Mashup Tools for Big Data Analysis in Maritime Surveillance

George Melillos*[ab], Kyriacos Themistocleous[ab], Chris Danezis[ab], Silas Michaelides[ab], Diofantos G. Hadjimitsis[ab] Sven Jacobsen[c], Björn Tings[c]

[a]Department of Civil Engineering and Geomatics, Faculty of Engineering and Technology, Cyprus University of Technology, Limassol, Archbishop 30 Arch. Kyprianos Str., 3036, Lemesos, Cyprus.
[b]ERATOSTHENES Centre of Excellence, 2-8 Saripolou,3036 Limassol, Cyprus
[c]German Aerospace Center, Remote Sensing Technology Institute, (DLR-IMF), Bremen, Germany

ABSTRACT

The growth of big data and its popularity in maritime surveillance has increased at an exponential rate. The amount of maritime information being collected every minute around the world exceeds the capacity of traditional databases. The development of real-time, Geospatial Web Applications e.g., MarineTraffic and VesselFinder AIS vessel tracking web sites, provide us with huge sets of structured and unstructured data that are too complex for traditional data-processing software. The aim of this paper is to exploit the benefits of query and mashup amounts of maritime data using mashup tools as a result to create a single, unique visualization. The results show that using mashup techniques in maritime surveillance could be used to monitor, compare, combine, manipulate and analyse Big Maritime data. Therefore, research on Maritime Data offers a huge potential and an opportunity to benefit from the advantages.

Keywords: Maritime surveillance, big data, mashup tools, python, web scraping, AIS

1. INTRODUCTION

Big Data is a common concept to define datasets whose size exceeds the processing capacity of traditional database systems [1]. While this is not the commonly agreed definition, Big Data is generally characterized by three V's: volume, velocity, and variety [2,3,4,5]. Volume dimension relates to the size of data from one or more data resources in tera-, peta-, or exabytes. The velocity dimension focuses on the data streams and how to store near real-time data, as well as handling the increasing rate of the data amount. The latter, namely, the variety dimension, is associated with the heterogeneity of data both at the schema-level and the instance-level [6].

Big Data brings innumerous challenges, commonly divided into four categories: (a) general dilemmas, such as the lack of consensus and rigor in the definition, models, architectures or benchmarks; (b) challenges related to the Big Data life cycle, from collection to analysis; (c) security, privacy and monitoring issues; and, finally, (d) organizational change, such as new required skills (e.g., data scientists) or changes in workflows to accommodate the data-driven mindset [7].

Working with Big Data implies knowledge from multiple disciplines; the term data science is frequently highlighted to designate the area responsible for dealing with Big Data throughout the stages of its life cycle, relying on the scientific method (defining hypothesis and validating conclusions) and on knowledge related to areas like machine learning, programming and databases, etc. [7].

Big Data is a research field involving a large number of collaborating disciplines [8]. The typical target group of Big Data solutions is knowledge able knowledgeable in different domains who are not familiar with the technical details of Big Data and data integration. As a result, there is a growing need to provide a solution with a smaller learning rate for such users.

We can consider mashup as an effective tool to support users in creating user-generated solutions based on available private/public resources and integrate several data sources with different formats easily [6]. As a result, both skilled programmers and non-skilled users are able to benefit from the large amounts of data [6] and solve any problems they may encounter.

The mashup approach allows users to build ad-hoc applications by combining several different data sources and services from across the web [6]. The aim is to combine these sources to create useful new applications or services. Content and presentation elements typically come in the form of RSS or Atom feeds, various XML formats, or as HTML or other graphical elements. Publicly available APIs (in JavaScript, for example) typically provide application functionality. Content, functionality, and presentation are then glued together in disparate ways: via JavaScript in the browser, server-side scripting languages such as Hypertext Preprocessor (PHP) or Ruby, or traditional languages such as Java or C# [9].

There are three approaches for the development of mashup solutions. First, the manual approach, which requires programming or scripting skills of users to integrate the data sources, generate visualizations, and create new functionalities. Second, the semi-automatic, which assists the users to build a mashup application using provided tools. Third, the automatic approach which allows creation of mashups without user's involvement, as the resources (data, visualization, as well as functionality) are chosen and invoked automatically by the following tools [10]:

1. spreadsheet-based tools, in which the users provide the data directly into a spreadsheet; The examples of this category are AMICO:CALC and MashSheet [11].
2. widget-oriented tools, allow users to create the mashup through a visual editor. Yahoo Pipes and Intel Mash Maker are examples in this category of mashups [6].
3. demonstration-based tools, allow users to mash up their data by providing examples and completing the data integration task via a visual step-by-step process. The instances in this category are Dapper and Karma [12].

The aim of this paper is to extract and compare AIS data from various webs sources by exploiting the benefits of web scraping [13] and mashup tools using Python programming language. Specifically, web scraping is the practice of gathering data through any means other than a program interacting with an API (or, obviously, through a human using a web browser). This is most commonly accomplished by writing an automated program that queries a web server, requests data, and then parses process that data to extract needed necessary information [13] in order to create a single, unique visualization.

The results show that using mashup techniques in the maritime surveillance could monitor, compare, combine, manipulate and analyse Big Maritime data. Maritime Data offers a huge potential, but further research is required in order to benefit from the advantages.

METHODOLOGY

As mentioned above, Python Programming Language was used to web scraping. Some advantages of the Python Language are: simple and easy to learn, free and open source software, works on different platforms, Python supports both: process-oriented function programming and object-oriented abstract programming, scalability and embeddability [14].

The **urllib** module was used to fetching URLs (Uniform Resource Locators), as shown in Figure 1. **Urllib** is a Python standard web request library that contains functions for network data requests, handling cookies, changing request headers and user agents, redirects, authentication, etc. [14]. As shown in Figure 1, in the script some essential functions were included in order to retrieve the appropriate data. These blocks of code are the URL of the website that contains the port name (e.g., https://www.fleetmon.com/ports/piraeus-athens_grpir_7251/). Moreover, the titles of the Fields - Columns names and how many Records-Rows will be extracted. It should be noted that each Vessel Tracking Website has its own database structure so the script should be modified accordingly.

Accessing data is only half of the problem; the other half is to automate this process. Some URLs may change or depreciate over time, which makes Python scripts run to an error, when a command such as **try** or **except** is used to check if the URL is still valid. Currently, to keep all versions of downloaded files, we have used the **datetime** module to add a timestamp to the output filename, too.

Additionally, the **xlwt** module was used in the script to import the data into a single and unique visualization. **Xlwt** is a library for writing data and formatting information to Excel files (i.e., .xls) [15]. In order to force automatic refresh of any data connections when we open Excel workbooks, a short snippet VBA code was used. These lines of code used the **ThisWorkbook Open event**, which points to the currently active workbooks. Next, a simple one-line of code that uses the **RefreshAll** method to refresh all of the connections that are contained within your workbook or worksheet was used.

Comparison of data from different web sources was made by using **Power Pivot** which is an Excel add-in to perform powerful data analysis and create sophisticated data models. With Power Pivot, you can mashup large volumes of data from various sources, perform information analysis rapidly, and share insights easily [17].

In both Excel and in Power Pivot, you can create a Data Model, as a collection of tables with relationships. The data model we see in a workbook in Excel is the same data model we see in the Power Pivot window. Any data we import into Excel is available in Power Pivot, and vice versa [17].

The overall methodology adopted in this study consists of six (6) processing steps briefly described below (see Figure 2). It must be emphasized that the methodology can be applied for any website which contains structured data.

Step 1: Identify Vessel Tracking Websites: Marine-Traffic, FleetMon and VesselFinder,

Step 2: Installing Python and Python Packages.

Step 3: Algorithm development and code writing.

Step 4: Running, Testing and Debugging the program.

Step 5: Data acquisition.

Step 6: Comparison of data from different data sources is performed using Power Pivot.

```python
import urllib.request
import urllib.error
import re
url_dept = 'https://www.fleetmon.com/ports/piraeus-athens_grpir_7251/'

try:
    req = urllib.request.Request(url_dept)
    with urllib.request.urlopen(req) as response:
        char_set = response.headers.get_content_charset()
        html = response.read().decode(char_set)
except urllib.error.HTTPError as e:
    print('Error HTTP:', e.code)
except urllib.error.URLError as e:
        print('Server Error')
        print('Reasons: ', e.reason)
else:
    h2_tags = re.findall(r"<h2\b[^>]*>(.*?)</h2>", html)
    count = 0
    for tag in h2_tags:
        code = re.findall(r"\[(.*?)\]", tag)
        if len(code)> 0 :
            code = code[0].strip()
            name = re.findall(r']:(.*)', tag)
            if len(name) > 0 : name = name[0].strip()
            else: name = ''
            print(code, name)
            count += 1
```

Figure 1. Part of Code for Data Extraction.

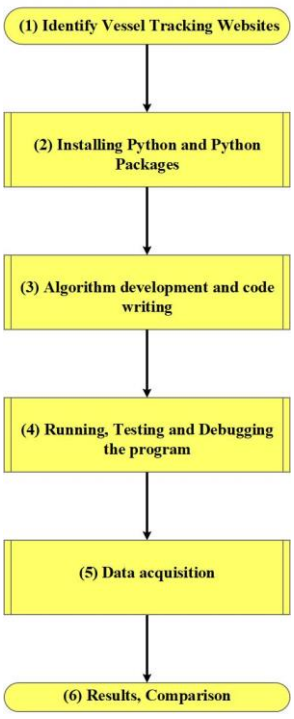


Figure 2. Methodology.

4. RESULTS

After running the code, the data were extracted in Microsoft Excel format, as shown in Figure 3. As mentioned above, the data were retrieved from Vessel Tracking Websites such as **Marine-Traffic**, **FleetMon** and **VesselFinder**, as shown in Figure 1. It is worth mentioning that the Area of Interest (AOI) is the Port of Piraeus, Greece. The extracted data lists the arrivals of ships in the port on 25 August 2020 from 01:00 am till 11:00 am.

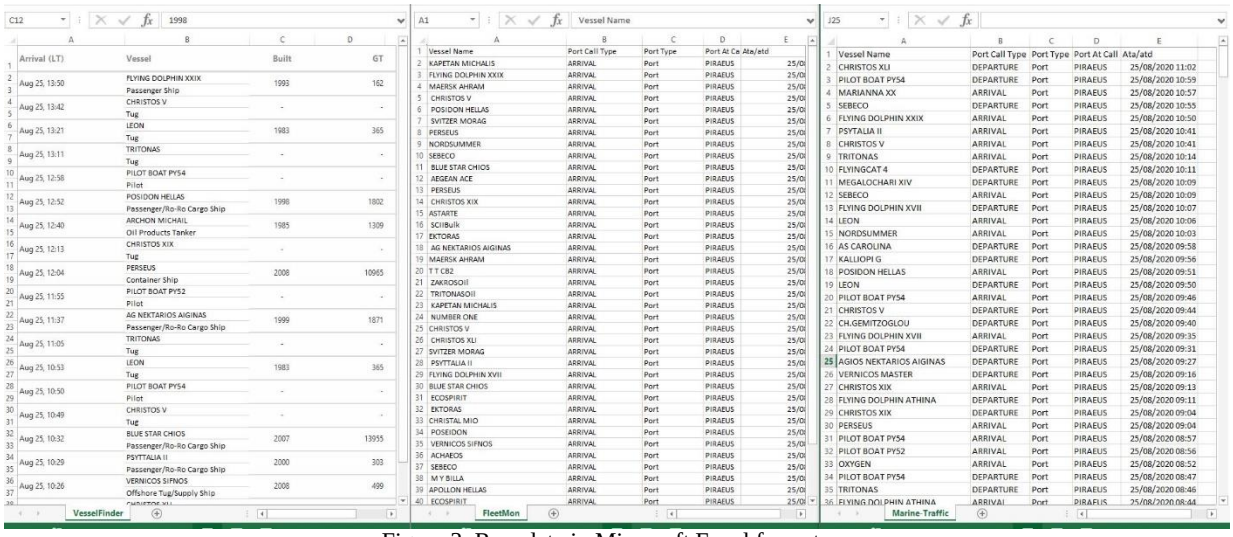
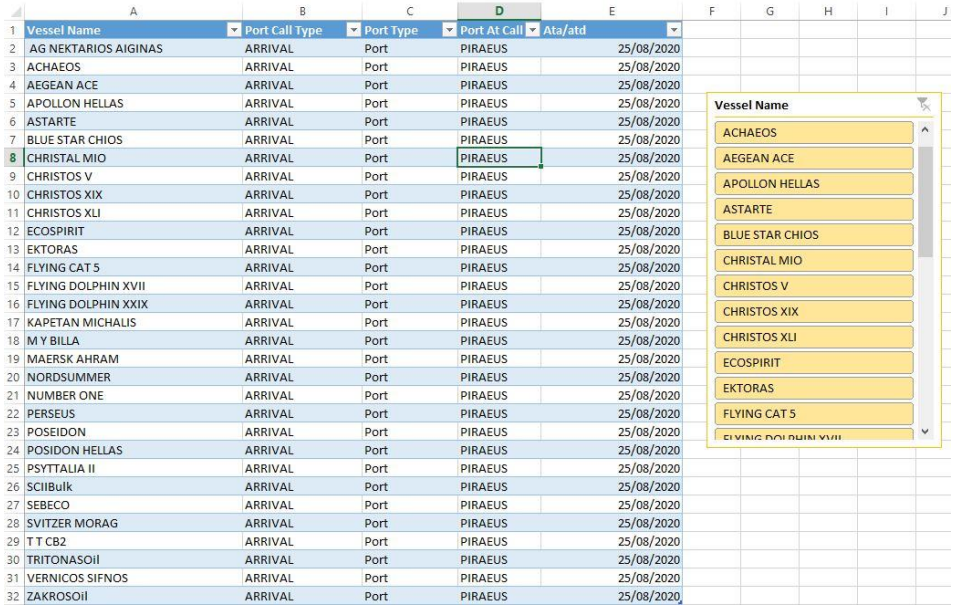


Figure 3. Raw data in Microsoft Excel format.

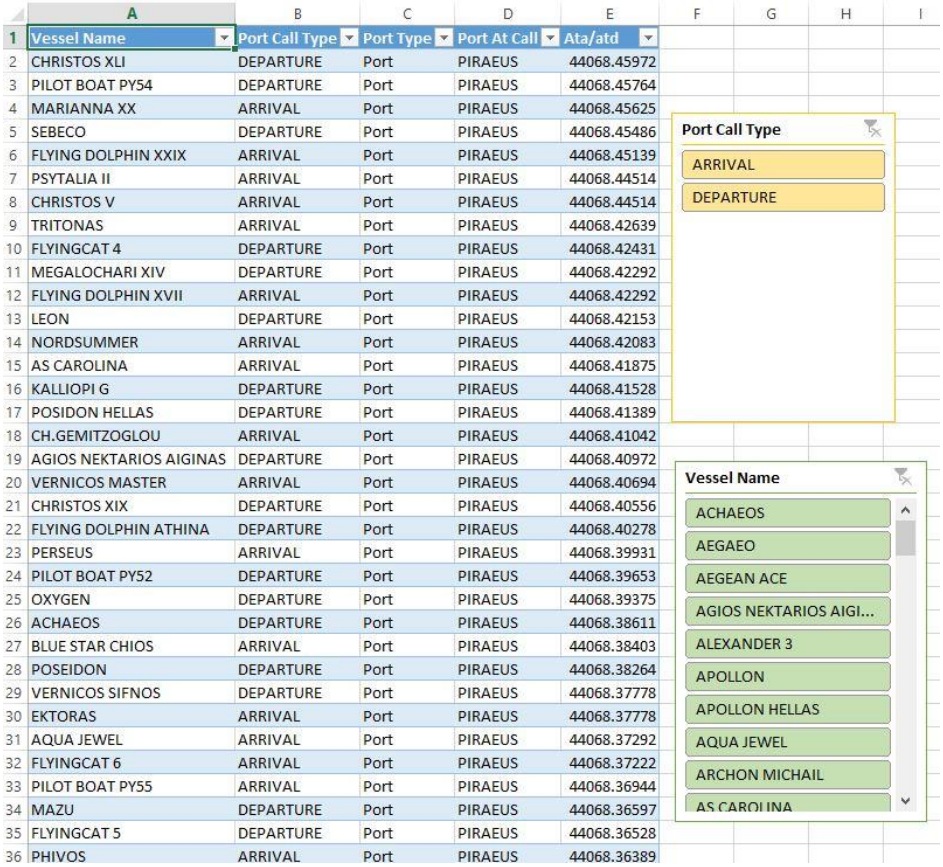
Furthermore, the basic issues involved in the mashup creation process are data retrieval, source modeling, data cleaning, data integration, and data visualization. Each of these issues is an area of research in its own and our goal is to prevent the end-user from delving into these underlying complexities during the mashup building process [18]. An improper visualization of the data could result in users wasting precious time to understand the data [18]. The Mashup environment should be user friendly, so that the user can monitor, compare, combine, manipulate and analyse Big Maritime data. Accordingly, a comparison of data was made by using the Power Pivot add-in.

Figures 4 and 5 present a typical example of the data results using Power Pivot data modelling technology. The Slicer tool allows to filter the information in the pivot table, by using one or more fields and the ability to let slicers “Show items with no data last” filter pivot tables. Using this tool, you can manipulate Big Data.



Vessel Name	Port Call Type	Port Type	Port At Call	Ata/atd
AG NEKTARIOS AIGINAS	ARRIVAL	Port	PIRAEUS	25/08/2020
ACHAEOS	ARRIVAL	Port	PIRAEUS	25/08/2020
AEGEAN ACE	ARRIVAL	Port	PIRAEUS	25/08/2020
APOLLON HELLAS	ARRIVAL	Port	PIRAEUS	25/08/2020
ASTARTE	ARRIVAL	Port	PIRAEUS	25/08/2020
BLUE STAR CHIOS	ARRIVAL	Port	PIRAEUS	25/08/2020
CHRISTAL MIO	ARRIVAL	Port	PIRAEUS	25/08/2020
CHRISTOS V	ARRIVAL	Port	PIRAEUS	25/08/2020
CHRISTOS XIX	ARRIVAL	Port	PIRAEUS	25/08/2020
CHRISTOS XLI	ARRIVAL	Port	PIRAEUS	25/08/2020
ECOSPIRIT	ARRIVAL	Port	PIRAEUS	25/08/2020
EKTORAS	ARRIVAL	Port	PIRAEUS	25/08/2020
FLYING CAT 5	ARRIVAL	Port	PIRAEUS	25/08/2020
FLYING DOLPHIN XVII	ARRIVAL	Port	PIRAEUS	25/08/2020
FLYING DOLPHIN XXIX	ARRIVAL	Port	PIRAEUS	25/08/2020
KAPETAN MICHALIS	ARRIVAL	Port	PIRAEUS	25/08/2020
M Y BILLA	ARRIVAL	Port	PIRAEUS	25/08/2020
MAERSK AHRAM	ARRIVAL	Port	PIRAEUS	25/08/2020
NORDSUMMER	ARRIVAL	Port	PIRAEUS	25/08/2020
NUMBER ONE	ARRIVAL	Port	PIRAEUS	25/08/2020
PERSEUS	ARRIVAL	Port	PIRAEUS	25/08/2020
POSEIDON	ARRIVAL	Port	PIRAEUS	25/08/2020
POSIDON HELLAS	ARRIVAL	Port	PIRAEUS	25/08/2020
PSYTTALIA II	ARRIVAL	Port	PIRAEUS	25/08/2020
SCIIBulk	ARRIVAL	Port	PIRAEUS	25/08/2020
SEBECO	ARRIVAL	Port	PIRAEUS	25/08/2020
SVITZER MORAG	ARRIVAL	Port	PIRAEUS	25/08/2020
T T CB2	ARRIVAL	Port	PIRAEUS	25/08/2020
TRITONASOil	ARRIVAL	Port	PIRAEUS	25/08/2020
VERNICOS SIFNOS	ARRIVAL	Port	PIRAEUS	25/08/2020
ZAKROSOil	ARRIVAL	Port	PIRAEUS	25/08/2020

Figure 4. Pivot Table: Extract data from FleetMon Vessel Tracking Website.



Vessel Name	Port Call Type	Port Type	Port At Call	Ata/atd
CHRISTOS XLI	DEPARTURE	Port	PIRAEUS	44068.45972
PILOT BOAT PY54	DEPARTURE	Port	PIRAEUS	44068.45764
MARIANNA XX	ARRIVAL	Port	PIRAEUS	44068.45625
SEBECO	DEPARTURE	Port	PIRAEUS	44068.45486
FLYING DOLPHIN XXIX	ARRIVAL	Port	PIRAEUS	44068.45139
PSYTALIA II	ARRIVAL	Port	PIRAEUS	44068.44514
CHRISTOS V	ARRIVAL	Port	PIRAEUS	44068.44514
TRITONAS	ARRIVAL	Port	PIRAEUS	44068.42639
FLYINGCAT 4	DEPARTURE	Port	PIRAEUS	44068.42431
MEGALOCHARI XIV	DEPARTURE	Port	PIRAEUS	44068.42292
FLYING DOLPHIN XVII	ARRIVAL	Port	PIRAEUS	44068.42292
LEON	DEPARTURE	Port	PIRAEUS	44068.42153
NORDSUMMER	ARRIVAL	Port	PIRAEUS	44068.42083
AS CAROLINA	ARRIVAL	Port	PIRAEUS	44068.41875
KALLIOPI G	DEPARTURE	Port	PIRAEUS	44068.41528
POSIDON HELLAS	DEPARTURE	Port	PIRAEUS	44068.41389
CH.GEMITZOGLOU	ARRIVAL	Port	PIRAEUS	44068.41042
AGIOS NEKTARIOS AIGINAS	DEPARTURE	Port	PIRAEUS	44068.40972
VERNICOS MASTER	ARRIVAL	Port	PIRAEUS	44068.40694
CHRISTOS XIX	DEPARTURE	Port	PIRAEUS	44068.40556
FLYING DOLPHIN ATHINA	DEPARTURE	Port	PIRAEUS	44068.40278
PERSEUS	ARRIVAL	Port	PIRAEUS	44068.39931
PILOT BOAT PY52	DEPARTURE	Port	PIRAEUS	44068.39653
OXYGEN	DEPARTURE	Port	PIRAEUS	44068.39375
ACHAEOS	DEPARTURE	Port	PIRAEUS	44068.38611
BLUE STAR CHIOS	ARRIVAL	Port	PIRAEUS	44068.38403
POSEIDON	DEPARTURE	Port	PIRAEUS	44068.38264
VERNICOS SIFNOS	DEPARTURE	Port	PIRAEUS	44068.37778
EKTORAS	ARRIVAL	Port	PIRAEUS	44068.37778
AQUA JEWEL	ARRIVAL	Port	PIRAEUS	44068.37292
FLYINGCAT 6	ARRIVAL	Port	PIRAEUS	44068.37222
PILOT BOAT PY55	ARRIVAL	Port	PIRAEUS	44068.36944
MAZU	DEPARTURE	Port	PIRAEUS	44068.36597
FLYINGCAT 5	DEPARTURE	Port	PIRAEUS	44068.36528
PHIVOS	ARRIVAL	Port	PIRAEUS	44068.36389

Figure 5. Pivot Table: Extract data from Marine-Traffic Vessel Tracking Website.

The Power Pivot can also establish, and graphically represent, relationships between the data included in the model. Figure 6 shows the Power Pivot window in Diagram view. The relationships are established between the three Vessel Tracking Website tables using the Vessel name. Relationships helped us to combine data from three different tables.

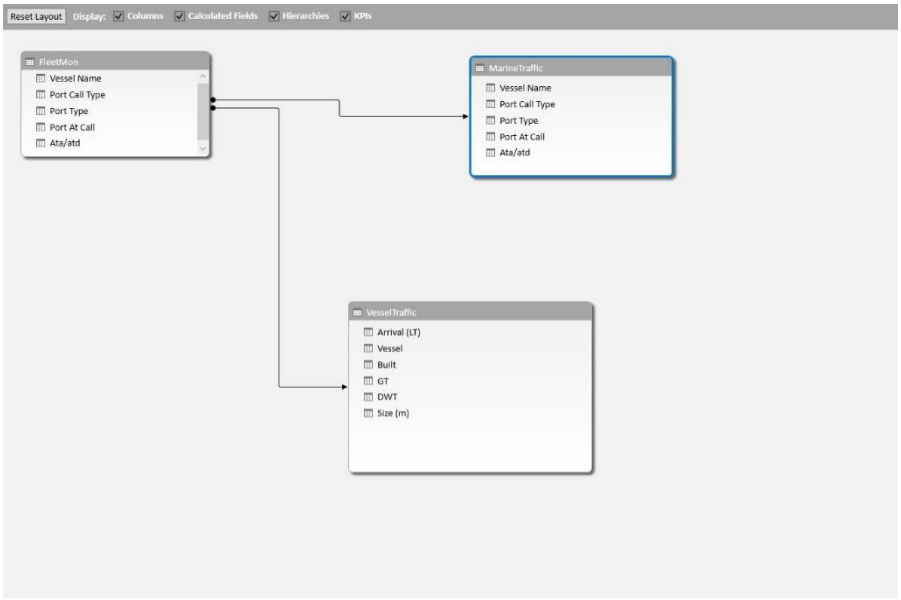


Figure 6. Relationships between the three tables in diagram view.

The results of comparing the data between the two Vessel Tracking Websites: Marine-Traffic and FleetMon after the above relationships are shown in Figure 7. Figure 8 depicts the results between the two Vessel Tracking Websites: Marine-Traffic and VesselFinder. The results (inside rectangular red boxes) show the differences between the two Vessel Traffic Websites. The following differences are due to the fact that are a result of data being gathered from different Automatic Identification System (AIS) equipped voluntarily by contributors in over 140 countries around the world [19].

Marine Traffic Data						FleetMon Data				
Vessel Name	**Port Call Type**	**Port Type**	**Port At Call**	**Ata/atd**	**Comparing Data**	**Vessel Name**	**Port Call Type**	**Port Type**	**Port At Call**	**Ata/atd**
ACHAEOS	DEPARTURE	Port	PIRAEUS	25/08/2020	ACHAEOS	AG NEKTARIOS AIGINAS	ARRIVAL	Port	PIRAEUS	25/08/2020
AEGAEO	DEPARTURE	Port	PIRAEUS	26/08/2020	Not Match	ACHAEOS	ARRIVAL	Port	PIRAEUS	25/08/2020
AEGEAN ACE	ARRIVAL	Port	PIRAEUS	27/08/2020	AEGEAN ACE	AEGEAN ACE	ARRIVAL	Port	PIRAEUS	25/08/2020
AGIOS NEKTARIOS AIGINAS	DEPARTURE	Port	PIRAEUS	28/08/2020	Not Match	APOLLON HELLAS	ARRIVAL	Port	PIRAEUS	25/08/2020
ALEXANDER 3	DEPARTURE	Port	PIRAEUS	29/08/2020	Not Match	ASTARTE	ARRIVAL	Port	PIRAEUS	25/08/2020
APOLLON	DEPARTURE	Port	PIRAEUS	30/08/2020	Not Match	BLUE STAR CHIOS	ARRIVAL	Port	PIRAEUS	25/08/2020
APOLLON HELLAS	DEPARTURE	Port	PIRAEUS	31/08/2020	APOLLON HELLAS	CHRISTAL MIO	ARRIVAL	Port	PIRAEUS	25/08/2020
AQUA JEWEL	ARRIVAL	Port	PIRAEUS	01/09/2020	Not Match	CHRISTOS V	ARRIVAL	Port	PIRAEUS	25/08/2020
ARCHON MICHAIL	DEPARTURE	Port	PIRAEUS	02/09/2020	Not Match	CHRISTOS XIX	ARRIVAL	Port	PIRAEUS	25/08/2020
AS CAROLINA	ARRIVAL	Port	PIRAEUS	03/09/2020	Not Match	CHRISTOS XLI	ARRIVAL	Port	PIRAEUS	25/08/2020
BLUE CARRIER 1	ARRIVAL	Port	PIRAEUS	04/09/2020	Not Match	ECOSPIRIT	ARRIVAL	Port	PIRAEUS	25/08/2020
BLUE HORIZON	DEPARTURE	Port	PIRAEUS	05/09/2020	Not Match	EKTORAS	ARRIVAL	Port	PIRAEUS	25/08/2020
BLUE STAR 1	DEPARTURE	Port	PIRAEUS	06/09/2020	Not Match	FLYING CAT 5	ARRIVAL	Port	PIRAEUS	25/08/2020
BLUE STAR CHIOS	ARRIVAL	Port	PIRAEUS	07/09/2020	BLUE STAR CHIOS	FLYING DOLPHIN XVII	ARRIVAL	Port	PIRAEUS	25/08/2020
BLUE STAR DELOS	ARRIVAL	Port	PIRAEUS	08/09/2020	Not Match	FLYING DOLPHIN XXIX	ARRIVAL	Port	PIRAEUS	25/08/2020
BLUE STAR NAXOS	ARRIVAL	Port	PIRAEUS	09/09/2020	Not Match	KAPETAN MICHALIS	ARRIVAL	Port	PIRAEUS	25/08/2020
BLUE STAR PAROS	ARRIVAL	Port	PIRAEUS	10/09/2020	Not Match	M Y BILLA	ARRIVAL	Port	PIRAEUS	25/08/2020
CH.GEMITZOGLOU	ARRIVAL	Port	PIRAEUS	11/09/2020	Not Match	MAERSK AHRAM	ARRIVAL	Port	PIRAEUS	25/08/2020
CHRISTOS V	ARRIVAL	Port	PIRAEUS	12/09/2020	CHRISTOS V	NORDSUMMER	ARRIVAL	Port	PIRAEUS	25/08/2020
CHRISTOS XIX	DEPARTURE	Port	PIRAEUS	13/09/2020	CHRISTOS XIX	NUMBER ONE	ARRIVAL	Port	PIRAEUS	25/08/2020
CHRISTOS XLI	DEPARTURE	Port	PIRAEUS	14/09/2020	CHRISTOS XLI	PERSEUS	ARRIVAL	Port	PIRAEUS	25/08/2020
CHRISTOS XXXIII	ARRIVAL	Port	PIRAEUS	15/09/2020	Not Match	POSEIDON	ARRIVAL	Port	PIRAEUS	25/08/2020
COSCO GLORY	ARRIVAL	Port	PIRAEUS	16/09/2020	Not Match	POSIDON HELLAS	ARRIVAL	Port	PIRAEUS	25/08/2020
DIONISIOS SOLOMOS	DEPARTURE	Port	PIRAEUS	17/09/2020	Not Match	PSYTTALIA II	ARRIVAL	Port	PIRAEUS	25/08/2020
ECOKEEPER	DEPARTURE	Port	PIRAEUS	18/09/2020	Not Match	SCIIBulk	ARRIVAL	Port	PIRAEUS	25/08/2020
ECOSPIRIT	DEPARTURE	Port	PIRAEUS	19/09/2020	ECOSPIRIT	SEBECO	ARRIVAL	Port	PIRAEUS	25/08/2020
EKTORAS	ARRIVAL	Port	PIRAEUS	20/09/2020	EKTORAS	SVITZER MORAG	ARRIVAL	Port	PIRAEUS	25/08/2020
ELYROS	ARRIVAL	Port	PIRAEUS	21/09/2020	Not Match	T T CB2	ARRIVAL	Port	PIRAEUS	25/08/2020
FLYING DOLPHIN ATHINA	DEPARTURE	Port	PIRAEUS	22/09/2020	Not Match	TRITONASOII	ARRIVAL	Port	PIRAEUS	25/08/2020
FLYING DOLPHIN XIX	ARRIVAL	Port	PIRAEUS	23/09/2020	Not Match	VERNICOS SIFNOS	ARRIVAL	Port	PIRAEUS	25/08/2020
FLYING DOLPHIN XVII	ARRIVAL	Port	PIRAEUS	24/09/2020	FLYING DOLPHIN XVII	ZAKROSOII	ARRIVAL	Port	PIRAEUS	25/08/2020
FLYING DOLPHIN XXIX	ARRIVAL	Port	PIRAEUS	25/09/2020	FLYING DOLPHIN XXIX					

Figure 7. Comparison of Data between Marine-Traffic and FleetMon.

FleetMon Data						VesselFinder					
Vessel Name	Port Call Type	Port Type	Port At Call	Ata/atd	Comparing Data	Arrival (LT)	Vessel	Built	GT	DWT	Size (m)
AG NEKTARIOS AIGINAS	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match	Aug 25, 12:04	AG NEKTARIOS AIGINAS	2008	10965	12558	140 x 23
ACHAEOS	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match	Aug 25, 12:52	ARCHON MICHAIL	1998	1802	891	86 x 14
AEGEAN ACE	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match	Aug 25, 12:05	BLUE STAR CHIOS				
APOLLON HELLAS	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match	Aug 25, 13:42	CHRISTOS V	-	-	-	15 x 5
ASTARTE	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match	Aug 25, 12:40	CHRISTOS XIX	1985	1309	2245	65 x 15
BLUE STAR CHIOS	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match	Aug 25, 11:57	CHRISTOS XLI				
CHRISTAL MIO	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match	Aug 25, 13:50	FLYING DOLPHIN XXIX	1993	162	16	34 x 6
CHRISTOS V	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match	Aug 25, 13:21	LEON	1983	365	89	34 x 10
CHRISTOS XIX	ARRIVAL	Port	PIRAEUS	25/08/2020	ARCHON MICHAIL	Aug 25, 12:13	PERSEUS	-	-	-	19 x 3
CHRISTOS XLI	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match	Aug 25, 13:11	PILOT BOAT PY54	-	-	-	13 x 4
ECOSPIRIT	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match	Aug 25, 12:58	POSIDON HELLAS	-	-	-	14 x 4
EKTORAS	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match	Aug 25, 11:55	PSYTTALIA II	-	-	-	14 x 4
FLYING CAT 5	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match	Aug 25, 11:05	SEBECO	-	-	-	13 x 4
FLYING DOLPHIN XVII	ARRIVAL	Port	PIRAEUS	25/08/2020	BLUE STAR CHIOS	Aug 25, 11:56	VERNICOS SIFNOS				
FLYING DOLPHIN XXIX	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match						
KAPETAN MICHALIS	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match						
M Y BILLA	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match						
MAERSK AHRAM	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match						
NORDSUMMER	ARRIVAL	Port	PIRAEUS	25/08/2020	CHRISTOS V						
NUMBER ONE	ARRIVAL	Port	PIRAEUS	25/08/2020	CHRISTOS XIX						
PERSEUS	ARRIVAL	Port	PIRAEUS	25/08/2020	CHRISTOS XLI						
POSEIDON	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match						
POSIDON HELLAS	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match						
PSYTTALIA II	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match						
SCIIBulk	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match						
SEBECO	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match						
SVITZER MORAG	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match						
T T CB2	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match						
TRITONASOil	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match						
VERNICOS SIFNOS	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match						
ZAKROSOil	ARRIVAL	Port	PIRAEUS	25/08/2020	Not Match						
					FLYING DOLPHIN XXIX						

Figure 8. Comparison of Data between FleetMon and Vesselfinder.

5. CONCLUSIONS

This paper presents a methodology for using web scraping to mashup maritime data using Python Programming Language and Microsoft Excel as a tool to import, combine and compare data. Mashup is an application development method which can be done be applied in a lightweight manner to mix information and automate processes. There has been a plethora of mashup tools in many shapes and forms [20]. In this paper, we analyzse a spreadsheet-based mashup tool with application to AIS data extracted from various web sources. We believe the spreadsheet environment has inherent advantages over other mashup environment due to its popularity and familiarity with the users [20].

The results of the study confirmed that the above software can help mashup programmers create mashups more efficiently and effectively [16]. We also see indications that it can improve reusability of code. In addition, the results show that using mashup techniques in the maritime surveillance could be adopted in monitoring, comparing, combining, manipulating and analysing Big Maritime data.

Other methods are planned in the near future for web scraping larger amounts of maritime data using Python Programming Language and Web tools to combine and compare big data.

ACKNOWLEDGMENTS

This paper is under the auspices of the activities of the 'ERATOSTHENES: Excellence Research Centre for Earth Surveillance and Space-Based Monitoring of the Environment'- 'EXCELSIOR' project that has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 857510 and from the Government of the Republic of Cyprus through the Directorate General for the European Programmes, Coordination and Development. From 1st of October 2019, the ERC group (Department of Civil Engineering and Geomatics) at the Cyprus University of Technology is on the way to be upgraded to ERATOSTHENES Centre of Excellence (ECoE) through 'EXCELSIOR' H2020 Widespread Teaming project (www.excelsior2020.eu).

REFERENCES

[1] Anjomshoaa, A., Tjoa, A.M. and Hubmer, A., "Combining and integrating advanced it-concepts with semantic web technology mashups architecture case study," In: Nguyen, N.T., Le, M.T., ´Swi _ atek, J. (eds.) ACIIDS 2010. LNCS, vol. 5990, pp. 13–22. Springer, Heidelberg (2010).

[2] Oracle: Information Management and Big Data A Reference Architecture.Tech. Rep. Available online: http://www.oracle.com/technetwork/topics/entarch/articles/info-mgmt-big-data-ref-arch-1902853.pdf (accessed on 30 June 2020).

[3] Analytics, I.B.M., "The real-world use of big data," Tech. rep. (2012).

[4] Hendler, J., "Broad Data: Exploring the Emerging Web of Data," Big Data 1(1),18–20 (2013), Available online: http://online.liebertpub.com/doi/abs/10.1089/big.2013.1506 (accessed on 30 July 2020).

[5] Hopkins, B., Evelson, B., Hopkins, B., Evelson, B., Leaver, S., Moore, C., Cullen, A., Gilpin, M. and Cahill, M., "Expand Your Digital Horizon With Big Data," Tech. rep. (2011).

[6] Anjomshoaa, A. and Tjoa, A. M., "Towards semantic mashup tools for big data analysis," In Information and Communication Technology-EurAsia Conference, pp. 129-138, Springer, Berlin, Heidelberg (2014).

[7] Costa, C. and Santos, M. Y., "Big Data: State-of-the-art concepts, techniques, technologies, modelling approaches and research challenges," (2017).

[8] Hu, J. and Zhang, Y., "Discovering the interdisciplinary nature of Big Data research through social network analysis and visualization," Scientometrics, 112(1), 91-109 (2017).

[9] Yu, J., Benatallah, B., Casati, F. and Daniel, F., "Understanding mashup development," IEEE Internet computing, 12(5), 44-52 (2008).

[10] Fischer, T., Bakalov, F. and Nauerz, A., "An overview of current approaches to mashup generation," In: Proceedings of the International Workshop on Knowledge Services and Mashups (2009).

[11] Hoang, D., Paik, H.Y. and Ngu, A., "Spreadsheet as a generic purpose mashup development environment," In: Maglio, P.P., Weske, M., Yang, J., Fantinato, M. (eds.) ICSOC 2010. LNCS, vol. 6470, pp. 273–287. Springer, Heidelberg (2010).

[12] Tuchinda, R., Szekely, P. and Knoblock, C.A., "Building mashups by example," In: Proceedings of the 13th International Conference on Intelligent User Interfaces, pp. 139–148. ACM (2008).

[13] Mitchell, R., "Web scraping with Python: Collecting more data from the modern web, "O'Reilly Media, Inc., (2015).

[15] Xlwt documentation. Available online: https://xlwt.readthedocs.io/ (accessed on 10 August 2020)

[14] Mei, H. and Liu, D., "Design and Implementation of Network Data Reptile," International Core Journal of Engineering, 5(9), 8-19 (2019).

[16] Kuttal, S. K., Sarma, A., Swearngin, A. and Rothermel, G., "Versioning for mashups–an exploratory study," In International Symposium on End User Development, Springer, Berlin, Heidelberg, pp. 25-41, (2011).

[17] Microsoft Support. Available online: https://support.microsoft.com/en-us/office/power-pivot-powerful-data-analysis-and-data-modeling-in-excel-a9c2c6e2-cc49-4976-a7d7-40896795d045 (accessed on 12 August 2020)

[18] Gupta, S. and Knoblock, C., "Building geospatial mashups to visualize information for crisis management," Proc. ISCRAM 2010 (2010).

[19] McCabe, R., "Site gives landlubbers trove of information about ships," (2015).

[20] Hoang, D. D., Paik, H. Y. and Benatallah, B., "An analysis of spreadsheet-based services mashup," In Proceedings of the Twenty-First Australasian Conference on Database Technologies-Volume 104, pp. 141-150 (2010).