An Online Analytical Processing (OLAP) Database for Agricultural Policy Data: a Greek Case Study

Michael Maliappis[1], Dimitris Kremmydas[2]

[1]Laboratory of Informatics, Department of Agricultural Economics & Rural Development, Agricultural University of Athens
[2]Laboratory of Agribusiness Management, Department of Agricultural Economics & Rural Development, Agricultural University of Athens

Abstract. Statistical data for agricultural policy analysis has certain unique features: a multitude of sources of very different nature; a variety of dimensional granularity; different end user requirements. The utilization of Data Warehouse technology would be valuable for overcoming the above data issues. In this paper, we describe the technologies involved and the data modeling requirements, making an exemplar implementation for few tables of the Greek agricultural census.

Keywords: Agricultural Data, Data warehouse, OLAP, Agricultural Policy.

1 Introduction

Data related to agriculture is of prime importance for agricultural policy research. Based on available data, policy makers are making qualitative judgments and researchers build their models. However, this kind of data bears certain features that need special attention. Firstly there exist many sources of information, e.g. Eurostat, FADN, national surveys, field surveys from universities, etc, none of which should be discarded because agricultural data is actually a scarce resource. Secondly, as a result of the nature of agricultural activity, the related data expand horizontally on many dimensions, e.g. biophysical (what is the effect on the soil of a certain crop, etc.), technical (what inputs a certain crop needs, etc.), economic (what is the cost per hectare for cultivating a certain crop, etc.), social (what is the age distribution of farmers in a certain area). Thirdly, the temporal and spatial dimensions are directly relevant and should always be attached; otherwise data loses its context thus shrinking its quality. Fourthly, almost any of the dimensions of agricultural data is of a hierarchical kind. For example the spatial dimension can extend from a small community to the whole EU and at the same time information regarding the finest geographical scale makes sense to be aggregated. Finally agricultural data for policy analysis are utilized by different kind of users, each having diverse needs. For example for a policy maker it would be sufficient to browse the data while for an

Proceedings of the 7th International Conference on Information and Communication Technologies in Agriculture, Food and Environment (HAICTA 2015), Kavala, Greece, 17-20 September, 2015.

agricultural policy modeller the data would ideally be directly imported to his / her model.

Moreover, the new CAP (2014-2020) is focusing on farm scale measures, thus the need for more low level data is emerging. The problem is that considering the above characteristics of agricultural data, a new approach for storing and presenting them should be considered. The Data Warehouse (DW) approach, followed by On-Line Analytical Processing (OLAP) system for data analysis, seems to be a natural choice (Boulil et al.,2014, Rai et al., 2008). Adoption of DW and usage of OLAP is a mean to move from data to information and then to knowledge.

All the above mentioned data are usually stored into conventional data storage means, following the relational database model. Moreover, they follow their own unrelated and incompatible data storage models. Relational Database models are optimized to handle simple transactions coming from a relative large number of users in real time. This orientation makes them unsuitable or less suitable to organize agricultural data for advanced data analysis. Advances in storage technology provide the means to effectively combine data coming from several incompatible and diverse sources into a DW. The storage structure of DW offers the proper organization of data to implement data analysis tools, such as OLAP, on huge amount of data.

There are several cases where a DW was introduced to agricultural statistical data. One of the earliest attempts was that of the US Department of Agriculture’s National Agricultural Statistics Service (Yost, 2000). Another attempt was that of the development of a central Data Warehouse at Indian Agricultural Statistics Research Institute (IASRI) at New Delhi (Chaturvedi et al., 2006).

In this paper, we propose an initial layout for a DW organizing Greek agricultural data and supporting a minimal implementation of an OLAP system for agricultural policy analysis. The paper describes the process towards the implementation of the DW. Section 2 discusses the technologies involved, section 3 describes the data modeling process and section 4 investigates the several difficulties identified during a case study on Greek Agriculture.

2 Data Warehouse Technology

To provide an effective data analysis for agricultural data several tools and technologies are needed. The data should be obtained from several sources, relational databases or flat files of several formats, transformed and loaded into a Data Warehouse (DW) (Kimball and Ross, 2013). From the DW several data marts can be created as a basis for the desired OLAP cubes and the final data analysis.

A **Data Warehouse** is meant to be the single, integrated, storehouse of data, mainly historical, that can be used for supporting an organization’s decision process. As such, it contains data covering a wide range of topics and business processes, for instance finance, logistics, marketing, and customer support. Often, a data warehouse cannot be accessed directly by end user tools. A **data mart**, in contrast, is meant for direct access by end users and end user tools, and has a limited specific analytical purpose, for instance Retail Sales or Customer Calls.

DW are constructed to answer "who?" and "what?" questions about past events using a huge amount of historical data. The development of DW is usually based on relational data base engines with specialized extensions to handle the intricacies and special needs of DW.

OLAP is a multidimensional view of aggregated data stored in a DW and corresponds to a specific data mart. This view allows analysts and managers to gain insight into data of interest quickly, consistently and with high interaction capabilities. OLAP analysis ranges from basic navigation and browsing, using slice and dice, to statistical analyses, to more serious analyses such as time series and complex modeling.

The implementation of OLAP data analysis is accomplished using **OLAP Cubes**. OLAP cubes are structures designed by using a dimensional model which represents the relationships between facts and dimensions. The facts are the measures of interest that are stored into the DW and dimensions are the qualitative variables concerning these measures. The dimensional model is usually implemented using the star schema. A star schema is a schema that allows the dimension tables to be joined directly with the fact table as is shown if Fig. 1.

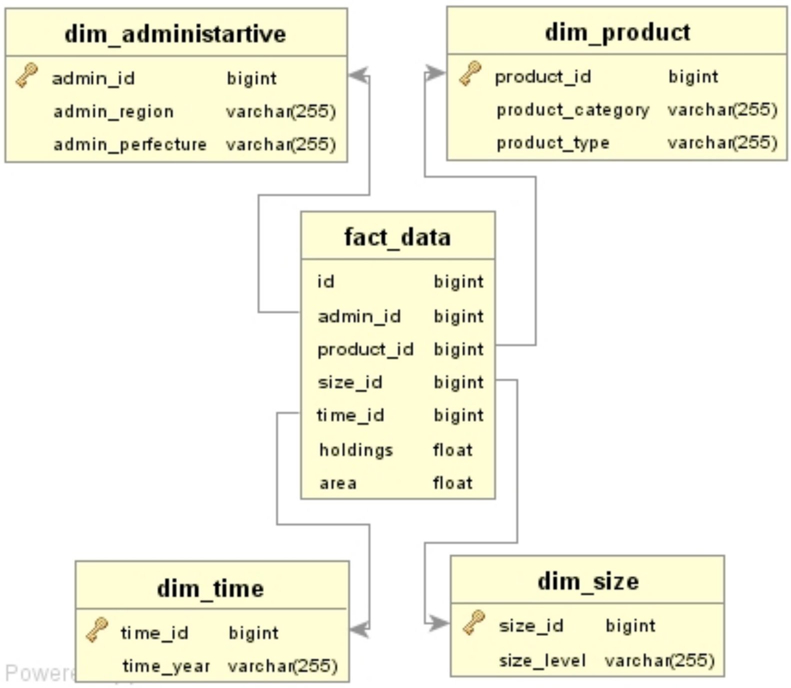

Fig. 1. Star Schema

The structure of OLAP cubes allows easy navigation through the dimensions of data using several operations, such as **slicing** which sets one dimension constant to show a two-dimensional table, **dicing** which creates a sub-cube, **drill down/up** which facilitates navigation from most summarized (up) to more detailed (down) levels and **roll-up** which summarizes the data along a dimension.

The implementation presented in this paper uses MySQL[1] as DW storage data base, Kettle[2] to facilitate collection, transformation and loading of data and Mondrian[3] to create the OLAP cube and apply data analysis. All of the above tools are distributed with free licenses.

3 Data Modeling

The process of DW development is simply the mapping of the source schemas contained in the source data model (structure of the underlying data sources and the relationships between them), to the target schema of the DW model and populate the target tables. This process follows several well defined steps. As is shown in Fig. 2, the data are collected from several sources, extracted in proper form, transformed as needed and loaded into the DW. Using the data of the DW the data marts are created as a basis to OLAP cubes and the other forms of data analysis (Casters et al., 2010).

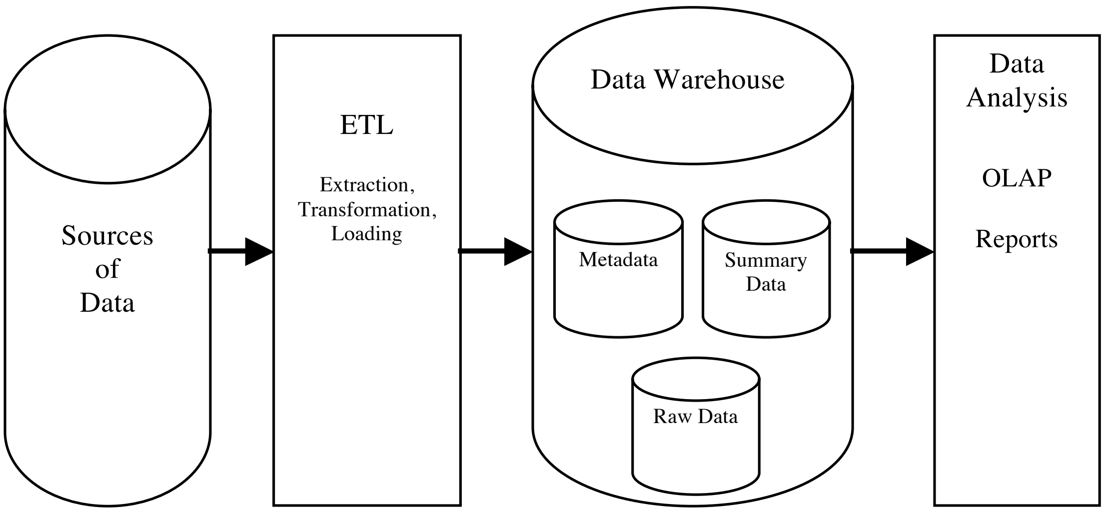

Fig. 2. Data Warehouse Development Process

Identification of sources and their types. The first step towards DW development is the identification of data sources. Usually, the sources are

1 https://www.mysql.com/
2 http://community.pentaho.com/projects/data-integration/
3 http://community.pentaho.com/projects/mondrian/

differentiated according to the mean of storage and the way that they are accessed. Each source, has its individual storage system and a different level of data quality.

ETL (extract, transform, and load) is a set of processes for getting data from several sources, such as OLTP systems, websites, flat files, e-mail databases, spreadsheets, and personal databases, such as Access, into a data warehouse. ETL is also used for loading data marts, generating spreadsheets, scoring customers using data mining models, or even loading forecasts back into OLTP systems. The main ETL steps, can be grouped into three sections:

- **Extract**: All processing required to connect to various data sources, extract the data from these data sources, and make the data available to the subsequent processing steps.
- **Transform**: Any function applied to the extracted data between the extraction from sources and loading into targets. These functions can contain the following operations:
    - Movement of data
    - Validation of data against data quality rules
    - Modification of the content or structure of the data
    - Integration of the data with data from other sources
    - Calculation of derived or aggregated values based on processed data
- **Load**: All processing required to load the data in a target system. This part of the process consists of a lot more than just bulk loading transformed data into a target table. Parts of the loading process include, for instance, surrogate key management and dimension table management.

Collection of large data volumes are a challenge. Extracting all the data from the source systems every time an ETL job is running is not feasible in most circumstances. Therefore there is a need to resolve the issue of identifying what has changed in source systems to be able to retrieve only the data that has been inserted, updated, or deleted. In some cases, this issue cannot be gracefully resolved and a brute force approach needs to be taken that compares the full source data set to the existing data set in the data warehouse.

Other challenges have to do with the way the data needs to be integrated; suppose there are many different systems where statistical data is stored, and the information in these systems is inconsistent or conflicting? How incomplete, inconsistent, or missing data are handled and compiled into a single DW ?

4 A Short Case Study on Greek Agriculture

The sources of statistical information for the Greek agriculture have been compiled on Table 1. The main provider is Hellenic Statistical Authority (EL.STAT) but the Farm Accountancy Data Network (FADN/RICA) is also an important source for microeconomic data on economic activity of farms, though this is limited to a specific range of farm sizes.

Ideally a Greek Agricultural Data Warehouse would use both sources to compile a single Data Warehouse. Such a DW would contain the following dimensions:

Administrative; Temporal; Agricultural Activity (nomenclature) and several different measures, while the OLAP cubes could be divided to themes like technical, economic, environmental, etc. There would also be the need for transforming the information, aligning where possible the granularity of time and space dimensions and also attaching the agricultural activity dimension. A discussion on dealing with such issues is made on Nilakanta et al. (2008).

Here we present a short proof-of-concept case where two tables of the Greek Census of Agricultural and Livestock Holdings (Agr.CENSUS) were parsed, transformed and imported to a DW and simple OLAP cubes were created using free-license tools.

Σελ. 1

Σ Ε Ι Ρ Α Α'　　　Π Ι Ν Α Κ Α Σ 7

Εκμεταλλεύσεις με αρδευόμενες και αρδευθείσες εκτάσεις

Εκτάσεις σε στρέμματα

ΣΥΝΟΛΟ ΕΛΛΑΔΟΣ ΜΕΓΑΛΕΣ ΓΕΩΓΡΑΦΙΚΕΣ ΠΕΡΙΟΧΕΣ ΥΠΑ ΓΕΩΓΡΑΦΙΚΑ ΔΙΑΜΕΡΙΣΜΑΤΑ ΝΟΜΟΙ ΔΗΜΟΙ & ΚΟΙΝΟΤΗΤΕΣ	Εκμεταλλεύσεις			Από αυτές με αρδευόμενες και αρδευθείσες εκτασεις			
	Σύνολο	Από αυτές με καλλι-εργού-μενη έκταση	Εκτάσεις	Εκμε-ταλλευ-σεις	Εκτασεις (περιλαμβανονται δασικες και αλλες μη καλλιεργουμενες εκτάσεις)		
					Συνολικές καλ/μενες	Αρδευ-όμενες	Αρδευ-θείσες
1	2	3	4	5	6	7	8
ΣΥΝΟΛΟ ΧΩΡΑΣ	861623	852466	33514060	565493	24088737	11256165	9383618
Δ Π ΠΕΡΙΦΕΡΕΙΑ ΠΡΩΤΕΥΟΥΣΗΣ	8768	8656	118719	3140	54817	40481	36176
Δ 0 ΛΟΙΠΗ ΣΤΕΡΕΑ ΕΛΛΑΣ & ΕΥΒΟΙΑ	136366	134585	5042993	82602	3572493	1725747	1412875
Δ 1 ΠΕΛΟΠΟΝΝΗΣΟΣ	163952	163609	5663055	101829	3857095	1427901	1183459
Δ 2 ΙΟΝΙΟΙ ΝΗΣΟΙ	29564	29554	690992	17281	437726	44100	33867
Δ 3 ΗΠΕΙΡΟΣ	48870	48237	1170676	36874	886422	405086	378171
Δ 4 ΘΕΣΣΑΛΙΑ	90023	87761	4128896	67037	3343238	2324832	1907569
Δ 5,6 ΜΑΚΕΔΟΝΙΑ	190591	187532	9061368	136233	6497853	3421291	2913163
Δ 7 ΘΡΑΚΗ	49193	48671	2662536	32752	2057623	1031442	861854
Δ 8 ΝΗΣΟΙ ΑΙΓΑΙΟΥ	56729	56437	2163920	28949	1240757	169702	130387
Δ 9 ΚΡΗΤΗ	87567	87424	2810905	58796	2140713	665583	526097
Μ 1 ΒΟΡΕΙΑ ΕΛΛΑΔΑ	329807	323964	15852800	236022	11898714	6777565	5682586
Υ 11 ΑΝΑΤΟΛΙΚΗ ΜΑΚΕΔΟΝΙΑ & ΘΡΑΚΗ	70441	69392	3560577	49996	2827165	1515873	1240884
ΝΟΜΟΣ ΔΡΑΜΑΣ	8499	8152	467147	6594	372544	218679	154134
ΝΟΜΟΣ ΚΑΒΑΛΑΣ	12749	12569	430894	10650	396998	265752	224896

Fig. 3. Raw data format

The Agr.CENSUS is taking place from 1961 every 10 years. We focused on 1991 and 2000 censuses and on data related to grain crops (soft wheat, durum wheat, etc,. table 7B). The first task was to find the source data. The EL.STAT website does not provide the census data in a real database format. One can download the report of the data in pdf format (1991, scanned and bad quality). We received the data of the census after contacting the corresponding EL.STAT office but again those data were not really database data, requiring us to spend time on transforming data to a processable format. Without any knowledge of the underlying IT infrastructure of EL.STAT, it is necessary that the data provided to the public is in a database format.

Table 1. Sources of statistical information for Greek Agriculture

Provider	Data Series Name	Type	Starting Year	Frequency	Geographical Coverage	Finest Geographical Resolution	Data Included	Data Availability
EL.STAT.[1]	Census of Agricultural and Livestock Holdings	Census	1961	every 10 years	Whole of Greece	Municipal districts	number of plant and animal agricultural holdings and their properties regarding their legal status, agricultural land tenure status, structural properties (type of crops / animal / activity), production methods	1961,1971,1981,1991 in printed form 2000,2009 in electronic form
EL.STAT.	Annual Agricultural Statistical Survey	Survey	1961	Annual	Whole of Greece	Municipalities (as defined in the “Kapodistrias” law)	agricultural utilized land per type of crop, volume of agricultural (plant and animal) production, utilization of agricultural machineries	Online from 1961 – 2006
EL.STAT.	Farm Structure Survey	Survey	1966	1966, 1977, since 1983 every 2 years (but not 1991 and 2000), since 2010 every 3 years	Whole of Greece	Municipal districts	number of plant and animal agricultural holdings and their properties regarding their legal status, agricultural land tenure status, structural properties (type of crops / animal / activity), production methods	Online since 2003
EL.STAT.	Survey on Crop Production (inclung permanent cultivations and grapeyards)	Survey		Grapeyards:Yearly survey , grains and other crops / Basic survey every 10 years for grapeyards / research every 5 years for permanent cultivations	Whole of Greece	Prefecture (NUTS-2)	Cultivating area per crop	Online since 2000

Provider	Data Series Name	Type	Starting Year	Frequency	Geographical Coverage	Finest Geographical Resolution	Data Included	Data Availability
EL.STAT.	Agriculture Input and Output Price Index	Index	1967	Monthly	Whole of Greece	760 (output) and 783 (input) price-collection-points, from all Greece	Index of output prices (subsidies and transport costs are excluded) for plant and animal products (as classified in European Economic Accounts) Index of input (products and services) prices	Online since 2001
EL.STAT.	Agriculture production factors' index (Cost Index)	Index	1975	Yearly	Whole of Greece	Whole of Greece / 155 points of price collection points	Index of production factor wage. It is comprised of three sub-indexes: labor (payment for one day), land (rent), and capital (loan interests and agricultural machinery rent)	Online since 2005
EU / MINAGRIC	FADN / RICA	Survey		Annual			Accountancy data	Fine detailed data is not freely distributed. Aggregated data is publicly available.
EUROSTA T	TRADE Database (COMEXT)	Detailed Data	1976	1976 – 1987 is annual, since 1988 is monthly	Whole of Europe	Intra is from direct collection of information from trade operators / Extra is from custom declarations	Value and quantity of goods traded between EU Member States (intra-EU trade) and between Member States and non-EU countries (extra-EU trade)	Since 2004 are free of charge http://ec.europa.eu/eurostat/web/international-trade/data/database

[1] Hellenic Statistical Authority (EL.STAT.)

The raw data received from the statistical office is shown in Fig 3. In order to transform the data to something manageable we pre-processed the tables with regular expression patterns in order to remove non-data characters (like dashes) and then converted the tables to records. The transformed data format is shown in Fig. 4.

	A	B	C	D	E	F
1	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	Μέχρι 4,9 στρ.	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΣΙΤΑΡΙ ΜΑΛΑΚΟ ΚΑΙ ΗΜΙΣΚΛΥΡΟ	164	23
2	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	Μέχρι 4,9 στρ.	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΣΙΤΑΡΙ ΜΑΛΑΚΟ ΚΑΙ ΗΜΙΣΚΛΥΡΟ	1576	60
3	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	Μέχρι 4,9 στρ.	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΣΙΤΑΡΙ ΣΚΛΗΡΟ	428	12
4	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	Μέχρι 4,9 στρ.	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΣΙΤΑΡΙ ΣΚΛΗΡΟ	4225	27
5	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	Μέχρι 4,9 στρ.	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΣΙΚΑΛΗ	36	4
6	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	Μέχρι 4,9 στρ.	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΣΙΚΑΛΗ	278	7
7	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	Μέχρι 4,9 στρ.	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΚΡΙΘΑΡΙ	224	6
8	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	Μέχρι 4,9 στρ.	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΚΡΙΘΑΡΙ	1542	16
9	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	Μέχρι 4,9 στρ.	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΒΡΩΜΗ	137	1
10	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	Μέχρι 4,9 στρ.	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΒΡΩΜΗ	686	2
11	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	Μέχρι 4,9 στρ.	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΑΡΑΒΟΣΙΤΟΣ ΑΜΙΓΗΣ ΓΙΑ ΚΑΡΠΟ	60	10
12	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	Μέχρι 4,9 στρ.	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΑΡΑΒΟΣΙΤΟΣ ΑΜΙΓΗΣ ΓΙΑ ΚΑΡΠΟ	335	23
13	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	Μέχρι 4,9 στρ.	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΡΥΖΙ	4	
14	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	Μέχρι 4,9 στρ.	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΡΥΖΙ	45	
15	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	Μέχρι 4,9 στρ.	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΛΟΙΠΑ ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	10	2
16	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	Μέχρι 4,9 στρ.	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΛΟΙΠΑ ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	32	6
17	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	5 - 9,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΣΙΤΑΡΙ ΜΑΛΑΚΟ ΚΑΙ ΗΜΙΣΚΛΥΡΟ	164	12
18	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	5 - 9,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΣΙΤΑΡΙ ΜΑΛΑΚΟ ΚΑΙ ΗΜΙΣΚΛΥΡΟ	1576	65
19	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	5 - 9,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΣΙΤΑΡΙ ΣΚΛΗΡΟ	428	17
20	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	5 - 9,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΣΙΤΑΡΙ ΣΚΛΗΡΟ	4225	91
21	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	5 - 9,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΣΙΚΑΛΗ	36	2
22	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	5 - 9,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΣΙΚΑΛΗ	278	12
23	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	5 - 9,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΚΡΙΘΑΡΙ	224	8
24	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	5 - 9,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΚΡΙΘΑΡΙ	1542	47
25	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	5 - 9,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΒΡΩΜΗ	137	2
26	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	5 - 9,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΒΡΩΜΗ	686	4
27	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	5 - 9,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΑΡΑΒΟΣΙΤΟΣ ΑΜΙΓΗΣ ΓΙΑ ΚΑΡΠΟ	60	10
28	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	5 - 9,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΑΡΑΒΟΣΙΤΟΣ ΑΜΙΓΗΣ ΓΙΑ ΚΑΡΠΟ	335	53
29	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	5 - 9,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΡΥΖΙ	4	3
30	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	5 - 9,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΡΥΖΙ	45	13
31	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	5 - 9,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΛΟΙΠΑ ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	10	3
32	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	5 - 9,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΛΟΙΠΑ ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	32	8
33	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	10 - 19,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΣΙΤΑΡΙ ΜΑΛΑΚΟ ΚΑΙ ΗΜΙΣΚΛΥΡΟ	164	28
34	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	10 - 19,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΣΙΤΑΡΙ ΜΑΛΑΚΟ ΚΑΙ ΗΜΙΣΚΛΥΡΟ	1576	274
35	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	10 - 19,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΣΙΤΑΡΙ ΣΚΛΗΡΟ	428	37
36	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	10 - 19,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΣΙΤΑΡΙ ΣΚΛΗΡΟ	4225	379
37	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	10 - 19,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΣΙΚΑΛΗ	36	6
38	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	10 - 19,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΣΙΚΑΛΗ	278	53
39	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	10 - 19,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΚΡΙΘΑΡΙ	224	2
40	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	10 - 19,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΚΡΙΘΑΡΙ	1542	15
41	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	10 - 19,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΒΡΩΜΗ	137	4
42	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	10 - 19,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΒΡΩΜΗ	686	27
43	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	10 - 19,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΑΡΑΒΟΣΙΤΟΣ ΑΜΙΓΗΣ ΓΙΑ ΚΑΡΠΟ	60	15
44	ΝΟΜΑΡΧΙΑ ΑΘΗΝΩΝ	10 - 19,9 "	ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	ΑΡΑΒΟΣΙΤΟΣ ΑΜΙΓΗΣ ΓΙΑ ΚΑΡΠΟ	335	105

Fig. 4. Record-format data

Fig.5 and Fig. 6. present two reports coming from the same OLAP cube. The cube has been created using the star schema of Fig. 1. Following this schema, the OLAP cube has been constructed with four dimensions and two measures. Two of the dimensions are flat. The *time* dimension contains only the year corresponding to the data and the *size* dimension represents the different sizes (in hectares) of agricultural holdings, from which the measures are coming. The other two dimensions are hierarchical. The *administrative* dimension contains the regions and the prefectures in each region and the *product* dimension contains the category and the crops in each category. The measures contained into the cube are the number of agricultural holdings and the cultivated area.

Using the appropriate queries to the DW, in a specialized language, it is possible to filter the data according the dimensions and reorder them in any desired manner. The report of Fig.5 shows the cultivated area of several crops for some of the regions and Fig.6 shows the cultivated area for a specific crop for some regions and several holding sizes. What is interesting, with OLAP analysis, is that all these different analyses are accomplished using the same cube and the same set of data.

	Measures				
	Areas				
	Administrative				
Product	All Administratives	Περιφέρεια Ανατολικής Μακεδονίας & Θράκης	Περιφέρεια Δυτικής Ελλάδας	Περιφέρεια Δυτικής Μακεδονίας	Περιφέρεια Θεσσαλίας
ΣΙΤΑΡΙ ΣΚΛΗΡΟ	11,743,728	1,688,125	134,021	1,185,269	1,921,813
ΑΡΑΒΟΣΙΤΟΣ - ΚΑΡΠΟΣ (ΑΜΙΓΗΣ)	3,666,133	1,154,703	605,947	301,510	368,990
ΣΙΤΑΡΙ ΜΑΛΑΚΟ ΚΑΙ ΗΜΙΣΚΛΗΡΟ	3,148,095	480,693	64,121	1,206,374	294,845
ΚΡΙΘΑΡΙ	1,959,958	202,347	179,768	413,645	286,002
ΒΡΩΜΗ	1,216,355	13,127	605,257	4,773	45,194
ΡΥΖΙ	389,418	11,198	13,621	102	915
ΣΙΚΑΛΗ	383,703	22,075	3,297	215,532	26,467
ΛΟΙΠΑ ΣΙΤΗΡΑ ΓΙΑ ΚΑΡΠΟ	49,837	7,515	10,551	1,630	7,986

Fig. 5. Example Report 1 of OLAP Cube

		Measures			
		Areas			
		Size			
Administrative	Product	All Sizes	1 (Μέχρι 4,9)	2 (5 - 9,9)	3 (10 - 19,9)
−All Administratives	ΑΡΑΒΟΣΙΤΟΣ - ΚΑΡΠΟΣ (ΑΜΙΓΗΣ)	3,666,133	107,728	271,567	592,356
+Περιφέρεια Ανατολικής Μακεδονίας & Θράκης	ΑΡΑΒΟΣΙΤΟΣ - ΚΑΡΠΟΣ (ΑΜΙΓΗΣ)	1,154,703	13,578	38,481	108,178
+Περιφέρεια Δυτικής Ελλάδας	ΑΡΑΒΟΣΙΤΟΣ - ΚΑΡΠΟΣ (ΑΜΙΓΗΣ)	605,947	17,789	50,041	124,096
+Περιφέρεια Δυτικής Μακεδονίας	ΑΡΑΒΟΣΙΤΟΣ - ΚΑΡΠΟΣ (ΑΜΙΓΗΣ)	301,510	4,040	11,833	27,063
+Περιφέρεια Θεσσαλίας	ΑΡΑΒΟΣΙΤΟΣ - ΚΑΡΠΟΣ (ΑΜΙΓΗΣ)	368,990	13,784	37,815	81,954
+Περιφέρεια Κεντρικής Μακεδονίας	ΑΡΑΒΟΣΙΤΟΣ - ΚΑΡΠΟΣ (ΑΜΙΓΗΣ)	772,332	16,541	71,658	155,144
+Περιφέρεια Στερεάς Ελλάδας	ΑΡΑΒΟΣΙΤΟΣ - ΚΑΡΠΟΣ (ΑΜΙΓΗΣ)	102,659	9,028	13,278	25,703

Fig. 6. Example Report 2 of OLAP Cube

5 Conclusions

Statistical data for agricultural policy analysis has certain unique features: a multitude of sources of very different nature; a variety of dimensional granularity; different end user requirements. The utilization of Data Warehouse technology would be valuable for overcoming the above data issues.

The first step towards this direction is the detailed reporting of all of the available sources, their properties (dimensions, measures, etc.) and of their availability format. Afterwards the star schema of the DW has to be crafted, containing the required dimensions taking also into account the end-user requirements. Finally the ETL process has to be designed and implemented in order to load data into the DW. There are several license-free tools, making the whole process cost-effective. We followed the above path and made a mini case study for the Greek Agricultural Data. Certain conclusions are coming out.

Primarily the quality of the EL.STAT distributed agricultural data should be vastly improved. Either a DW approach should be incorporated for handling their source data or if this is already the case an OLAP application should go online for disseminating processed information. Also it seems that some of the EL.STAT early historical agricultural data are not available at all in electronic format, which also hardens their handling from researchers.

Secondly, for creating a Greek agricultural DW, the design of the star schema will not be a straightforward process. There are several issues that should be resolved. The administrative division of the Greek territory has changed more than a couple of times and the designer has to align all inter-temporal differences. In order for the OLAP extracted data to be consistent with Eurostat standards, additional information has to be incorporated, like NUTS-to-administrative units mapping and alignment with agricultural activity nomenclature.

Another important issue is the integration of different levels of data detail in the DW. All data sources are referring to some kind of administrative unit and to a specific kind of agricultural activity, and those two could be the connecting element. Micro-level farm data (e.g. the cost of production collected from FADN) could be presented next to more low granular data (e.g. area of a certain cultivation) if those two dimensions were consistent across data sources.

Finally the use of OLAP cubes and Web Services is very important for the usability of the DW. For instance, an agricultural policy model could use a Web Service directly instead of maintaining its own database.

As far as the future work is concerned, the need for a more expanded case study is evident. The consolidation of data from Farm Structural Surveys, Farm Census and FADN data will be very useful to agricultural policy modelers. From the diversity of those data sources will, probably, arise the issues of dimension integration and conflicting or missing data which will have to be addressed.

1. References

2. Boulil, K., Le Ber, F., Bimont, S., Grac, C., and Cernesson, F., (2014) Multidimensional modeling and analysis of large and complex watercourse data: an OLAP-based solution, Ecological Informatics 24 pp.90–106.
3. Casters, M., Bouman, R. and van Dongen, J. (2010) Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data -Integration, Wiley Publishing, Inc.
4. Chaturvedi, K.K., Rai, A., Dubey, V.K. and Malhotra, P.K. (2008) On-line Analytical Processing in Agriculture using Multidimensional Cubes. Journal of the Indian Society of Agricultural Statistics. Vol. 62(1), pp 56-64
5. Kimball, R., Ross, M. (2013) The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3rd ed. John Wiley & Sons, Inc.
6. Nilakanta, S., Scheibe, K., and Rai, A. (2008) Dimensional issues in agricultural data warehouse designs. Comput. Electron. Agric. 60, 2 (March 2008), 263-278. DOI=10.1016/j.compag.2007.09.009 http://dx.doi.org/10.1016/j.compag.2007.09.009
7. Rai, A., Dubey, V., Chaturvedi, K.K. and Malhotra, P.K. (2008) Design and development of data mart for animal resources, Comput. Electron. Agric. 64, pp. 111–119.
8. Yost, M. (2000) Data warehousing and decision support at the National Agricultural Statistics Service. Soc. Sci. Comput. Rev. 18, 4 (October 2000), 434-441. DOI=10.1177/089443930001800406 http://dx.doi.org/10.1177/089443930001800406