Introducing dAUTObase: a first step towards the global scale geoepidemiology of autoimmune syndromes and diseases

Vassiliki A. Gkantouna[1], Manousos E. Kambouris[2], Emmanouil S. Viennas[1], Zafeiria-Marina Ioannou[1], Michael Paraskevas[3], George Lagoumintzis[2], Zoi Zagoriti[2], George P. Patrinos[2], Giannis E. Tzimas[3,*] and Konstantinos Poulas[2,*]

[1]Department of Computer Engineering and Informatics, University of Patras, Rio Patras, 26500, Greece, [2]Department of Pharmacy, University of Patras, Rio Patras 26500, Greece, [3]Department of Computer and Informatics Engineering, Technological Educational Institute of Western Greece, M. Alexandrou 1, Koukouli Patras 26334, Greece

Associate Editor: Janet Kelso

ABSTRACT

Motivation: An autoimmune disorder occurs when the immune system mistakenly attacks and destroys its own healthy body tissues. The initiation of a geoepidemiological database, for recording autoimmune incidents with a focus to clinical manifestations, demographic parameters and geographic background is crucial to detect correlations.

Results: The dAUTObase collects an ever increasing number of publications—currently counting 435—on autoimmune diseases' frequencies in various populations and ethnic groups. The respective data have been hosted by a web application developed for the task. It uses three data visualization tools: the PivotViewer, the Disease Treemap and the Disease World Map, to assist the effective data querying.

Availability and implementation: The dAUTObase 2.0 version (www.biodata.gr/dautobase) needs no registration for querying, but data entry and modification is reserved for registered users (curators-administrators).

Contact: kpoulas@upatras.gr or tzimas@cti.gr

1 INTRODUCTION

The autoimmune diseases result from a dysfunction of the immune system due to which elements of the organism (cells, tissues or whole organs) are subjected to immune attack by agents of its own immune system. The immune response is complicated and entails numerous, different and complex mechanisms of control and regulation; still, its core discriminative power stems from the specificity of the antigen–antibody interaction. Most of the classified autoimmune diseases target nervous tissue, but the cumulative prevalence lies with endocrine system, which is implicated in more high-prevalence autoimmune syndromes (Hayter and Cook, 2012; Invernizzi, 2010).

Autoimmune disorders are mostly of unknown etiology, as they constitute a heterogeneous group as far as demographic profile and primary clinical manifestations are considered, but with a common denominator: the effector of tissue and organ injury is an erroneous immune response to self antigens.

Cases of comorbidity (Invernizzi, 2010) might link more than one results to a single cause, which is in stark contrast to the usual approach of multifactorial diseases, whence a number of causes play a contributory role towards one, given result/syndrome (Cooper *et al.*, 2009). This clustering allowed the notion of 'autoimmune disease class', where syndromes of similar pathogenesis but different anatomic manifestation or epidemiology are grouped together (Hayter, 2012). Considerable evidence supports a role for genetic, environmental and other factors in inducing autoimmune disorders (Baranzini, 2009; Cooper and Stroehla, 2003; Mayes, 1999; Selmi *et al.*, 2009).

The epidemiological indices for the same disease vary between patients from different geographical regions—up to $10\times$, (Youinou *et al.*, 2010)—and in many cases variations are observed within the same country, as happens with psoriasis in Australia were the native (Aborigines) population is much less affected than recent immigrants from Europe (Chandran and Raychaudhuri, 2010).

Geoepidemiology aims at analyzing the spatial distribution of disease incidents to identify risk populations and possible causal factors. The first requirement from epidemiological surveys is to estimate disease incidence in order to focus the investigation on high-incidence areas and subsequently to determine spatial risk factors such as UV and natural radioactivity exposure, inhalation of iodine, ambient temperature and locally available foodstuff (Shapira *et al.*, 2010). Although different geographic distribution is a well-substantiated fact in many cases, such as the almost complete lack of psoriasis in the natives of the Andes mountains compared with the high incidence in Europe amongst native Europeans (Chandran and Raychaudhuri, 2010), rate variation across the map may also reflect differences in the quality of data (e.g. in the diagnosis, classification or reporting of the incidents as exemplified by Cooper *et al.*, 2009) rather than true differences in incidence rates (Best and Wakefield, 1999). Additionally, people and communities tend to cluster in space in systematic ways that may be highly predictive of disease risk (Zeki *et al.*, 2010). Such spatially organized socioeconomic patterns may have important influence on the rates of diseases observed in small areas (Dolk *et al.*, 1995), while communities aggregating different populations show complicated indices: The

*To whom correspondence should be addressed.

Asian-Americans and African-Americans are in higher risk of systemic lupus erythematosus and scleroderma but in lower risk of Type I Diabetes than Americans of European origin, while rheumatoid arthritis shows no variation (Cooper and Stroehla, 2003).

The dAUTObase is the first attempt to register a great number of epidemiological studies on autoimmune diseases, to combine them with geographical and statistical parameters and to visualize the results in order to mine valuable data, that is, relations between specific autoimmune diseases and regions/populations/society mechanics. Such correlations may greatly assist decisions on public health policies, encourage sensible research prioritization and form the basis for a temporal versus spatial/population database of the evolution of such diseases from an epidemiological point of view. Recent advances in data availability and analytic methods have created new opportunities for investigators to improve on the traditional reporting of disease at national or regional scale. They also present new challenges because, as the scale of the investigation becomes broader, local variations in data quality are likely to be shadowed in the greater picture (Youinou *et al.*, 2010), whereas at the small-area scale, these variations could lead to serious biases if not reported.

By employing data mining techniques, a great number of epidemiological studies were screened with initial emphasis on 22 autoimmune diseases. Data mining was performed on specific indices: incidence, prevalence, mortality, age at onset etc. Similar attempts for the compilation and development of epidemiological databases are numerous and proven, but the emphasis of such endeavours rests mostly on infectious diseases (Gray *et al.*, 2009).

To date, there are several data repositories that fall under the banner of autoimmune diseases databases, such as 'The Autoimmune Disease Database' (Karopka *et al.,* 2006) and 'The ARRC Research Database Register' (The Autoimmune Resource & Research Centre Research Database Register. Available at: http://www.autoimmune.org.au/arrc-research-participant-database/.aspx). However, these databases, which are mostly literature derived, focus on the documentation of autoimmune diseases and their causative background, mainly genetic or/and the clinical manifestations; they provide no epidemiological aspect as to population distribution, much less to geographical distribution, national, regional or global. To this end, we are currently experimenting with various visualization techniques in an effort to promote a novel and promising approach in the study of autoimmune diseases, called 'geoepidemiological databases'. The aim of such databases is the study of the descriptive epidemiology of autoimmune diseases and the risk assessment as to their comorbidity in both individual and population levels.

In this work, we present the development of interactive web-based data visualization and querying tools for dAUTObase (http://www.biodata.gr/dautobase). These tools allow users to compare the epidemiological indices of the documented in dAUTObase autoimmune diseases among different populations and thus identify hidden relationships between individual pieces of information (Chandran and Raychaudhuri, 2010; Youinou *et al.*, 2010). More specifically, we have built an elegant web-based multimedia frontend which is based on a software tool launched by Microsoft, namely the PivotViewer (Microsoft

Live Labs—Pivot Available at: http://www.getpivot.com/silverlight/pivotviewer), in order to provide a high-level visualization of the data collection and the mining process. Additionally to the PivotViewer, there are also two alternative visualization querying interfaces; the one based on the Flare visualization toolkit (Flare—Data Visualization For The Web, Available at: http://flare.prefuse.org/) and the JQV map (JQVmap, Available at: http://jqvmap.com/) providing two extra visualization types of the underlying data collection, namely the Disease Treemap and the Disease World Map; these allow users to query disease distributions and correlations among populations (Papadopoulos *et al.*, 2014).

2 MATERIALS AND METHODS

After initial screening of more than 1000 publications in PubMed, a large number (435 and still rising) of epidemiological studies on autoimmune diseases, which were recording the respective frequencies in various populations around the globe, have been selected and manually imported to the database by the curators- the team of Asst Prof. K. Poulas, in the University of Patras. Data were not collected from other databases, in order to simplify the implementation of selection criteria and the importing of records with consistency. Special emphasis was given to the identification of incidence and prevalence rates, while additional information, if available, was registered under the respective fields (number of patients, mortality, morbidity, age at onset). Suitable publications were identified by periodical scans of the PubMed using as keywords the term autoimmune and/or the name of the autoimmune syndrome. Case reports were not included had they not been, according to their authors, a novelty in either of the syndrome/geography or syndrome/population schemes. Moreover, entries were validated and used only when directly stating at least one of the abovementioned epidemiological indices.

2.1 System architecture

The dAUTObase (www.biodata.gr/dautobase) is a web application recording solely epidemiological data of autoimmune diseases in various populations around the globe. It is public and there are no registration requirements for data querying.

The dAUTObase is based on a relational database developed with Microsoft SQL Server (Microsoft SQL Server, Available at: http://www.microsoft.com/sqlserver/en/us/default.aspx). Database records include the population, the geographic region, the relative publication and a number of epidemiological indices like incidence, prevalence, mortality, morbidity, age at onset etc. The entire database schema is depicted in Figure 1. The overall system architecture is based on a three-tier model, which is robust and flexible enough to aggregate multiple information sources and integrate modular development (Eckerson, 1995). This way, dAUTObase supports scalability. As the dAUTObase dataset grows, the expansion of the existing data types will be possible in order to include new data fields. This will be possible upon decision of the curators and the administrator in order to proceed with an upgrade of the database schema.

2.2 Technologies used

The dAUTObase provides a web-based multimedia visualization environment for population-based autoimmune diseases epidemiological data collection and retrieval utilizing state-of-the-art visualization technologies implemented in three alternative data querying and visualization interfaces. The dAUTObase main querying interface is based on Microsoft Silverlight PivotViewer control (Microsoft Live Labs—Pivot Available at: http://www.getpivot.com/silverlight/pivotviewer), a free web browser plug-in that displays full, high-resolution content without long loading
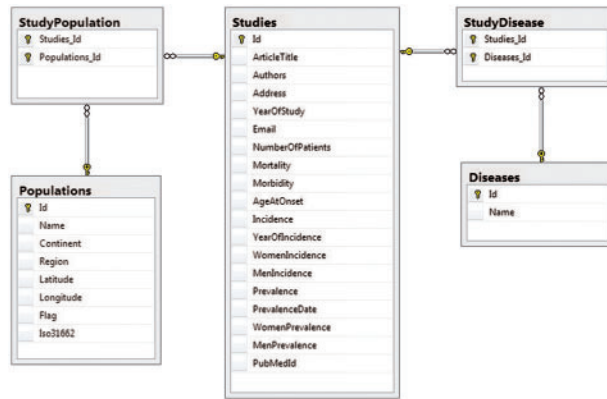
Fig. 1. The data schema of the dAUTObase

Fig. 2. dAUTObase card item: **(A)** A card represents a single Autoimmune Disease in the underlying collection occurring in a specific population. The card consists of the Autoimmune Disease name, information about the incidence and prevalence rate and a flag corresponding to the particular population. **(B)** A sidebar information panel is appearing when users zoom-in on a card providing in-depth information about the card currently selected

times, while the animations and natural transitions provide context and prevent users from feeling overwhelmed by large quantities of information.

The dAUTObase PivotViewer application enables users to interact with thousands of objects at once, and categorize and browse data in multiple ways revealing hidden associations and discovering tracks and trends.

The dAUTObase queries can also be performed utilizing the Flare Visualization Toolkit, an open-source library. Because it includes a wide variety of features, ranging from basic charts to complex interactive graphs, it supports data management, visual encoding, animation and interaction techniques. In dAUTObase, Flare implements an alternative representation of the underlying data collection, the Disease Treemap, an easy way of analyzing large amounts of data in a small space (Treemap: Treemap—Available at: http://www.cs.umd.edu/hcil/treemap/).

Finally, dAUTObase queries output can also be visualized by using an alternative interface based on the JQVMap which enables the change of the look and feel of the maps by adding colours, events, and interactivity. We have utilized the JQVMap technology to implement the Disease World Map.

3   IMPLEMENTATION

Because dAUTObase is the first attempt for a geoepidemiological database for autoimmune diseases, our primary goal was to provide scientists with a powerful tool for the study of the epidemiology of autoimmune diseases. The first step was the selection of relative publications with suitable in volume and context epidemiological data to be included in the dAUTObase. The publications are selected through PubMed scans by the reference of the autoimmune syndrome in their title. The database does not use controlled terminology; the autoimmune syndromes are mentioned and reproduced by the most frequent orthography/ nomenclature forms, which were selected to make the database friendlier to as many users as possible. To this end the 'The Autoimmune Disease Database' (Karopka et al., 2006) is followed. On the other hand, ethnicities are recorded as in the original publications, provided they are not in discord with UN mandates and decisions.

The next step was the efficient visual representation of the collected information in a way that would give researchers the opportunity to explore, analyze and understand the screened data. The three querying interfaces mentioned earlier, all support
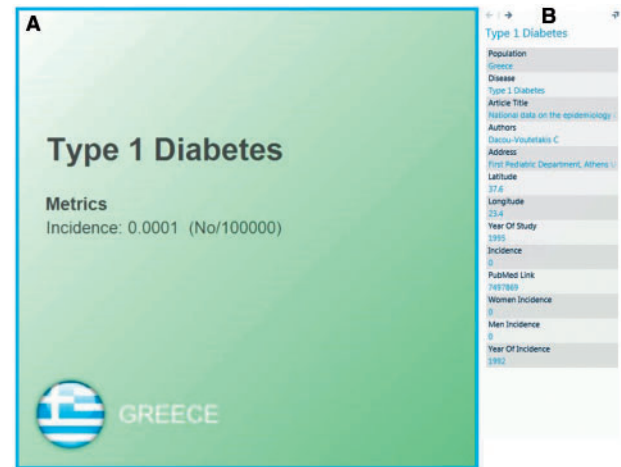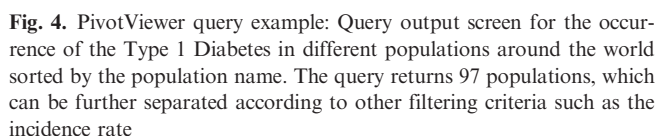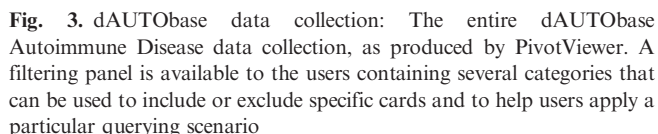
sophisticated data exploration and allow users to apply advanced filtering criteria upon a set of multiple aspects of the same underlying data collection, giving them an active role in the mining process. Moreover, they offer users the possibility to zoom-in from the extensive dAUTObase datasets to particular disease-specific and/or population-specific data. This way, users can detect the links between apparently unassociated and/or remote data entities and handle them in a rather intuitive way reflecting their semantic proximity. It is noteworthy here that user experimentation with different querying scenarios sometimes results in making incidental discoveries of potentially high medical importance, such as the discovery of demographic patterns and the extraction of population-based patient characteristics.

The form of a card (Fig. 2A) is used for the representation of each autoimmune disease, in order to provide a more human-centric visualization approach. This card is accompanied by a sidebar information panel (Fig. 2B) which provides in-depth information on the particular autoimmune disease and population upon users' zooming-in on the card. The entire dAUTObase data collection, as generated by PivotViewer (http://www.biodata.gr/ dautobase/pivot/), is shown in Figure 3. A data filtering panel (displayed on the left side; Fig. 3) is available to offer a variety of filtering criteria applicable to the underlying data collection. Moreover, by clicking on the disease name or the 'PubMed Link' at the sidebar information panel, the user can have access to the original publication of the selected card item. This way, users are able to dynamically sort, organize and categorize data according to common characteristics that can be selected from the data query menu and then zoom-in for a closer look, by either filtering the collection to get a subset of information or clicking on a particular card. There is a 'Download dataset' link at http://www.biodata.gr/dautobase/ allowing users to have full access to the data displayed by the visualizations tools. More specifically, all the recorded studies are

Fig. 3. dAUTObase data collection: The entire dAUTObase Autoimmune Disease data collection, as produced by PivotViewer. A filtering panel is available to the users containing several categories that can be used to include or exclude specific cards and to help users apply a particular querying scenario

Fig. 4. PivotViewer query example: Query output screen for the occurrence of the Type 1 Diabetes in different populations around the world sorted by the population name. The query returns 97 populations, which can be further separated according to other filtering criteria such as the incidence rate

Fig. 5. Disease Treemap: **(A)** the entire dAUTObase Autoimmune Disease data collection, as produced by Flare visualization toolkit. **(B)** Display of the incidence rate of Type 1 Diabetes in Germany. Only the countries/populations with similar incidence (less than 10% difference) are coloured

Fig. 6. Disease World Map: Display of **(A)** the overall, **(B)** the women and **(C)** the men incidence rate of Type 1 Diabetes in different populations around the world. By hovering or clicking on a specific country (in this case China), an additional information panel appears on the left

exported to a Microsoft Excel spreadsheet, including a reference to the respective PubMed Identifier.

Such an example would be the query of the occurrence of the Type I diabetes in different populations around the world, sorted by the population name (Fig. 4).

The dAUTObase data collection can also be launched in an alternative representation, namely the Disease Treemap (http://www.biodata.gr/dautobase/DiseaseTreeMap/) (Fig. 5). It is a treemap representing the incidence rates of autoimmune diseases over populations, with each rectangle representing the specific incidence rate of an autoimmune disease for a specific population. The colour of each rectangle corresponds to a different population. Whenever more than one rectangles are included within a country's rectangle, the subordinate rectangles refer to different autoimmune diseases occurring within the same population or within regions of the same country. The area of each rectangle indicates the incidence rate of the autoimmune disease in the respective population. By clicking on a rectangle, dAUTObase Disease Treemap displays all the relative populations having <10% difference of incidence on average. Additional information, such as the population identity, the disease, the region name and the corresponding value of the incidence of the specified autoimmune disease, become accessible when users hover the mouse over a rectangle. A typical query example is shown in Figure 5B presenting the incidence rate of
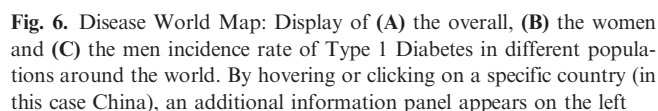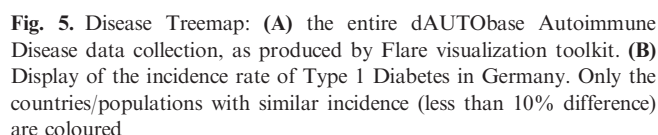
Type 1 Diabetes in different populations and regions around the world.

The Disease World Map (http://www.biodata.gr/dautobase/DiseaseWorldMap/) (Fig. 6) is a visual depiction of the worldwide map enriched with a variety of epidemiological indices per population. Through two drop-down lists, the user has the ability to make a selection of the specific autoimmune disease and the type of epidemiological indices to be displayed. When these selections are made, the Disease World Map dynamically updates the data appearing on the map. The colour scale, corresponding to each country, depends on the screened data and reflects the value of the selected index for the specified disease and population. By hovering or clicking on a specific country, an additional information panel appears on the left, presenting details like the name, the flag and the value of the selected index for the specified population (the selected country is always denoted by the green colour). A typical query example is shown in Figure 6 presenting the incidence rate (among men and women) for Type 1 Diabetes around the world.

4   DISCUSSION

The databases are important tools for the huge load of data gathered around the globe in various fields of medical and biomedical interest: genomics, proteomics, structural biology, epidemiology feature prominently among so many others. The geoepidemiological aspect of autoimmune diseases entails the dynamic to assist in both the elucidation of pathogenetic mechanisms and the emergence of population (Chandran and Raychaudhuri, 2010; Kochi *et al*., 2009; Selmi and Tsuneyama, 2009), ethological (Cooper and Stroehla, 2003; Shapira *et al*., 2010; Youinou *et al*., 2010) and other, community-derived patterns (Shapira *et al*., 2010). Recording incidents meticulously and with an epidemiological focus might help unravel the complex interactions between the various genetic contributors and environmental or social factors that alleviate or bolster such hereditary background (Chandran and Raychaudhuri, 2010). Moreover, better population-scale planning of healthcare services could be achieved (Shapira *et al*., 2010). dAUTObase will be further expanded to become a dynamic tool for all the researchers who deal with autoimmunity. The thorough and systematic study of the epidemiology of autoimmune diseases can pave the way to new perspectives in terms of medical care utilization and patient quality of life. In this work, we present the integration of novel visualization tools within the dAUTObase 2.0 version in an attempt to provide a state-of-the-art data and querying visualization environment for the epidemiological study of autoimmune diseases.

The net utility of the dAUTObase lies in three discrete applications, the audience of which might diverge: (i) The solid recording and compilation of published autoimmune epidemiological data in a single platform and the compilation of the source references so as to be reviewed, compared and downloaded by the user. Such an application, which is the most basic, is helpful in identifying gaps in epidemiological research and vigilance and the design of further research projects by healthcare policymakers, epidemiologists and population biologists/population geneticists. (ii) The second application is the classification and organization of epidemiological data on a geographical basis, were population groups and nationalities are secondary parameters in sorting. This way, a world map of autoimmunity is morphing, for the first time. Through this map commercial and public interests can prioritize the offering of products and services (research and healthcare, the latter in both preventive and therapeutic contexts) in different areas. The third application (iii) is entwined with the very design and the tools provided by the database; the different visualization tools allow access to different levels of compiled data in an intuitive manner and facilitate combination and/or correlation of different databits and bytes in order to form hypotheses, draw conclusions or simply detect tendencies in autoimmune diseases' aetiology and patterns- such as co-morbidity patterns and geography-based incidence biases. Except their potential to increase our basic knowledge by observation and extrapolation such data initiate new fields and contexts for researchers of autoimmune physiology, genetics and therapeutics as much as physicians, health professionals and policymakers.

The current version of the dAUTObase complies to no standards of granulation for ethnicity and other population groups, nor with standardized terminology and systematic reference; it is devoted in recording data published in PubMed-indexed papers, as a measure of quality control. At present, the dAUTObase offers a solid and massive recording of epidemiological indices of autoimmune syndromes sorted by natural and political geography; priority is given to incidence and prevalence, with four fields each (general incidence/prevalence, incidence/prevalence for men and women separately, year of incidence/date of prevalence) but other indices, (such as mortality, morbidity, age at onset) are also included as mentioned earlier.

Although by themselves general incidence and prevalence rates provide a rather poor context of geoepidemiology, they are adequate for initial applications and drawing of rough conclusions. As the database is scalable and additional data types can be imported within a short time upon decision, other indices, to be added as more data become available and recordable, are to further the usefulness of the database to more elegant and specific issues.

Meticulously kept epidemiological records, easily accessible to all interested parties, will allow an 'alert' function whenever a study produces significantly different indices than the ones recorded in the dAUTObase. In such cases, the publication might be subject in a quality control step, and if found valid, scientists and authorities should scrutinize the reason for such deviations. Further still, records will allow comparative studies, where comorbidity tendencies and wide-scale biases might be detected faster and in more intuitive and direct ways than with traditional tools. Of course the granulation will never suffice for establishing comorbidity traits within families or, even less, for substantiating a siblings' concordance and other family bias as in Myasthenia Gravis (Bach, 2012; Zagoriti *et al*., 2014); population-wide traits, though, as the North–South gradient which is instrumental for the Hygiene Theory (Bach, 2012) will be easily detectable, if the number of studies in developing countries offers an *argumentum ex silentio*, meaning that no publications on autoimmune diseases means no such incidents and rather than failure to report them (Best and Wakefield, 1999; Cooper *et al*., 2009).

The issue of comorbidity is just the tip of the iceberg: significant deviations in similar but geographically distant populations

would point to significant environmental or ethological biases and causative agents, whereas similarities in colocated, mixed population groups would advocate against significant genetic contribution in syndrome aetiology. On the contrary, similarity in indices observed in ethnically relative populations located in distant regions or in a dispersed pattern, would point towards an increased genetic contribution to the disease phenotypes; such are Crohn's disease and IPEX-Immunodysregulatory Polyendocrinopathy and Enteropathy X-linked (Bach, 2012). Last, but not least, such retrospective, whole-population queries might provide insight in long term, low incidence effects of different environmental or technological factors in generating disease phenotypes, which in many an autoimmune syndrome are due to acquired deregulation/dysregulation without serious genetic background. Of course, as the database is a recording and sorting tool rather than an exploratory method, its usefulness depends on the input achieved. And this is subject to funding and policy issues, especially among the developing countries. However, the emergence of the dAUTObase enables prompt use of high-volume and high-quality data, as they become published and available, for compiling larger and more coherent pictures on the issue of autoimmune diseases.

The integration of the current visualization tools leaves enough to be desired, as the tools stem from different technologies, a fact synonymous to hardship in seamless integration. However, the preparation of a user manual to guide users through the functionalities of the visualization tools is now in progress. Furthermore, in the near future, we are planning to experiment on more querying options and alternative visualization types and techniques capable of manipulating multidimensional data, in order to create a global visualization framework which is to allow the conduct of multifaceted comparative studies of the querying results by the users.

The administration environment of dAUTObase is another aspect which we consider of great importance. The system supports multiple user profiles, permitting scaled access to the information with users divided into three main groups according to their privileges, namely the administrators, the curators and the simple users. Data entry and modification in dAUTObase is possible only for registered users that is, administrator and curators; the data entered into the system by curators are incorporated into the dAUTObase collection only after their successful approval by the administrator who is responsible for their validity.

Querying needs no registration. This way, researchers worldwide could concurrently contribute the epidemiological data of their country of origin and thus dAUTObase can become the first worldwide up-to-date and comprehensive geoepidemiological data repository. An external user can become a curator so that he/she can upload data in the dAUTObase collection, by contacting the administrator via the website (http://www.biodata.gr/dautobase/). After approval by the administrator, he/she can acquire data entry and modification rights only for those data entered by him/her and under no circumstances can alter data entered by others.

Data mining techniques to further enhance and automate the discovery of valuable information are also due for incorporation and the dAUTObase dataset is going to be exposed through web services based on the open data protocol (Open Data Protocol, Available at: http://www.odata.org/) giving the ability to the research community to freely exploit our data.

Conflict of interest: none declared.

REFERENCES

Baranzini,S.E. (2009) The Genetics of autoimmune diseases: a networked perspective. *Curr. Opin. Immunol.*, **21**, 596–605.

Bach,J.F. (2012) The etiology of immune disease: the case of myasthenia gravis. *NY Acad. Sci.*, **1274**, 33–39.

Best,N.G. and Wakefield,J.C. (1999) Accounting for inaccuracies in population counts and case registration in cancer mapping studies. *J. R. Stat. Soc. Ser. A*, **162**, 363–382.

Chandran,V. and Raychaudhuri,S.P. (2010) Geoepidemiology and environmental factors of psoriasis and psoriatic arthritis. *J. Autoimmun.*, **34**, J314–J321.

Cooper,G.S. and Stroehla,B.C. (2003) The epidemiology of autoimmune diseases. *Autoimmun. Rev.*, **2**, 119–125.

Cooper,G.S. *et al.* (2009) Recent insights in the epidemiology of autoimmune diseases. *Autoimmunity*, **33**, 197–207.

Dolk,H. *et al.* (1995) A standardization approach to the control of socioeconomic confounding in small area studies of environment and health. *J. Epidemiol. Community Health*, **49**, S9–S14.

Eckerson,W.W. (1995) Three tier client/server architecture: achieving scalability, performance, and efficiency in client server applications. *Open Inform. Syst.*, **10**, 3–20.

Gray,D.J. *et al.* (2009) An innovative database for epidemiological field studies of neglected tropical diseases. *PLoS Negl. Trop. Dis.*, **3**, e413.

Hayter,S.M. and Cook,M.C. (2012) Updated assessment of the prevalence, spectrum and case definition of autoimmune disease. *Autoimmun. Rev.*, **11**, 754–765.

Invernizzi,P. (2010) Geoepidemiology of autoimmune liver diseases. *J. Autoimmun.*, **34**, J300–J306.

Karopka,T. *et al.* (2006) The Autoimmune Disease Database: a dynamically compiled literature-derived database. *BMC Bioinformatics*, **7**, 325.

Kochi,Y. *et al.* (2009) Genetics of rheumatoid arthritis: Underlying evidence of ethnic differences. *J. Autoimmun.*, **32**, 158–162.

Mayes,M.D. (1999) Epidemiologic studies of environmental agents and systemic autoimmune diseases. *Environ. Health Perspect.*, **107**, 743–748.

Papadopoulos,P. *et al.* (2014) Developments in FINDbase worldwide database for clinically relevant genomic variation allele frequencies. *Nucl. Acids Res.*, **42** (1), D1020–D1026.

Selmi,C. and Tsuneyama,K. (2009) Nutrition, geoepidemiology, and autoimmunity. *Autoimmun. Rev.*, **9**, A267–A270.

Shapira,Υ. *et al.* (2010) Defining and analyzing geoepidemiology and human autoimmunity. *J. Autoimmun.*, **34**, J168–J177.

Youinou,P. *et al.* (2010) Geoepidemiology and autoimmunity. *J. Autoimmun.*, **34**, J163–J167.

Zagoriti,Z. *et al.* (2014) Recent advances in genetic predisposition of myasthenia gravis. *BioMed. Res. Int.*, **2013**, 1–12.

Zeki,A.A. *et al.* (2010) Geoepidemiology of COPD and idiopathic pulmonary fibrosis. *J. Autoimmun.*, **34**, J327–J338.