OTTAWA 2023
64TH WORLD STATISTICS CONGRESS

Statistics and GIS Journey in PCBS: From Data Collection to Results Dissemination

Mrs. Raghad K. Abulail

Email: rabulail@pcbs.gov.ps

Palestinian Central Bureau of Statistics, Ramallah, Palestine

Abstract

The world we live in is getting increasingly challenging. Population growth, over-consumption, and sustainability—these are some of the key global issues that we can consider understanding with GIS and data. Statistics based on data collection, analysis, and interpretation need a geographical background to understand it more, and thus to make more accurate/informed decisions. Where by statistics alone are not relevant without being associated with a specific place/location.

The Palestinian Central Bureau of Statistics (PCBS) has embarked on a special experience in the field of GIS, starting with establishing geo-spatial databases at all statistical geographic levels (DBMS), moving on collecting the statistical primary data from the fieldwork, and joining it with its spatial coordinates, to represent it on statistical thematic maps, and visualize and disseminate it through online interactive platforms, atlases, and open data sites. This made it easier for researchers, decision-makers, and strategic planners in Palestine to easily access statistical indicators data and analyze it based on spatial location and info-graphics provided by the GIS Web-based platforms and portals.

In this paper, a lengthy and detailed presentation will be made about the effectiveness of using GIS applications in serving statistical data in the surveys and censuses projects of PCBS, and the outputs that were produced as data collection techniques, geo-databases, thematic maps and atlases, interactive and story maps platforms, and open data sites to provide high-quality datain an understandable manner.

Keywords: GIS, Statistics, Fieldwork, Data Collection, Data Representation, Data Visualization, Dissemination, Geo-databases.

1. Introduction

GIS stands for Geographic Information Systems, and it is defined as a computer-based system that aims to store, manage, analyze, and display geospatial data. As it relies on technological hardware such as high-speed computers with large storage capacity, Global Positioning System (GPS)in the fieldwork, printers, and scanners to print maps, in addition to commercial programs or open sources software packages for computing geospatial data and joining them with their respective statistical data [1]. Furthermore, the GIS system needs digitizers, specialists, data collectors, and organizations that analyze data, make decisions, and develop strategic plans.

ESRI Technology is the industry leader in GIS, providing a suite of commercial software packages offering various tools and features to serve this field. Each of these systems is composed of applications and extensions with three license levels. These applications are ArcMap, ArcGIS Pro, ArcCatalog, ArcScene, and ArcGlobe, with basic, standard, and advanced license levels.

PCBS was able to use all the applications provided by ESRI to create an integrated GIS system since 2008. It made a long journey in keeping pace with the updates, new applications and developments in the field of GIS to produce a comprehensive system with a high level of efficiency and organization until the year 2023, all with the aim of statistical data service from the stage of its collection from the field, to the stage of representation and exploration, until the stage of disseminating and visualizing it in a modern interactive manner to the world.

2. Methodology

This research is qualitative research that presents the experience and expertise of PCBS at the administrative level and technological infrastructure. The data was collected through conducting semi-structured interviews with the Director of the Department of Maps and Geographic Information Systems at PCBS, and by referring to previous scientific papers that present part of similar global and local experiences (literature review), and finally through the observations and experience of the author of the study, as she is an expert in the Department of Geography and GIS and has more than four years of practical experience in several projects that include all the stages covered in this paper. The following is the set of questions that the study will answer:

1-What is the history and techniques of establishing and maintaining geo-databases in the GIS Department to create a base map for the State of Palestine in PCBS?

2-What is the mechanism\techniques for collecting raw statistical data from the field within the framework of high quality collection and control?

3-What is the mechanism\techniques for displaying the treated statistical results based on spatial coordinates?

4-What are the outputs developed to display and disseminate the results through Web-Based (cloud) technologies?

3. Results (Statistics and GIS Journey in PCBS)

This chapter talks about the stages that statistical data undergoes based on the GIS. Figure (3.1) displays a representation of the data journey, followed by a breakdown of each phase:

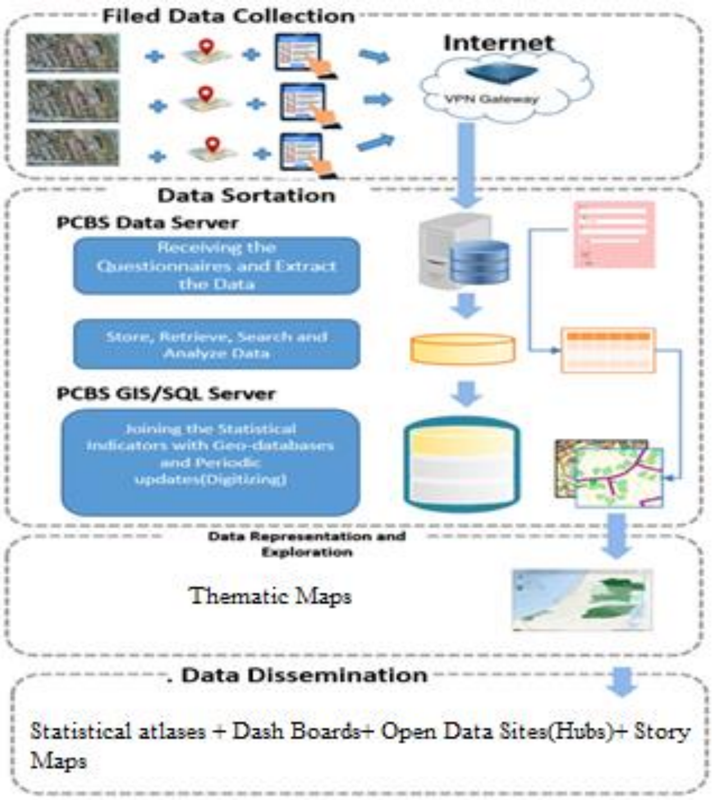

Figure (3.1): The Stages in the Statistical Information Journey Based On GIS

3.1. Data Collection and Storage

Undoubtedly, fieldwork is the starting point for obtaining statistical data in any survey or census, which requires organized human and technological efforts to collect high-quality, real-time, and accurate data, through advanced technological equipment connected to databases that store raw statistical data immediately from the field. Thus, the computerization of the statistical data collection system based on GIS will help in monitoring the progress of work and ensure a high degree of monitoring over fieldwork enumerator, in addition to the reduction of human errors in the field, such as duplicating data, missing data, wasting field survey, or lack of clarity and accuracy, in addition to saving time, effort, and even money.

Recently, the use of mobile communication devices, such as smart phones, tablets, and cellular phones in field data collection is increasing due to the emergence of embedded Global Position Systems (GPS) and Wi-Fi Internet access. In order to project the experience of PCBS, the following is a discussion in two parts:

3.1.1. Data Storage (Geospatial Databases)

According to a personal interview with the head of the geography and GIS Department at PCBS; at the beginning of 2008, the Department has started a project of computerizing (digitizing) paper maps into geospatial databases, using high-resolution aerial photographs that can be obtained annually from the Ministry of Local Government, to computerize all landmarks and give them geospatial coordinates and store them in relational databases.

Speaking of what these geospatial databases are; they follow the relational database management system (DBMS). The files that include this type of databases are named "Shape files", and the transformation system “WGS 1984” according to ESRI's applications. Shape files are stored on a dedicated SQL and GIS servers and serviced by ArcMap and ArcGIS Pro applications, in the provision graphical user interface to carry out countless multiple tasks, the main task is creating tables (named layers) that include descriptive and statistical data for each landmark/geographic level on the one hand, and geospatial data (coordinates) that act like the ID of the landmark on the other hand (unique shape length and shape area or x and y for each landmark without duplication). With the ability to modify these layers and perform various statistical operations on the data joined to them, analyze them, and beyond that.

ESRI's principles have been followed in the representation of each family of features, as each feature layer has its own representation, such as Point representation (x and y coordinates) for buildings layer, a Line for streets and road layers, Polygon for geographical boundaries layer, and Polyline. All geographical statistical levels layers have been targeted, including a country, region, governorate, localities, statistical district, enumeration area, streets, roads, and pathways. In addition to layers with a political dimension such as barriers, the annexation and expansion wall, Oslo divisions, settlements, and outposts. Whose shape files were obtained from external parties to be included in the computerized geo-databases within the GIS department, and to adopt a formal base map for Palestine based on the mentioned layers. In order to ensure the comprehensiveness of the work during the engagement in the fieldwork, in addition to the possibility of joining it later with statistical indicators.

3.1.2. Data Collection and Monitoring (Fieldwork)

To present what is happening in the fieldwork, two main applications will be discussed at this stage; the application of data collection and the fieldwork management application dedicated to monitoring enumerators and workflow in the field. As these two techniques formed a system to ensure the quality of statistical work in the field according to international standards.

For data collection, each field enumerator is provided with a tablet device linked to internet service (SIM card, WIFI), GPS, and VPN technology, and supported with a synchronized specific application linked with pre-defined databases on the main central dedicated server to collect the data. This application relies on the digitized geographical layers stored within the geospatial databases, and a tile package files (TPK) that are specially prepared by ESRI’s applications to display the offline aerial image in all its details and landmarks to facilitate the enumerator’s task in inference, and it also relies on data entry technique linked with the digitized buildings taken from the geospatial databases, since each questionnaire in the sample will not be possible to start filling it out unless there is a match between the GPS coordinates in the field with the pre-computed spatial coordinates of the target building. The application also includes logical checks and warning messages regarding filling out each questionnaire. All of that ensures that the enumerator will reach the target buildings in the EA's (Enumeration Area) assigned to them and complete the questionnaire filling with low probability of human errors. Thus insures higher data quality and credibility. In addition, there is a possibility for fieldworkers to update geospatial data from the field of work through adding, deleting or modifying buildings, landmarks, building characteristics (such as the number of floors or the number of housing units), and starting point or the external boundaries in a way that match the reality on the ground. This is followed by obtaining an updated geo-database and submitting to many topological rules (such as EAs mustn't overlap and shouldn't have any gaps between them) to be used as formally revised layers in the established geo-databases in the GIS department [2]. Moreover, VPN technology enables the enumerator to send real-time data (raw data) to the previously mentioned server, once the enumerator finishes filling out a questionnaire during work.

As for fieldwork tracking, the fieldwork management application has been developed, relying mainly on aerial photographs and geospatial databases, in addition to other management and monitoring features to serve fieldwork. This enables managers of statistical projects to follow up on the workflow and daily achievements of each supervisor, crew leader, and enumerator through a decentralized monitoring, distribute job tasks among them and follow up on their progress. As the application is supported by a GPS-based technology to limit the flow of the enumerator according to the EAs assigned to them by the supervisors [3].

3.2. Results Representation and Exploration

The aforementioned geospatial databases can be joined with the data of statistical indicators whose raw data were collected from the field and analyzed in the competent departments. The joining technique takes place through the existence of a common field (primary key like governorate code) between the geospatial databases and the statistical indicators data files, which results in tables (Shape files According to ESRI) that include statistical and spatial data, so that each statistical information is associated with a specific spatial dimension.

One of the main applications of GIS is related to the possibility to display statistical information on the maps according to the joining mechanism mentioned above, so that the users can have a geographical perspective on the represented phenomenon. The main representation technique used in GIS department is the “thematic maps”. There are several techniques implemented in ArcGIS, used to classify some items in accordance with the value of one indicator (attribute). To achieve this, it’s a must to identify disjoint intervals of values and the indicator used to form groups. For example, the simplest way to create ranges of values is the division of the values used to make the classification, into equal intervals. The Natural Breaks method, used to split the ranges, is based on the shaping of the groups of values, so that to have a minimum dispersion inside the group and a maximum dispersion between groups [4]. In ESRI’s software, colors could be defined individually. In this case there are special classes for various color models, like RgbColor class used for RGB model, and as collections of colors, managed by color ramp classes, having the name ColorRamp.Collections of colors can be generated by the programmer or he can usethe default collections that are defined by ESRI’s style gallery. These results in a map with different color gradations, or with graphic shapes like charts, or other graphics that express the values of the indicator according to the geographical location compared to another geographical location in the same map, which helps in exploring the value of the indicators and their interpretation based on the spatial dimension. ArcMap and ArcGIS Pro are the applications that serve this type of representation.

3.3. Results Dissemination and Visualization

ESRI offers a cloud-based applications package through the ArcGIS Online subscription. Through these applications, statistical indicator data can be displayed on an online thematic map and analyzed through multiple infographics (as bar charts). Also, these applications allow the technique of downloading statistical data according to spatial coordinates in any extension desired by the user.

3.3.1 Statistical Atlases

The thematic maps discussed in the data presentation section were used to create statistical atlases supported by graphic figuresby using the aforementioned geospatial databases, which constitute the base map of Palestine, and joining its spatial coordinates with the statistical indicators data that were processed in the competent departments. As a result, paper atlases were produced, and they were published using PDF technology with the ability to download them through the official website of PCBS. Among these atlases are the statistical atlases [5], poverty atlas [6], the agricultural atlas, the establishments atlas, and others.

3.3.2 Dashboards

Dashboards are available with all ArcGIS Online subscriptions, and designed to display multiple visualizations that work together on a single screen. They offer a comprehensive view of the statistical data and provide key insights for at-a-glance decision-making. Thematic maps are one of its most important elements, supported by a set of interactive figures, images and charts. PCBS prepares an interactive map for each indicator, in both Arabic and English, and then similar indicators are collected within sections and displayed within a platform dedicated to interactive maps linked to the PCBS page [7]. Another platform similar in technology is SDGs platform which has been produced and published through PCBS website [8].

3.3.3 Hubs and Open Data Sites

ArcGIS Hub is another application available with all ArcGIS Online subscriptions. It is a cloud-based engagement platform that enables PCBS to communicate more effectively with their communities and public audience. Through this technology, a comprehensive platform was produced to display all the outputs of the GIS department, including interactive maps, electronic atlases, story maps, and others. Another platform was also produced to download the data of statistical indicators, which acts an open data site to provide the data by tabulating indicators with similar content and displaying its tables with the ability to download all its data joined to geospatial coordinates. Display and download techniques are supported by a filtering system for the rows and columns desired by the user, and the data download technology is supported by multiple extensions as desired by the user, from shape files, CSV, JSON, KML, etc [9]. These platforms enabled PCBS to publish its data to the public for any user at any time and place without any limitations.

3.3.4 Story Maps

Story maps use GIS tools to combine geospatial data with photos, video, audio, and text to visualize a theme or sequential events. Early Marriage in Palestine [10] and Women's Empowerment in the Labor Market [11] are the popular story maps produced by the Department of GIS in PCBS.

4. Conclusion and Recommendations

The science and applications of GIS have revolutionized the world of statistics, as they serve statistical data and information in terms of collecting, storing, updating, displaying, visualizing, and disseminating. All these are within the framework of a very high quality that conforms to international standards and with few human and technological errors compared to the old traditional methods. Furthermore, it saves the efforts of manpower, the funds of institutions, and a lot of the time required in managing statistical projects such as surveys and censuses. And beyond that, serving decision makers, strategic planners, and researchers. Thus, this working paper advises statistical agencies around the world to keep up with what is new in GIS, including programs, applications, training courses, technical missions, and new technologies. And try to continuously develop the phases of surveys and censuses by restructuring them with GIS.

5. References

[1] K.Chang, INTRODUCTION TO GEOGRAPHIC INFORMATION SYSTEMS, New York, 2019. Accessed: Jan. 2, 2023. [Online].Available: https://www.mheducation.com.

[2] A. Amro and M.Abu alhayja, "A The Effect of Using New Technology and Geographic Information System on The Quality of Official Statistics: The Implementation of The General Palestinian Census 2017 as a Case Study ", PCBS, Ramallah, Palestine, March. 12, 2019. Accessed: Feb. 10, 2023. [Online]. Available: https://www.pcbs.gov.ps/papersar.aspx?lang=ar.

[3] A. Amro and M.AbuAlhayja, "A The Effect of Using New Technology and Geographic Information System on The Quality of Official Statistics: The Implementation of The General Palestinian Census 2017 as a Case Study ", PCBS, Ramallah, Palestine, March. 12, 2019. Accessed: Feb. 10, 2023. [Online]. Available: https://www.pcbs.gov.ps/papersar.aspx?lang=ar.

[4] A. REVEIU and M. DÂRDALĂ, "Techniques for Statistical Data Visualization in GIS ", Journal of Informatica Economică, vol. 15, no. 3/2011, March.2011. Accessed: Feb. 1, 2023. [Online]. Available: https://www.academia.edu/.

[5] https://www.pcbs.gov.ps/pcbs_2012/Publications.aspx

[6]] https://www.pcbs.gov.ps/site/lang__ar/1220/default.aspx?lang=ar.

[7] https://interactive-maps-pcbs.hub.arcgis.com/.

[8] https://www.pcbs.gov.ps/MainSDGs.aspx.

[9] https://geospatial-file-download-center-pcbs.hub.arcgis.com/.

[10] https://interactive-maps-pcbs.hub.arcgis.com/apps/a167b495c4a54128be132ee0d3ceec9d/explore

[11] https://interactive-maps-pcbs.hub.arcgis.com/apps/4d04de3f0c124a89aa915d27e4bf3f67/explore.