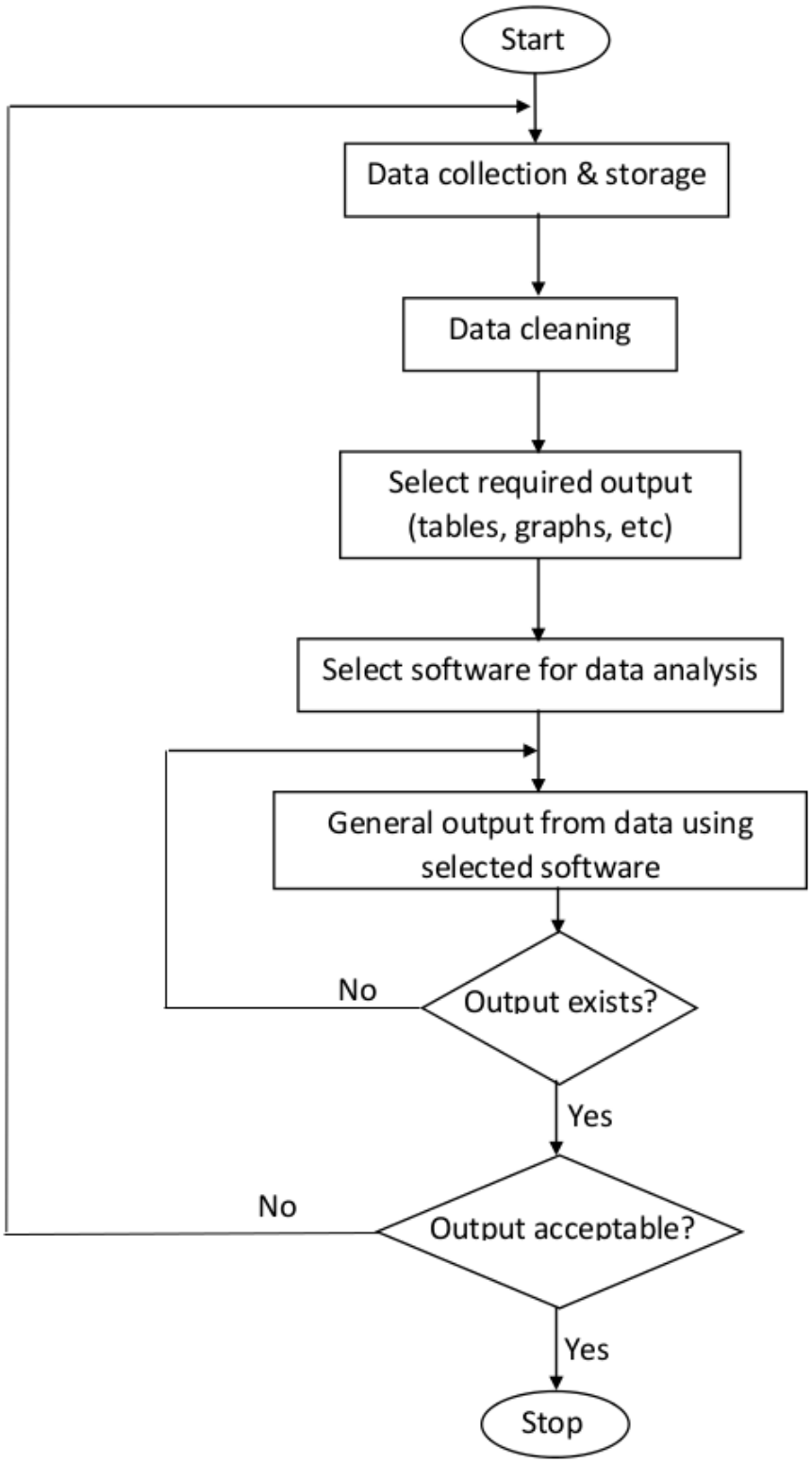

Figure 4: Data analysis steps

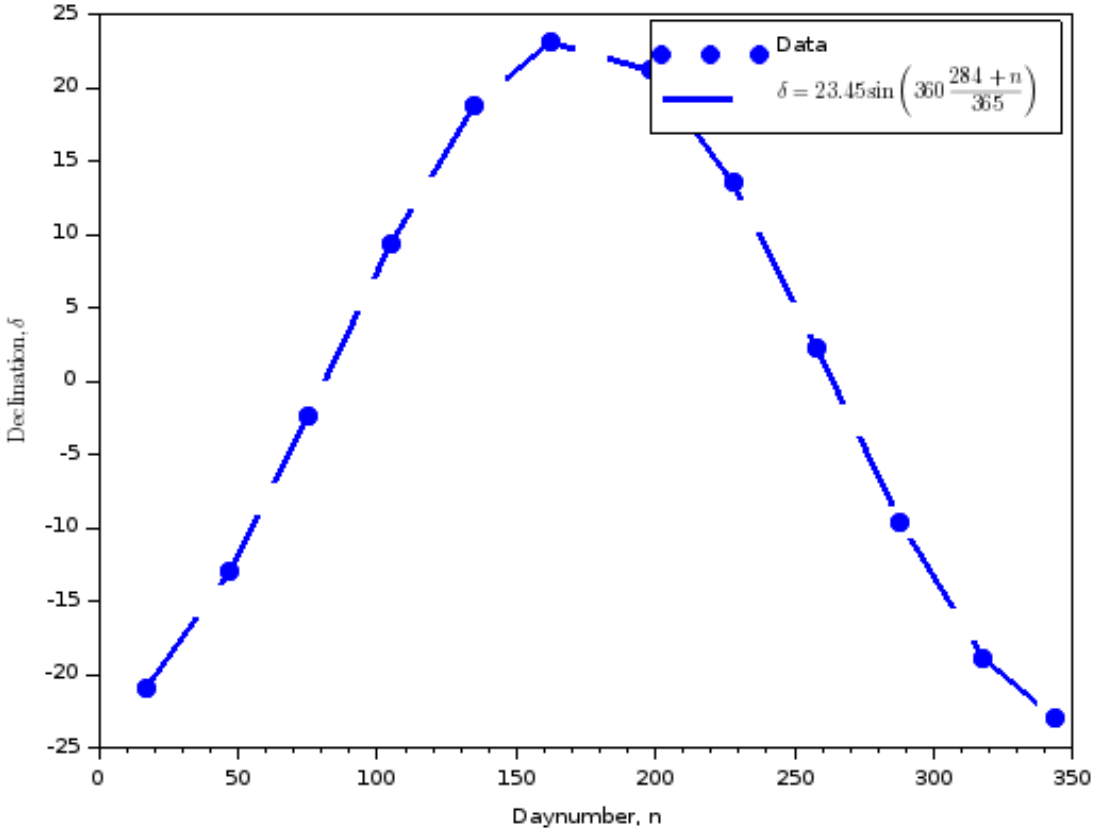

Figure 5: Graph of solar declination with daynumber

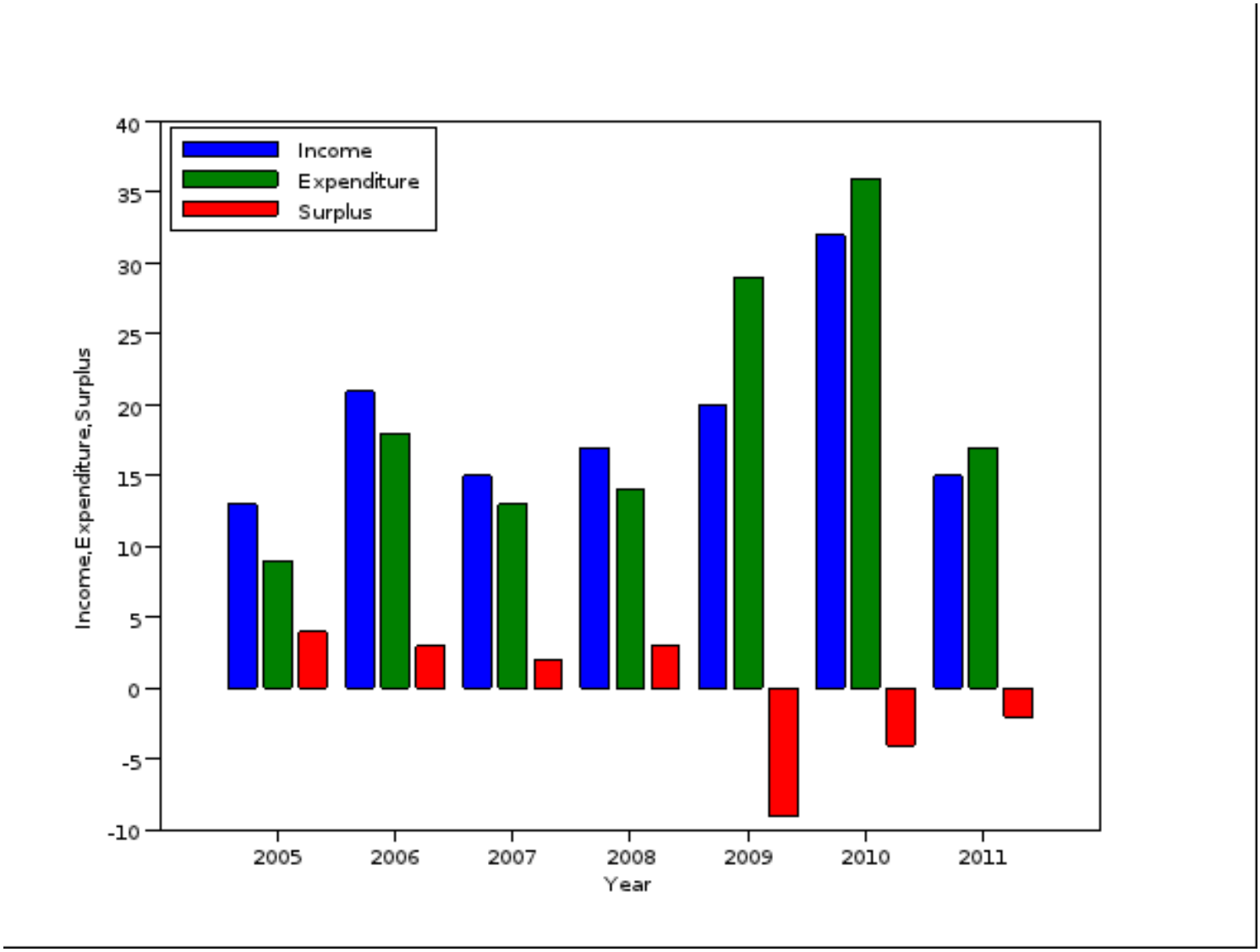

Figure 6: An example bar chart produced from the income and expenditure data in table 2

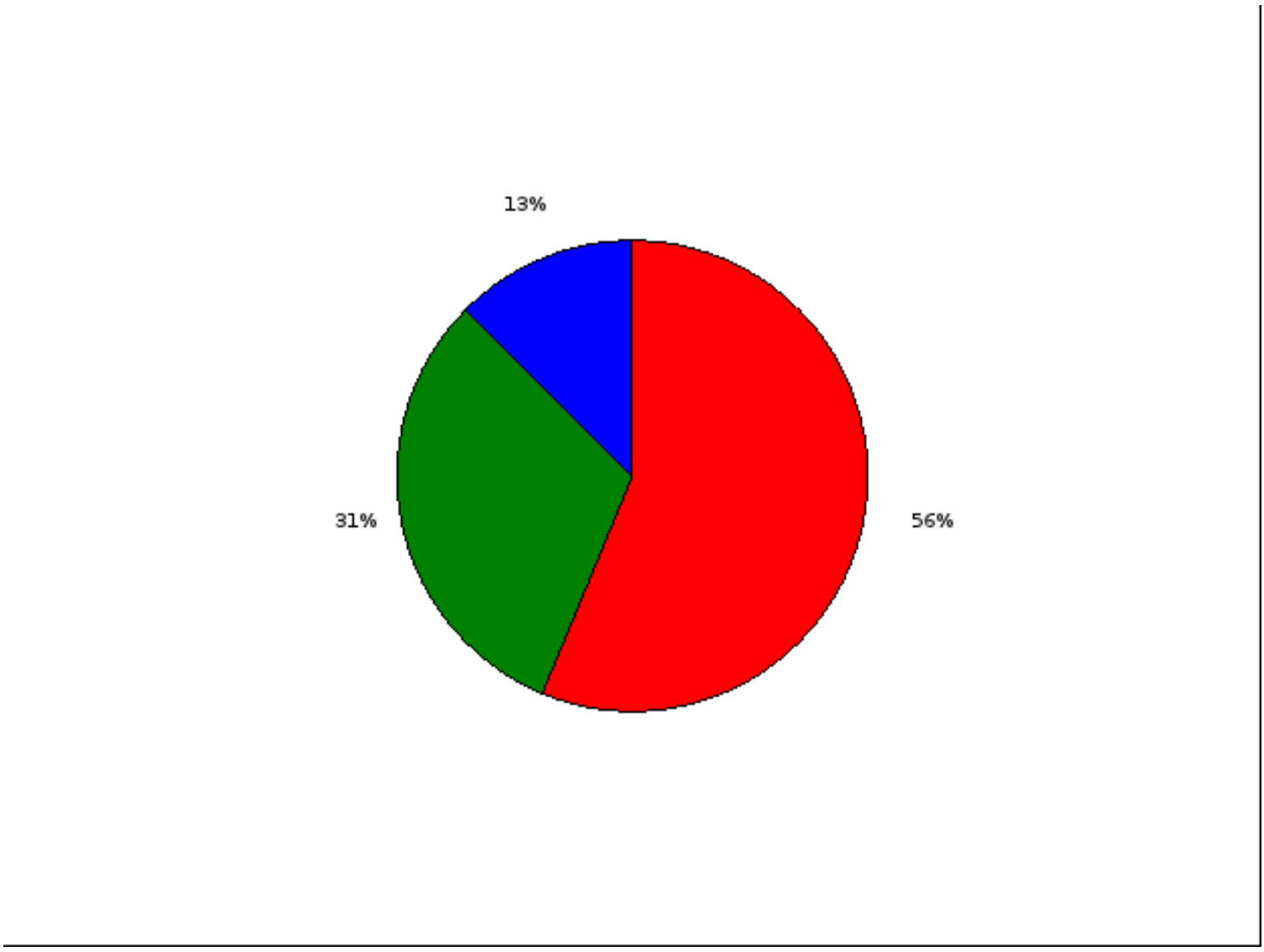

Figure 7: Pie chart using the data of table 3

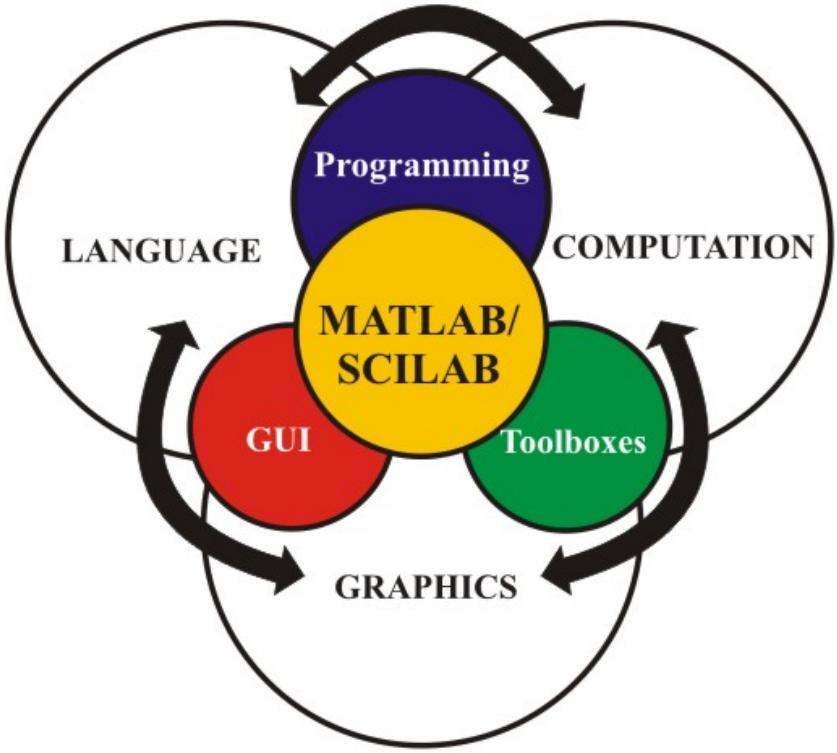

Figure 8: MATLAB/SCILAB features
Source: Enibe (2013)

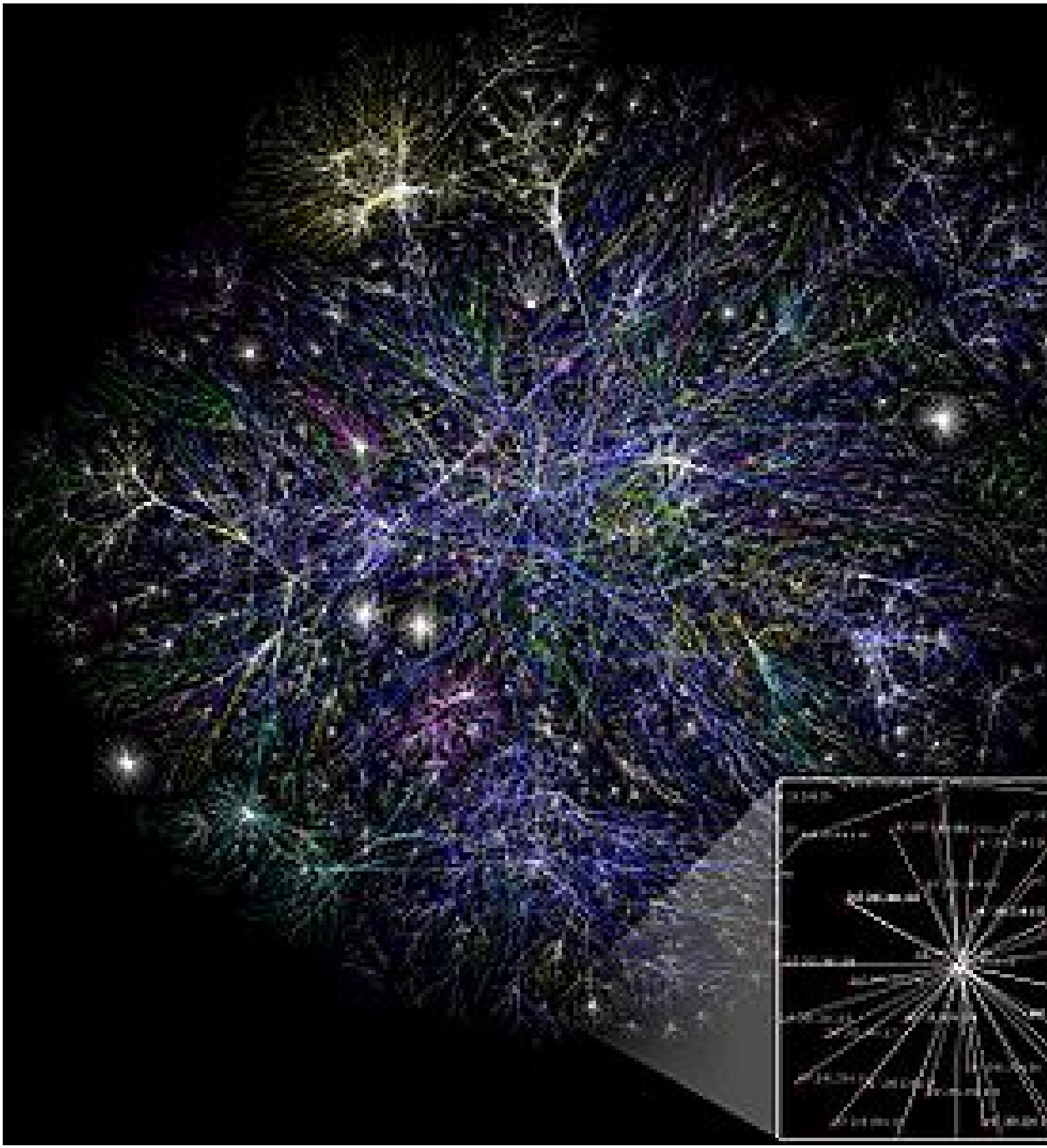

Figure 9: A visual representation of various routs through a portion of the Internet
Source: Wikipedia (2011)

Table 6: Major social networks

Name	Description/Focus	Date launched	Registered users	Registrati
Facebook	General: Photos, Videos, Blogs, Apps.	February 2004	1,000,000,000[80]	Open to peopl and older
Twitter	General. Micro-blogging, RSS, updates	15 July 2006	500,000,000[299]	Open to all ag [300]
Google+	General	28 June 2011	500,000,000[135]	Open to peopl and older [136
Qzone	General. In Simplified Chinese; caters for mainland China users		480,000,000[242][243]	Open to the ge public
Sina Weibo	Social microblogging site in Mainland China.	14 August 2009	300,000,000[256]	Open
Formspring	social Q&A website	November 2009	290,000,000[98]	Registration p responses and questions

Source: Wikipedia (2013)