Jurnal
Teknologi

Full Paper

Development of Web-Based Application for GIS Data Format Coordinate System Conversion

Article History
Received
14 May 2018

Elysonia Alim

Bachelor of Geoinformatics Science,
Faculty of Geoinformation and Real Estate,
University of Technology, Johor, Malaysia.

elysoniaalim@gmail.com

Abstract

It is difficult to process GIS vector data when they are not aligned with one another. The need for different coordinate systems rose from the fact that some coordinate systems are better fitted to describe the phenomenon happening in a specific area. The purpose of this study is to develop a web application capable of converting the coordinate system of a GIS data format such as a shapefile for Peninsular Malaysia. The web application named Coordinate Conversion Application (CCA) was developed using Django and Python and is capable of 5 coordinate transformations namely WGS84 to GDM2000 and vice versa, WGS84 to MRSO (old), MRSO (old) to Cassini (old) and vice versa. Results obtained were compared with existing software such as GDTS and ArcGIS, and analysis shows that CCA has achieved satisfactory accuracy.

1.0 INTRODUCTION

Coordinate systems are important in GIS as for different layers to be used together, they must align spatially [1]. Before the arrival of GNSS, coordinate reference systems around the world have been using classical coordinate reference systems that uses a local datum and span a particular region only. Following advancements in GPS, countries all over the world including Malaysia strove to adopt a global geocentric coordinate reference system [2]. This means a coordinate system with a common datum for all regions on the globe. One such system is WGS84 or the World Geodetic System 1984.

GIS users usually work with shapefiles. Shapefiles store geometrical and attribute information for the spatial features in a data set, where geometries are stored as shapes comprised of sets of vector coordinates (ESRI, 1998). Locations on the map features are represented by x- and y-coordinates, based on plane coordinate system [1]. For shapefiles to align, these coordinates must be reprojected into a common spatial reference system. Failure in doing so will cause inaccurate analysis results. It is important to define references for coordinates, because they are defined not by the law of nature but are artificially formulated.

GIS datasets can come in many different geocentric coordinate system, be it a free online GIS data or surveyed data. Depending on the goal of the data user or GIS analyst, the original coordinate system may be kept as it is. However, for region-specific analysis or mapping, it is preferable to use the local geocentric datum (i.e. GDM2000 for Malaysia), since it fits the geoid of the region better [3].

1.1 Problem Statement

There are commercial GIS software that can convert coordinate systems into another but the accuracy of the converted coordinates are questionable. Moreover, they use generalized formulas in their conversion algorithm and are different from JUPEM standards, which Malaysia uses. MapInfo and ArcGIS are two software that is commonly used to convert coordinates. Based on the study by [4], who studied the difference between the level of accuracy in georeferencing using MapInfo and ArcGIS, found that both showed error of around 0.1 m compared to the input point coordinate differences obtained from DCDB. This error also affected the processes that was done afterwards such as area calculation. Therefore, it can be said that even commercial software are not completely error-free. [4] also compared the gap magnitude between three layers of lot parcel. To do so, lot parcels from DCDB, MapInfo and ArcGIS were overlapped and the gap magnitude between them was measured. The result was that ArcGIS yielded a smaller gap compared to MapInfo. The smaller the gap magnitude between the raster dataset and the reference dataset yields better alignment result [4]. In relation to the focus of this research, spatial reference system affects the gap magnitude of datasets hence the alignment as well.

ArcMap, which is a part of the ArcGIS 8.8 suite for geospatial data processing, has no tool to find or calculate datum transformation parameters from measured or observed control data [5]. The study conducted by [5] had an objective of finding a reliable set of datum transformation parameters by comparing old and new topographic maps. While the results for Netherlands were satisfactory, when the same method was applied to find the datum shift parameters for regions in Mozambique, it was found that the datum transformation parameters between the local system and GPS measurement defined in the WGS84 system weren't accurate enough. Therefore, the inaccuracies in coordinate conversion may not be due to difference in formula, it is because of the difference in predefined parameters. In the same paper, [5] stated that parameters of local datums provided by the national survey are becoming more precise, and that they are also subject to updates once in a while. Hence, [5] developed a custom VBA program to counter the absence of user-defined datum transformation using ArcObject. This supports the purpose of this research, which is to develop an application capable of converting coordinate systems using parameters defined by JUPEM, which is the government body responsible for monitoring changes in datum origins for Malaysia, rather than relying on current software's parameters that are most likely haven't been updated according to current datum conditions that has changed due to disasters such as big earthquakes.

2.0 Literature Review

2.1 Background to projections, datums and coordinate systems

Coordinate systems are man-made as it is based on centuries of accumulated knowledge. Thus, before moving onto specific coordinate systems, it is appropriate to first examine its origin.

2.1.1 Reference Surfaces

According to [6], geodesy is the study of the size and shape of the Earth; the measurement of the position and motion of points on the Earth's surface; and the configuration and area of large portions of the Earth's surface whereas the general definition from Merriam-Webster is "a branch of applied mathematics concerned with the determination of the size and shape of the earth and the exact positions of points on its surface and with the description of variations of its gravity field". So, it can be said that geodesy deals with the geometry of the Earth, including characteristics of everything in and on it, and that it uses mathematic formulas in order to describe those geometries and other characteristics.

There are 4 different ways we can view Earth's shape in geodesy, or reference surfaces, according to [6]. The first is the terrestrial surface which is Earth's topography, or the terrain which we experience on Earth. It is uneven and complex which makes it unsuitable as reference for precise mathematical calculation. The second reference surface is the geoid. The geoid coincides with the mean sea level (MSL) and its equipotential quality makes it suitable as a reference surface for heights. Due to the force of the Earth's rotation, the shape of the Earth has been found to be more oval than spherical: "squashed" at the poles and bulges in the equator [6]. The third one is the ellipsoid, which is a mathematical approximation of the geoid that allows precise geometry determination on the surface of the Earth. The fourth and final way is the sphere, which is merely a simplified ellipsoid, defined by only a radius. When the spherical model of the Earth is used, the radius is 6378137.00 meters. Figure 1 illustrates the reference surfaces.

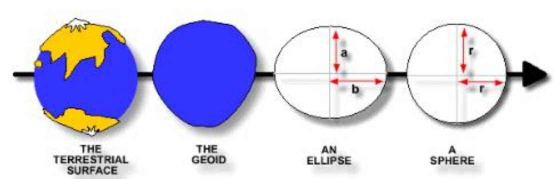

Figure 1: Four Different Ways of describing Earth's Shape

(Source: [6])

In the matter of precise coordinate determination, the two most instrumental reference surfaces would be the Geoid and the Ellipsoid. Figure 2 shows the shape of the geoid and ellipsoid compared to the terrain and mean sea level (MSL).

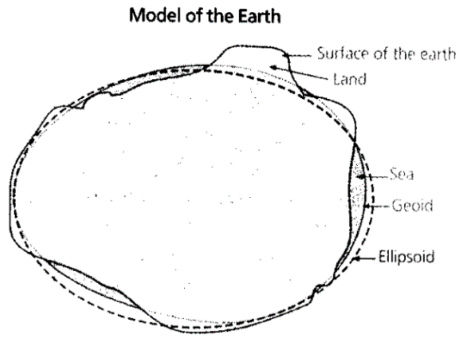

Figure 2: Model of Earth

(Source: [7])

The geoid is a closed, continuous surface and its curvature shows discontinuities abrupt density variations. It is an undulating surface, due to Earth's geological evolution which caused irregular distribution of crustal rocks having different densities [8]. This surface is established utilizing the mean sea level, assuming that the waters of the ocean is a free-moving entity influenced by the force of Earth's gravity. At the point of equal force, these waters align with the gravity field, hence the conception of the geoid. It takes averaged ocean water levels registered over intervals of more than a year to estimate the geoid. Therefore, MSL represents an approximation to the geoid. While it is far from being a mathematical surface hence unsuitable for position computations, it makes a good reference for heights defined in the gravity field.

The ellipsoid closely resembles the natural shape of the Earth as it rotates about its axis, around the Sun, and the ellipsoid of rotation can also be called a spheroid. Hence, it is commonly used to as a base define the geometry of datums. These datums are more regularly used than Cartesian coordinate systems. The importance of using an ellipsoid or spheroid to represent the shape of the Earth is that while the sphere model may be appropriate for smaller scaled maps (e.g. 1: 5,000,000) as there is no detectable difference between a sphere and an ellipsoid, for larger scale maps (e.g. 1: 1,000,000) which show more detail it is necessary to use ellipsoids to represent the shape of the Earth [6].

2.1.2 Ellipsoids: Terrestrial Reference Frames

ITRF (International Terrestrial Reference Frame) is a definition of geocentric system adopted and maintained by IERS (International Earth Rotation and Reference Systems Service). The ITRF is a set of points observed by VLBI, LLR, GPS, SLR and DORIS with their 3D Cartesian coordinates and velocities, that realizes an ideal reference system, the International Terrestrial Reference Frame. Updated almost every year, and is still the best reference system currently [9].

WGS72 (World Geodetic System 1972) was the third geocentric reference frame developed by the United States of Defense Mapping Agency (DMA) to support its activities. It was used until 27 January 1987, with the GPS system and before 27 January 1989 it was used for the Transit Doppler navigation system broadcast ephemeris. Table 1 describes the WGS72 parameters.

Table 1:WGS72 Ellipsoid Parameters

Semi Major Axis (a)	6,378,135 meters
Inverse Flattening (1/f)	298.26

WGS84 (World Geodetic System 1984) is the default coordinate system for most GPS receivers used in the world today. The United States National Geospatial-Intelligence Agency (NGA) defines and maintains this datum. Since its establishment, it has been revised several times and is presently accurate at the centimeter level to the ITRF, which mainly guarantee scientific integrity and compatibility with international standards and conventions. WGS84 coordinates of tracking stations for computing GPS broadcast orbits are annually adjusted for plate tectonic motion to an epoch at the half year mark. For most users, the differences between the ITRF and WGS84 are negligible.

According to [10], in the manual written by the Intergovernmental Committee on Surveying and Mapping, the WGS84 definition include the following items:

- WGS84 Cartesian axes and ellipsoid are geocentric; meaning their origin is the center of mass of the whole Earth including oceans and atmosphere.
- Their orientation (the directions of the axes; orientation of the ellipsoid equator and prime

meridian of zero longitude) coincided with the equator and prime meridian on the midnight of New Year's Eve 1983.

- Since 1984 at January 1st, when the time was precisely at 00:00, orientation of the axes and ellipsoid has changed such that the average motion of the crustal plates relative to the ellipsoid is zero. This ensures that the Z-axis of the WGS84 datum coincides with the International Reference Pole, and that the prime meridian of the ellipsoid (plane with Z and X Cartesian axes) coincides with the International Reference Meridian.
- The shape and size of the WGS84 biaxial ellipsoid is defined by the semi-major axis length a = 6378137.0 meters, and the reciprocal of flattening.
- Conventional values are adopted for the standard angular velocity of the Earth for time measurement, and for the Earth gravitational constant to define relative scale.

Correct and consistent geodetic datums important for work involving precise locations such as highway constructions, cadastral survey and coastal inundation planning. To get points in different datums to align, there are three steps that must be followed:

1. Origin Alignment: Positioning the coordinates into close proximity.
2. Axis Rotation: Rotating the axis until the coordinates align.
3. Scale Factor: Scaling axes until the coordinates match.

The ellipsoid GRS80 (Geodetic Reference System 1980) is recommended by the International Association of Geodesy (IAG) to use as the reference frame for datums [11]. In the same report, it is said that the parameters of the WGS84 ellipsoid "*…are identical to those for the GRS80 ellipsoid with one minor exception. The coefficient form used for the second degree zonal is that of the WGS84 Earth Gravitational Model rather than the notation J2 used with GRS80*." The only difference between WGS84 and GRS80 is the value of its inverse flattening; but in most cases, it is negligible. Thus, it can be said that WGS84 and GRS80 are the same.

2.1.3 Coordinate Systems

A coordinate system may be used and set up using three concepts, according to Ordnance Survey's "A Guide to Coordinate Systems in Great Britain", which are datum, datum realization and type of coordinates. Table 2, taken from the same document, summarizes these concepts.

Table 2: Coordinate System Concepts (Source: [10])

Coordinate system concept	Alternative name	Role in positioning
Datum	Terrestrial Reference System (TRS)	The set of parameters which defines the coordinate system and states its position with respect to the Earth's surface
Datum realization	Terrestrial Reference Frame (TRF)	The infrastructure of 'known points' that makes the coordinate system accessible to users
Type of coordinates		The way we describe positions in the coordinate system

Generally, a coordinate is a set of at least 2 numbers specifying the position of a point, line or other geometric figures based on a reference system. There are only two types of coordinate systems which are geographic and projected coordinate systems. A geographic coordinate system (GCS) use the latitude and longitude coordinates on a spherical model of the Earth's surface while a projected coordinate system (PCS) use mathematical formulas to transform GCS which is 3D into planimetric 2D surfaces and is usually expressed in Eastings and Northings. Spatial features on the surface of the Earth may be represented using one of these two coordinate systems.

2.2 Background on Malaysia Coordinate Systems

The realization of classical and modern datums are different in methods and concept. Classical datums were established using traditional survey techniques and procedures, and local datums had limited coverage but was adequate for national mapping requirements [2]. A classical datum is a model of the earth because it is defined by the position, orientation, size and shape of the reference ellipsoid used to define latitudes and longitudes by projecting points to it from the Earth's surface [12]. Different datums and reference ellipsoids suit different parts of the globe, and at the time the technology to establish a truly geocentric datum had not yet been developed. Therefore, maps were hard to align because each region used a different origin to determine their coordinates. The incompatibility of classical datums with one another made it hard for maps in projected old coordinate systems to be used to make a world map that was accurate and reliable.

[6] mentioned in his writing that "historic" data is data acquired before the conception of the Australian geocentric datum. Therefore, "classical" datum in this case can be considered as datum before the emergence of geocentric satellite positioning. WGS84 is a widely used modern datum or coordinate system, as clarified in the previous section. Because of a minor difference in the value of its inverse flattening, the recommended ellipsoid to use as base in local datum formulation is the GRS80. It is used as the reference ellipsoid for Malaysia's geocentric datum GDM2000 as well as many other national datums around the world. Among the countries using GRS80 as their reference ellipsoid are Japan and Australia.

Work done in the 1880s was instrumental in initiating trigonometrical works in Malaya which led to trigonometrical survey across Malaya that laid the foundation for the existing control framework. However, these traditional coordinate systems were inconsistent and unconvincing in terms of reliability that it had to be revised in order to comply with the modern standards at the time [2]. This is why Malayan Revised Triangulation 1968 (MRT68) for Peninsular Malaysia was established. MRT68 was to replace the old Primary or Repsold Triangulation for Malaya. The work to establish Sabah's primary triangulation work began with a project known as the Borneo West Coast Triangulation (1930 – 1942) whereas for Sarawak and Brunei, around 1935 [2]. The task to readjust the primary triangulation of Borneo was undertook by the Directorate of Overseas Survey (DOS), who later commenced the East Coast Triangulation project that was the basis for the Borneo Triangulation 1968 (BT68). The origins of MRT68 and BT68 are Kertau, Pahang and Timbalai, Labuan respectively.

The reference ellipsoid for these two triangulation networks is Modified Everest or Everest 1830 Modified, adopted in 1967 for use in West Malaysia. The EPSG Geodetic Parameter Registry defines its semi-major axis to be 6377304.063 meters and its inverse flattening as 300.8017 unity. It applied Benoit 1898 inch-meter ratio of 39.370113 to the original definition of the semi-major axis and 1/f of Everest 1830 but with the semi-major axis taken to be in British rather than Indian feet (Source: [13]). This means that the values of the parameters were the same but used the conversion factor of 39.370113 inch = 1 meter to alternate between inch and meter measurements while assuming that the original parameters are in British imperial measurement system. The original ellipsoid, Everest 1830 (1937 Adjustment) was used for the readjustment of the Indian Triangulation. The conversion factors of Indian- British foot and British-Indian foot were rounded and applied to the ellipsoid definition. Therefore, taking the parameters of the original ellipsoid in British foot would not cause severe distortion that would result in coordinate misalignment.

[14] as mentioned in the paper presented at the Seminar on GDM2000 Kuala Lumpur, Malaysia, emphasizes the importance of providing a standardized or homogenous geodetic infrastructure as the basis for integration of spatial data for sustainable development decision making [12]. The regional nature of MRT68 and BT68 are not aligned with global geocentric coordinate frames such as WGS84, so JUPEM made effort to fully utilize space-based technology by establishing a GPS network of 238 stations in Peninsular Malaysia which was called the Peninsular Malaysia Geodetic Scientific Network 1994 (PMGSN94); and then in East Malaysia, where 171 GPS stations were established to form the East Malaysia Geodetic Scientific Network 1997 (EMGSN97) [2]. The origins for East and West Malaysia are still Timbalai, Labuan and Kertau, Pahang, respectively. The origins were used as fixed points to do minimally constrained adjustments on their respective geodetic networks.

2.3 Django Web Framework for Web Development

Django is a high-level Python Web framework that stimulates fast development and clean, pragmatic design. It was invented to meet deadlines while satisfying the needs of professional Web developers. It has its own modules that may be imported to use in the website but custom modules may also be used. It is free and open source, meaning the public may access its source code without paying. The main benefits of using Django to build a website with Python are [15]:

1. Speed: The Django framework is well suited for those who want to put up a website on the internet as quickly as possible. Therefore, for a project with a strict deadline such as this, using Django is time-saving and cost effective. Django also comes with excellent documentation which helps even the very beginners to understand its structure.
2. Feature rich: Django allows developers to access extra features such as sitemaps, content development and user authentication, as well as having GeoDjango, which give Django GIS capabilities.
3. Security: Django eliminates the need to worry about security resulting from common mistakes such as cross-site forgery and scripting.
4. Versatile: It can be used in various types and scale of web development, from small scale projects such as this as well as big organization web development.

According to The Django Book website in a webpage written by [16] there are a few solid reasons why Django is a good choice:

1. Its written in Python, a programming language that's considered the beginner's language with its natural-language constructs that's easy to learn and understand. Python is the most wanted language among developers, beating JavaScript for first place in 2016.
2. Batteries included – meaning Django comes with common but complex processes that may be implemented and encapsulated easily without compromising power. Django's "batteries" are located in the *contrib package*, which contain the module *GIS* that adds geospatial capabilities to Django, and *postgres* which contain PostgreSQL database specific procedures.
3. Solid foundation – unlike its competitor Node.JS, Django provides everything to build an application upon from the start, resulting in a better awareness of what is in the application.

According to a comparison made between Django and Node.JS, which is a platform built on Chrome's JavaScript runtime, in the website [17] which is a website dedicated to comparisons, Django trumps over Node.JS in terms of being more beginner-friendly as it includes everything from libraries to framework for the whole process of website development, instead of having to pick another framework. Furthermore, Django supports a wider range of databases that are commonly used in GIS such as PostgreSQL, MySQL, Oracle, Apache and Microsoft SQL Server 2005. Another reason why Django is chosen for this project is because it is highly documented. All across the internet, it is far easier to find tutorials, Ebooks and forums on Django as it has been around much longer than Node.JS. Node.JS is an excellent choice for advanced programmers with plenty of experience in JavaScript, or doing a project without strict time constraints, both of which this writer does not have.

Django employs the model-view-controller client/server code reuse which involve the model layer, templates layer and views layer. Since this project involves geospatial data input and output, the forms layer and GeoDjango (GIS functionality for Django, accessed through the *contrib.gis* package) is also important. Models contain the data definition of each field in a database, specifying its data type and length of character. Views contain functions that control what the end user sees and interacts with. It may also contain functions that process the data or even make calculations, but it can also be used to connect to a custom module in the same repository that performs the processing or calculation work. Forms handle the processes for data input and output, which includes reading, writing and saving data. The Geographic Framework, or GeoDjango is a package inside Django that comprises of geospatial data handling and processing. However, GeoDjango will not be used to build this application for it's too complex and hard to implement for shapefiles with different table structures.

The general workflow to start coding in Django comprises of these steps:

1. Creating a Django Project
2. Running the development server
3. Creating an application
4. "Installing" or adding the application in settings.py
5. Modifying urls.py in the project folder.
6. Adding urls.py in the application folder.
7. Controlling what the user sees with views.py
8. Templating.
9. Creating models (for data-driven applications).

3.0 Methodology

This section will explain the steps taken that was used in this project, or methodology. There are five phases to this project, that was followed from start to end. They generally follow the GIS project management workflow but with minor differences to fit the kind of project this research is about. Each phase is accompanied by a brief description of the activities that was done in those phases. In this project, the methodology chosen is one fit for system development. The SDLC methodology comprises of the main stages related to one another through succession and regression. A stage must be completed before proceeding onto the next one, while examining each and every requirement for advancement, which may lead to the process being repeated or pushed back one stage in the case of unsatisfied prerequisites.

3.1 Research Approach

For this project, the methodology is comprised of five phases which are the preliminary study, planning, implementation, validation and conclusion. The preliminary study consists of the initial research to formulate the problem, which requires plenty of reading. This phase also includes further research on the implementation method and the focus of the study, for better understanding of the topic and more comprehensible general writing direction. The planning stage is the most important stage of them all, since it includes the designing and identification of work that must be done in order to achieve project objectives. The implementation stage comprises of tasks described in the planning stage, but more involved and practical rather than theoretical. For this project, testing is important and will be done along with the coding process. The analysis stage is critical in determining the success or failure of the project. In this stage, comparisons between results from different software are made in order to analyse the difference. The conclusion consists of statements made that validate or not validate the success of the project at completion according to objectives achieved. Recommendations for improvement are also given in this stage.

Figure 3 shows the research approach for this project.

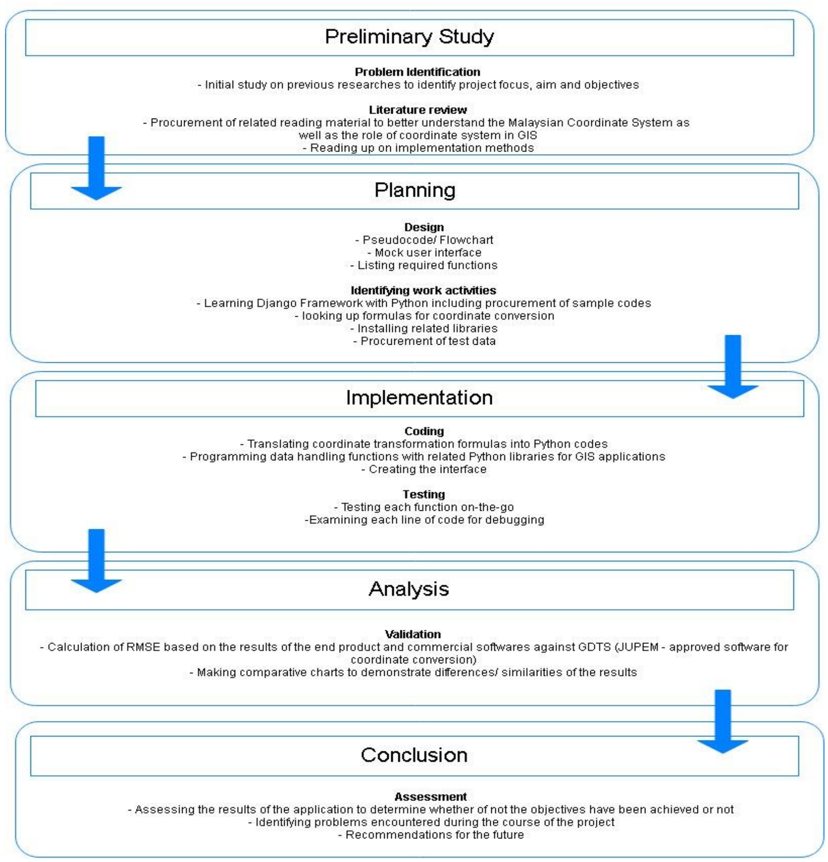

Figure 3: Research Approach

3.2 Development of Coordinate Conversion Application

One of the methods that can be used to plan a project is the System Development and Life Cycle (SDLC). SDLC is a conceptual model that is used in system development. It gives the project the opportunity to combine two or more project methods for the bet outcome. For this project, the methodology used is illustrated in Figure 4.

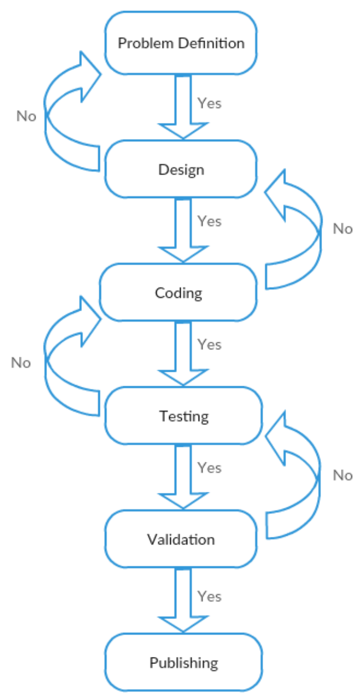

Figure 4: System Development and Life Cycle

3.2.1 Problem Definition

This stage involves defining the problem and identifying activities required to accomplish the goals. Chapter 1: Problem Statement and Objectives has accomplished the first task which is identifying the problem and determining the goal and activities involved in this project. According to [18], there are three components that must be prioritized which are:

1. The input; for this project, it would be an ESRI Shapefile already in the WGS84 system.
2. The processes; coordinate transformation which are the formulas described in the Section 2.3 and data handling functions incorporating the Django framework for Python Web Programming.
3. The output; an ESRI Shapefile that is either WGS84, GDM2000, RSO or Cassini coordinate systems that is saved in a different file from the input.

3.2.2 Design

The next step in application development is the design. This stage includes designing a solution algorithm which uses the user requirements defined in the problem definition stage and 3 control structures that are sequence control, selection and repetition to produce a logical structure that is always correct, irrespective of programming language. A clear and concise way to represent an algorithm is by creating a flowchart.

The general process of the application would start from the user choosing whether they want to do a Datum Transformation or Map Projection. In Datum Transformation, they would have to choose whether to convert from WGS84 to GDM2000 or vice versa, whereas in Map Projection, they would have to choose either WGS84 to MRSO (old) and MRSO(old) to Cassini (old) (and vice-versa). Overall, the user would have 5 conversions to choose from. After that, the user would input all the files of a shapefile so it may be saved in a virtual folder for easy processing. Once saved in the virtual folder, the coordinate transformation takes over by firstly exporting it into GeoJSON format, extracting its coordinates and run it through transformation formulas and then rewriting it in GeoJSON. When all points have been converted, the shapefile writer function takes over and writes a shapefile based on the GeoJSON data. The new shapefile will be saved in a virtual folder before being turned into a .zip file. A download link for the .zip file will then be generated for the user to download the new shapefile.

The interface of this application would be direct and simple to minimise confusion and maximise comprehension. It would only consist of forms to choose coordinate conversion functions and forms to upload and download data.

Figure 5 shows the flowchart of the solution algorithm for the web application.

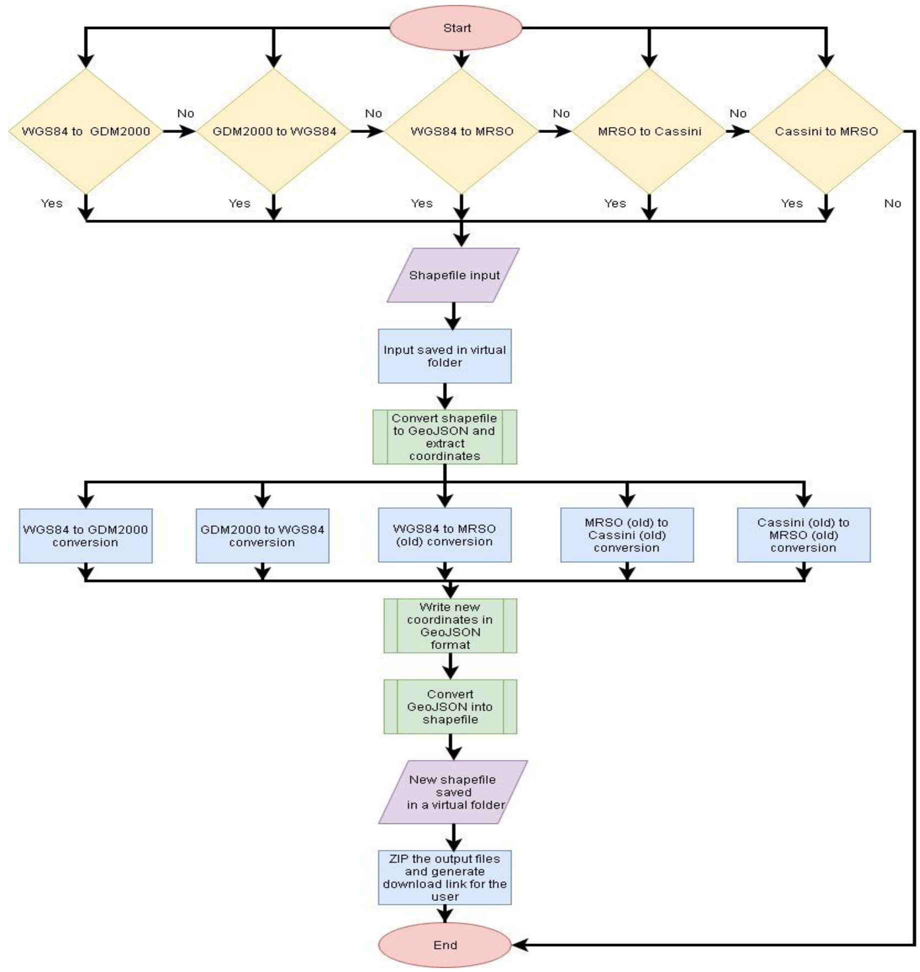

Figure 5: Application Flowchart

3.2.3 Coding

This part of the project follows the Django framework procedures strictly. At project creation, Django creates default modules that serves as framework to build a web application from at project initiation. After that, custom modules may be added for the specific purpose the application has. Figure 6 summarises the organization of files involved, their respective purposes and interaction.

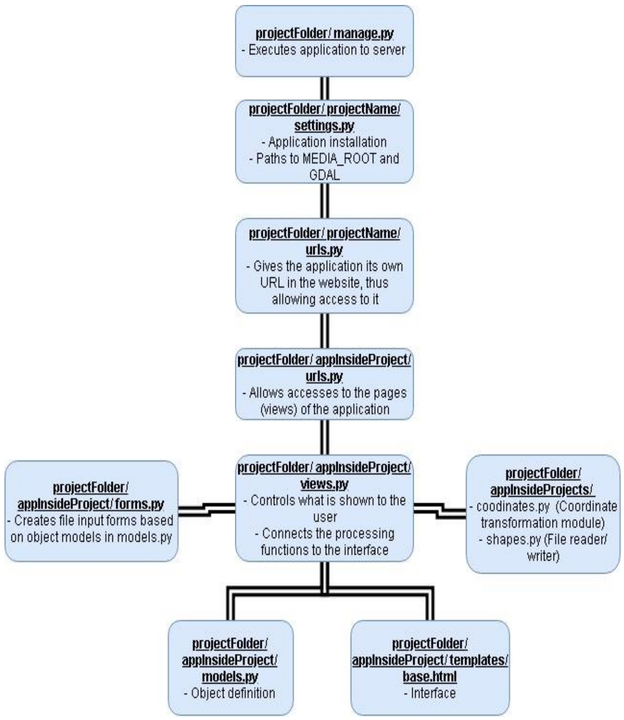

Figure 6: General File Structure and Interaction

3.2.4 Testing

There are a few methods available in software testing: Black-Box, White-Box and Grey-Box. The Black-Box testing is a method used when the tester or user is completely clueless about the encapsulated functions and architecture. In this method, only the interface is tested; the user provides input and validates output. White-Box testing involves detailed and thorough examination of the internal logic and structure of code. It is also called glass-testing or open-box testing and requires a knowledgeable tester as this method focuses on debugging. The Grey-Box testing is the intersection between the previous two methods, in which the tester has a limited knowledge of the program. Unlike the Black-Box method, the tester has access to design documents and the database so they can prepare test data and scenarios while making test plans. Therefore, in software testing, it is crucial for the tester to know as much as possible about the program. For this application, White-Box testing will be used.

The criteria to be considered in testing are as follows:

1. Coordinate transformation functions – the results must be the same as the test data.
   a. WGS84 to GDM2000
   b. GDM2000 to WGS84
   c. WGS84 to MRSO (old)
   d. MRSO (old) to Cassini (old)
   e. Cassini (old) to MRSO (old)

2. Data handling functions – it must be capable of the following:
   a. Accepting ESRI Shapefile input
   b. Passing data to the coordinate transformation functions for conversion
   c. Writing new coordinates while preserving original attribute data
   d. Saving new ESRI Shapefile as output

3. Control functions
   a. Sequencing processes from data input to process to output.
   b. Selecting which coordinate systems to convert from and to, and selecting which process to apply based on user's choice.
   c. Repetition of processes (looping) in converting each and individual coordinates.

3.2.5 Validation

The method chosen to validate the results of this project will be the RMSE method or Root Mean Square Error method. This method is a measure of the error around the regression line, the same way the standard deviation is the measure of how varied the data is around the mean. The formula for calculating RMSE is:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x - x_i)^2 + (y - y_i)^2}{n}}$$

Where,

x, y	=	Predicted Coordinates
x_i, y_i	=	Observed coordinates
n	=	Number of points

RMSE of the result from the coordinate conversion application will be compared with the RMSE of the result from a commercial software. Hooi (2011) mentions that for a JUPEM approved level of accuracy, the RMSE value must be less than 0.2.

The conversions that would have to be validated are:

a. WGS84 to GDM2000
b. GDM2000 to WGS84
c. WGS84 to MRSO (old)
d. MRSO (old) to Cassini (old)

e. Cassini (old) to MRSO (old)

4.0 Analysis and Results

As a final product, CCA is deployed through a private server. Before it is suitable for deployment, its results must be validated to evaluate the accuracy and whether or not it complies with the standards set by JUPEM. The analysis was done by comparing results obtained with CCA v2.0 and ArcGIS with GDTS v4.01 using the mathematical formula for RMSE described in 3.3.5. The result between CCA and GDTS showed that CCA is capable of achieving a similar accuracy level with GDTS.

4.1 Coordinate Transformation from WGS84 to GDM2000

A shapefile in WGS84 coordinate system was converted into GDM2000 using CCA v2.0 and a customised trasformation with ArcGIS 10.3 since it does not have the transformation from WGS84 to GDM2000. Meanwhile, a text file containing the coordinates of the features of the shapefile obtained by exporting its attribute table using ArcGIS 10.3 was used as the input data for GDTS v4.01. Results obtained from CCA v2.0 and ArcGIS were compared with results from GDTS v4.01. Table 3 depict the result.

Table 3: WGS84 to GDM2000

	GDTS – CCA		GDTS – ArcGIS (Custom)	
	Lat	Lon	Lat	Lon
Max. Error	2.42144×10^{-13}	2.46475×10^{-9}	2.3752×10^{-13}	2.46451×10^{-9}
Min. Error	1.87888×10^{-17}	1.28135×10^{-12}	8.38056×10^{-17}	1.27577×10^{-12}
RMSE	0.00003		0.00003	

Referring to Table 4.1, CCA had achieved an RMSE of 0.00003" which was similar to ArcGIS against GDTS. The maximum and minimum latitude error for CCA are 2.42144×10^{-13} and 1.87888×10^{-17} respectively whereas for ArcGIS it was 2.3752×10^{-13} and 8.38056×10^{-17}. For longitude, the maximum and minimum errors for CCA were 2.46475×10^{-9} and 1.28135×10^{-12} respectively whereas for ArcGIS it was 2.46451×10^{-9} and 1.27577×10^{-12} respectively. 0.00003" is within the acceptable margin of positional error. Both software achieved an RMSE of less than 0.2 which was the suggested error allowance by JUPEM. Hence, this conversion can be considered a success.

4.2 Coordinate Transformation from WGS84 to MRSO (old)

A shapefile in WGS84 coordinate system was projected into MRSO (old) using CCA v2.0 and ArcGIS 10.3. Its coordinates in text format was converted from WGS84 to MRT then MRSO with GDTS v4.01. The results of CCA v2.0 and ArcGIS were compared with GDTS and Table 4 shows the result.

Table 4: WGS84 to MRSO (old)

	GDTS – CCA		GDTS – ArcGIS	
	N	E	N	E
Max. Error	0.000363583	5.8128×10^{-7}	0.174436534	0.625394691
Min. Error	0.000327288	3.2185×10^{-12}	0.16202074	0.609678311
RMSE	0.018537576		0.885626497	

Referring to Table 4.2, CCA had achieved an RMSE of 0.018537576" which was much smaller than ArcGIS which was 0.885626497 against GDTS. The maximum and minimum Northing error for CCA are 0.000363583 and 0.000327288 respectively whereas for ArcGIS it was 0.174436534 and 0.16202074. For Easting, the maximum and minimum errors for CCA were 5.8128×10^{-7} and 3.2185×10^{-12} respectively whereas for ArcGIS it was 0.625394691 and 0.609678311 respectively. 0.018537576" is within the acceptable margin of positional error. Only CCA achieved an RMSE of less than 0.2 which was the suggested error allowance by JUPEM. Hence, this conversion can be considered a success.

4.3 Coordinate Conversion from MRSO (old) to Cassini (old)

A shapefile of the contours of Cameron Highlands in MRSO (old) coordinate system was projected into Cassini (old) using CCA v2.0. Its coordinates in text format was converted from MRSO (old) to Cassini (old) with GDTS v4.01. The results cannot be compared with ArcGIS since ArcGIS does not offer coordinate conversion for the old Cassini (non-geocentric). The results of CCA v2.0 were compared with GDTS and Table 5 shows the result.

Table 5: MRSO (old) to Cassini (old)

	GDTS – CCA	
	N	E
Max. Error	0.000371889	8.94101×10^{-7}
Min. Error	0.000310872	1.07288×10^{-10}
RMSE	0.018536393	

Referring to Table 4.3, CCA had achieved an RMSE of 0.018536393" against GDTS. The maximum and minimum Northing error for CCA are 0.000371889 and 0.000310872 respectively. For Easting, the maximum and minimum errors for CCA were 8.94101×10^{-7} and 1.07288×10^{-10} respectively. 0.018536393" is within the acceptable margin of positional error. CCA achieved an RMSE of less than 0.2 which was the suggested error allowance by JUPEM. Hence, this conversion can be considered a success.

5.0 Conclusion

To achieve the objectives, research questions corresponding to each objective must be answered.

For the first objective: To review the coordinate systems used in Malaysia, its research questions have been answered in Chapter 2: Literature Review. This chapter explored the coordinate systems used in Malaysia, their realisation, formulas involved in converting coordinate systems and the need to convert coordinates.

The second objective: To develop a web application that is capable of reading GIS data format, convert its coordinate system and save it in the new coordinate system, has been achieved by the scope description in Chapter 1: Introduction and through implementation of Chapter 3: Methodology. Chapter 1 discussed the software and hardware used for this research, answering the first research question. Chapter 3 elaborated extensively the methodology followed to program the application which involved Django and Python. The input, process and output has been described in section 3.3.1: Problem Definition, thefore answering the third research question. Section 3.3.2: Design answered the fourth research question, as it depicts the interface of the application. Finally, section 3.3.5: Validation answered the fifth research questions for the second objective, as it gave a list of conversions that must be validated, therefore setting the limit of what the application should do to be considered a success.

The third objective: To validate the accuracy of the converted coordinates, is answered in Chapter 4: Results and Analysis. Each of the subsections desribed the data used, the software used for comparison, the RMSE obtained and whether or not the result is within the error allowance.

In brief, the objectives for this research have been achieved.

References

[1] Chang, K.-T. 2014. Introduction to Geographic Information Systems, McGraw-Hill, New York.
[2] JUPEM(a). 2009. Technical Guide to the Coordinate Reference Systems, in: J.U.d.P. Malaysia .
[3] Navipedia. 2012. Regional Datums and Map Projections.
[4] Hooi, T.S. 2011. Comparison of Georeferencing Quality by Using Different Geographic Information Systems (GIS) Software, Faculty of Geoinformation and Real Estate, Universiti Teknologi Malaysia.
[5] Hendrikse, J.H.M. 2003. Use of the Spatial Reference Object Model to enhance Projection and Datum Transformation, Esri User ConferenceSan Diego (USA).
[6] Lowry, J. 2004. WGS-AGD-GDA: Selecting the correct datum, coordinate system and projection for north Australian applications, in: D.o.t.E.a. Heritage (Ed.)Australia.
[7] sbg-systems. 2013.Altitude Reference Models-Ellipsoid/Geoid/MSL.
[8] Maling, D.H. 1992. Coordinate Systems and Map Projections, 2nd ed. ed., Pergamon Press, Great Britain.
[9] Matsumura, S., Murakami, M. and Imakiire, T. 2004. Concept of the New Japanese Geodetic System, Bulletin of the Geographical Survey Institute.
[10] Ordance Survey. 2016. A Guide to Coordinate Systems in Great Britain, Crown, Southampton.
[11] Intergovernmental Committe on Surveying and Mapping. 2014. Geocentric Datum of Australia Technical Manual.
[12] Kadir, M., Ses, S., Omar, K., Desa, G., Omar, A. H., Taib, K., Teng, C. H., Mohamed, A., Chang, L.H., Saleh, R. and Nordin, S. 2003. Geocentric Datum GDM2000 for Malaysia: Implementation and Implications, Seminar On GDM2000 Kuala Lumpur, Malaysia.
[13] Galdos Systems Inc. 2017.EPSG Geodetic Parameter Registry.
[14] Kadir, M., Ses, S. and Hisam, A. 2017. Current and Future Geodetic Activities in Malaysia, GNSS ForumHong Kong, 2002.
[15] M. Edwards, Should I learn Django or Node.JS?.
[16] Big-nige. 2017. The Django Book: Why Django?.
[17] VsChart. 2017. Node.JS Vs. Django.
[18] L.A. Robertson. 2000. Simple Program Design: A Step-By-Step Approach, Third ed., Course Technology, Cambridge.