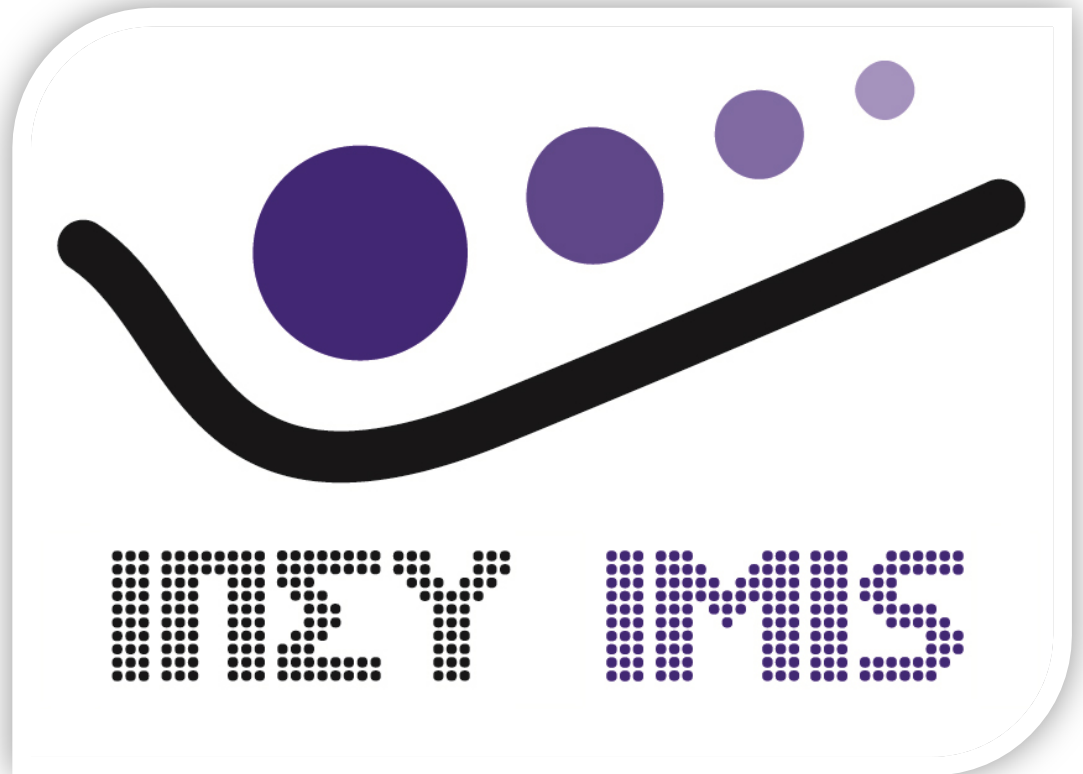
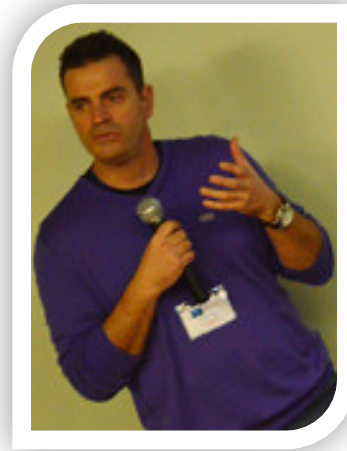


IMIS annual report 2016



Preface

Welcome to the Institute for the Management of Information Systems (IMIS) of the ATHENA Research and Innovation Centre! Established in 2007, IMIS is today one of Greece's premier research centers in the areas of large-scale information systems and Big Data management. Over the past few years, IMIS researchers have been very successful in attracting and implementing numerous cutting-edge research & development projects, at both the national and international level; furthermore, IMIS has created strong collaborative ties with top European research institutions and has successfully promoted the development and use of state-of-the-art information systems in both local industry and various Greek government organizations.



As in previous years, in 2016, IMIS has significantly expanded the scope of its research efforts in a number of focus areas, including Big Data and Scalable Data Analytics, Web Data Management and Semantic Web Technologies, Geospatial Data Management, and Digital Curation and Research Infrastructures. In addition, IMIS researchers have led and/or participated in numerous activities promoting research and educational excellence in the areas of information systems and data management, as well as the development of novel software platforms and services, made available to the research community and employed by both local and international users.

Over the next few years, IMIS aims to continue to strengthen its collaborative ties with local and international industry and academia, promote the transfer of state-of-the-art information technology to national organizations and industry, and continue to strive for excellence further increasing the visibility of its research efforts and results.

Prof. Minos Garofalakis

Director, Institute for the Management of Information Systems (IMIS)

ATHENA Research and Innovation Centre

Athens, Greece, 10/11/2017



Contents

- Preface1
- Profile3
- Infrastructures6
- Highlights8
- Research Directions10
- Projects20
- Publications36
- Dissemination Activities.....41
- Systems.....43
- Education49
- Facts and Figures51
- Contact57



Profile

The Institute for the Management of Information Systems (IMIS) is a research institute within the Research and Innovation Center in Information, Communication and Knowledge Technologies ATHENA, which is the first research center in Greece with a focus exclusively on Information Society. Established in Athens in 2003, it is a research and technology organization supervised by the General Secretariat for Research and Technology of the Ministry of Education and Religion. IMIS was founded in 2007 with the mission to conduct research in the area of data management and large-scale information systems.

The research at IMIS has a strong collaborative aspect, and ranges from basic to applied research. The collaborative aspect is expressed in that research is conducted with national and international partners from industry as well as academia, often also in the context of novel and innovative projects.

The mission of IMIS is to conduct research, develop applications and products and to offer services in the areas of information management and large-scale information systems. IMIS is particularly interested in the areas of database technology, software engineering and development methodologies, managing and exploiting information resources, as well as their large-scale applications.

The objectives of IMIS include:

- Participating and carrying out research and development projects in the area of information technology and to cooperate with the academic community in topics of research, education and knowledge transfer to the industry.
- Producing experimental and industrial prototypes, and developing new innovative products in cooperation with industry.
- Transferring knowledge on research and development topics and offering training and certifications on information technology and related processes for the management of information and information systems.

To attain these objectives, IMIS is structured as follows:

- The Department of Database and Business Intelligence is responsible for basic and applied research in (a) database management systems, and (b) business intelligence systems. The focus is on design, optimization, scalability, security, privacy and high availability of databases and data warehouses, data mining and ETL.



- The Department of Distributed and Web Information Systems conducts basic and applied research on data management problems rising in Web applications and distributed computing in general. Focus is on methods and technologies to support search and exploration on the Web during a creativity cycle: from the abstraction describing the search domain, the information harvest and retrieval tasks and the adaptation of results to user needs, to the classification of results and exchange with other users. One of the use-cases we adopt is based on R&D for Web applications for biosciences. Strong interest also exists for modeling and data management issues for the Semantic Web and ontologies, heterogeneous data source integration problems, web services, and semistructured data storage and querying. Finally, focus is also given on real-time information systems, like sensor networks and p2p/grid systems.
- The Department of Geoinformatics conducts basic research and development of prototypes in the areas of geospatial data management, i.e., Geographic Information Systems, Spatial Data Infrastructures, Location-based services, geospatial data management on the Web, and human-machine interaction with spatiotemporal information.
- The Department of Scientific Databases and Simulation conducts basic and applied research on the modeling and efficient implementation of data management systems that support scientific applications and simulation processes. The focus is on biological data, environmental data, the management of evolution in scientific databases, and the management of imprecise information. Moreover, the Department concentrates on issues of design, development, and management of Digital Libraries.
- The Department of Software Engineering and Project Management Software is responsible for developing innovative applications that integrate cutting-edge research results and responding to needs of the Public and the Private Sector, based on professional standards and methodologies. It also conducts studies and provides consulting services and technical support to entities involved with IMIS. It also offers certifications to individuals and institutions with respect to the proficiency in software development and management of software project lifecycle (technical project description, project contract, project coordination, project delivery). Furthermore, it has the responsibility of training programs in specialized software issues, software development environments and project management software.
- The Department of Digital Curation (DCU) has the mission to conduct research, develop technologies and applications, provide services and training, and act as a national focus point in the field of digital curation. Digital curation encompasses a set of activities aiming at the production of high quality, dependable digital assets; their organization, archiving and long-term preservation; and the generation of added value from digital assets by means of resource-based knowledge elicitation. To ensure the adequate capture of the context of digital resources and their subsequent creative and effective use, the DCU adopts a multidisciplinary approach



that considers the full lifecycle of digital assets, such as records, digital surrogates and scholarly/ scientific datasets.

The activities of IMIS departments are supported by the Accounting Office, the Central Administration Office and the Procurement Office.



Infrastructure

IMSI IT Infrastructure provides the necessary resources and services to support the organizational needs of the Institute and the development and support of quality IT services and solutions. It consists of the hardware, software and network components that are used in order to achieve the above goal.

In an effort to combine the best services, IMSI Infrastructure is built upon both cloud computing and traditional locally implemented solutions. Local IT Infrastructure consists of servers hosted in IMSI owned facilities in a dedicated server room and is largely based on Virtualization technologies in order to achieve server consolidation and maximize the hardware's efficiency. It is the base upon which most of the Institute's IT services are built while at the same time it provides the required resources for a large number of projects the Institute participates in. It is also used by affiliated researchers and students for larger-scale experimentation and research.

Along with maintaining its own private local servers, IMSI also takes advantage of the benefits of cloud infrastructures. The resources made available to the Greek academic and research community through the GRNET Okeanos IAAS Service are heavily used by the Institute's users for research and development purposes. Also, in collaboration with Microsoft and its Academic program Office 365 is used for providing mail and collaboration services.

Through the described infrastructures IMSI provides to its members and affiliates a variety of services, such as:

- Mail Services
- Directory Services used for centralized authentication and authorization
- Source Control
- Shared storage
- Virtual Private Network
- Web publishing
- Project Management and Collaboration
- Communication Services

The above services are provided and implemented using both commercial and open source operating systems and software, such as but not limited to:



- Operating Systems: Debian Linux, Ubuntu Linux, CentOS Linux. Microsoft Windows Server
- Virtualization Software: KVM
- Database Server: PostgreSQL, MySQL, MariaDB
- Distributed Processing: Apache Hadoop

IMSI network infrastructure provides high-speed connectivity to its users and the provided services. IMSI network connects to the Internet using a 1Gbps fiber optics connection to GRNET. It consists of several Gigabit switches that offer wired connectivity and takes advantage of the Research Center's Wireless Infrastructure to provide high speed and reliable Wireless Connectivity. Also, through the Research Center's participation to the eduroam initiative, IMSI members can use their account to gain wireless Internet access in research and academic institutions in more than 70 locations worldwide.



Highlights

Events

- The **final review for the EU (FP7-IP) DIACHRON project** took place in Luxemburg on 24 May 2016, and concluded that the project showed “**Excellent progress**”, which is the highest achievable mark. This result is especially important for IMIS, since IMIS had a leading role in the whole lifetime of the project, from proposal writing through design / planning and finally to implementation, and was one of the main contributors in the collaborative effort. Project scientific coordinator: Yannis Ioannidis. Project coordinator for IMIS: Yannis Stavrakas.
- The **final review for the EU (FP7) GeoKnow project** took place in Luxemburg on 21 January 2016, and concluded that the project showed “**Excellent progress**”. GeoKnow’s results further strengthened, and disseminated to third parties, the scientific and technological expertise that IMIS has built on the fields of Linked Data and Geospatial Data Management. The project’s outcomes are further exploited and extended during the EU (Horizon2020 – IA) project SLIPO. Project coordinator for IMIS: Spiros Athanasiou.
- The EU (H2020-MSCA-ITN-2015) **project ARCADES started** on January 1st, 2016. The coordinating member of the project is IMIS. Scientific coordinator: Ioannis Emiris.
- The project ARCADES has successfully completed its recruitment campaign in 2016 with the **selection of 13 ESR fellows** out of a total of 54 candidates. Marie Sklodowska-Curie fellows must satisfy a number of mobility requirements and have less than 4 years experience since their first University degree. Moreover, funding of MSCA Networks is dependant upon the successful continuation of the fellows' doctoral studies. The recruitment campaign comprised of a Kick-off meeting and Recruitment Event at T.U. Wien, attended by 18 international candidates, and an open call widely advertised in a variety of platforms and media including the Euraxess website, announcements in local and international mailing lists. IMIS has recruited two ESR fellows in May and October 2016, respectively.
- The **final review for the National (“SYNERGASIA”) AI4B project** took place at ATHENA R.C. on 18 July 2016, and was successful in every aspect. IMIS was the coordinating partner for this project. Project scientific coordinator: Theodore Dalamagas.
- The **final review for the National (“SYNERGASIA”) RealEstate2.0 project** took place at ATHENA R.C. on 25 April 2016, and was successful in every aspect. IMIS was the coordinating partner for this project. Project scientific coordinator: Yannis Stavrakas.
- The **final review for the National (“KRIPIS”) SoDaMap project** took place at the National Centre for Social Research on 27th of May 2016 and was successful. IMIS was one of the two partners in this project (along with NCSR), responsible for the



development of two tools: www.linked-statistics.gr and www.socioscope.gr. Project coordinator for IMIS: George Papastefanatos.

- The **final review for the National (“KRIPIS”) MEDA: Efficient Management for Big Data – Challenges, Methods and Techniques** project took place at ATHENA R.C. on the 23rd of May 2016 and was successful. The project was carried out by IMIS, and the scientific coordinators were Timos Sellis and Manolis Terrovitis.

Awards

- **Best paper award:** Gavrilis D., Nomikos V., Kravvaritis K., Angelis S., Papatheodorou C., Constantopoulos P. (2016). MORE: A Micro-service Oriented Aggregator, Metadata and Semantic Research. Proceedings 10th Int Conf Communications in Computer and Information Science -CCIS (No. 672: pp. 15-26), Göttingen, Germany: Springer-Verlag.

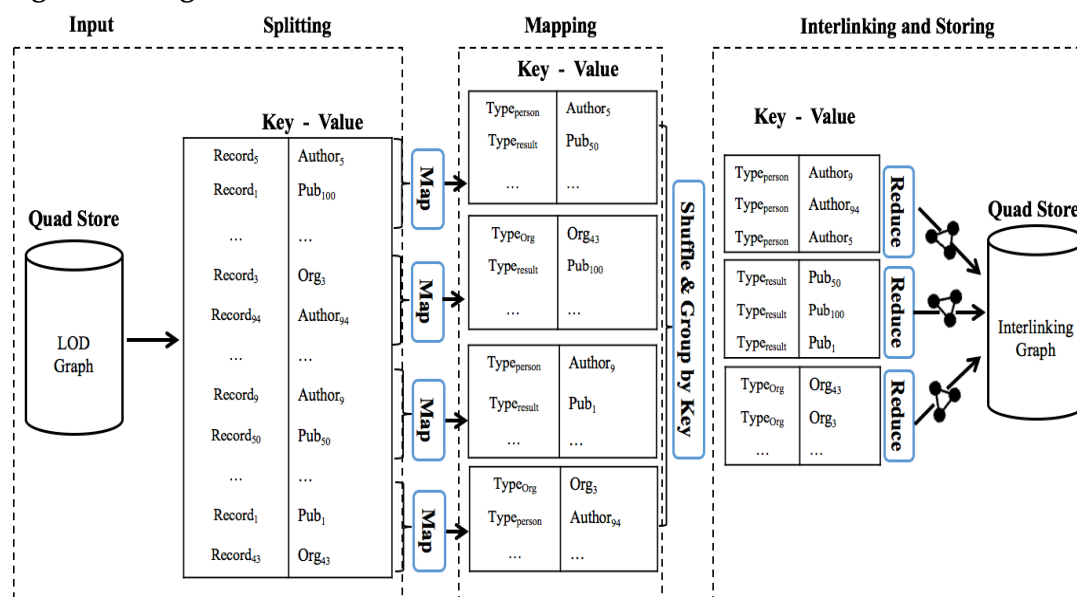


Research Directions

Big Data and Scalable Data Analytics

Current (ongoing) research directions in this area include:

- **Big Data Integration**



An area of interest concerns *Entity Resolution* in Big Data Integration settings, which can be used for duplicate detection and entity interlinking. Entity resolution constitutes a crucial task for many applications that involve integration of heterogeneous and noisy data; but it has an inherently quadratic complexity making brute-force methods inefficient to scale in big datasets. In order to enable entity resolution to scale to large volumes of data, blocking (cluster similar entities into groups and perform comparison only within groups) and metablocking techniques (clean overlapping blocks) are employed. In this direction, we have proposed novel metablocking techniques that aim to increase precision by orders of magnitude at a small cost in recall and exploited parallelization techniques (Map-reduce and Spark) to further improve the overall performance (in terms of efficiency and accuracy) of data integration. These methods have been implemented into a scalable framework for interlinking scholarly data, in the context of the *OpenAire* project. Still these methods operate in a more or less static mode, without considering fast updates on the data, or diverse user needs on the quality or the accuracy of the matched results.

- **Data Visualization & Visual Analytics**

One of the major challenges in the Big Data era is that it has realized the availability of a great amount and variety of big datasets for analysis by non-expert data

analysts, such as research scientists, data journalists, policy makers, SMEs and individuals. The level of difficulty in transforming a data-curious user into someone who can competently access, analyze and consume that data is even more burdensome now for a great number of users with little or no support and expertise on the data (pre-) processing part. Thus, data exploration and visualization methods and tools for visual analytics are of great importance nowadays. Modern systems must address the challenge of on-the-fly scalable visualizations over large dynamic sets of data, offering efficient exploration techniques, as well as mechanisms for information abstraction, summarization and meaningful visualization of different types of data (graph, geo, biological, etc). In this context, we have developed a tool, called *GraphVizDB* that enables the visualization and exploration of very large graphs using spatial indexing techniques. Furthermore, we have developed *Socioscope* (www.socioscope.gr), a visual analysis tool, used by social scientists, for the visualization and exploration of social and political data. Another application developed within the project *CitySense* aims to integrate city data from disparate sources and provide a visual way to combine them and filter city areas.

- **Mathematical Modeling and Analysis**

Mathematical modeling of the physical world is crucial in a number of applications. Despite the complexity of three-dimensional models, current algorithms and software are making enormous progress in efficiently representing, handling, exchanging, and operating on such models. The ARCADES Network contributes in this direction by exploiting cutting-edge research in mathematics and algorithm design so as to design and implement robust methods in Computer-Aided Design and manufacturing (CAD/CAM). However, geometric modeling is facing new challenges in modern engineering analysis, simulation, manufacturing, and construction. This is becoming evident in new sectors such as the movie and game industry, where CAD methods are not penetrating fast enough, or are facing new challenges arising from massive and fast point acquisition (e.g. by laser scanners), big data and mobile computing. This captures precisely the challenge taken up by the ARCADES Network, namely to build the next generation of CAD software based on strong mathematical foundations from computer algebra, geometric computing, numerical analysis, and algorithm design. The crux of our method relies on algebraic representations, understood in the widest sense of the term, namely relying on polynomial expressions, and including parametric, implicit, and semi-algebraic representations, which can be converted to/from further representations such as point clouds or subdivision surfaces. Compared to classical discrete representations, algebraic representations not only drastically reduce the size of the data, but also provide a small number of parameters (e.g. control points). The Network participants represent a multidisciplinary and multisectoral spectrum for implementing this vision, thus also offering an excellent opportunity for career development to the ARCADES fellows.



- **Scalable Analytics for Social Data**

The increasing use of online social networks and microblogging platforms, such as Facebook and Twitter, and the content generated by millions of users, has led to the development of a multitude of innovative applications based on the analysis of big data from social networks. The selection of an appropriate sample for each different application data analysis is critical to the quality of the analysis results. In this direction, technologies supporting efficient collection and modeling of social network data are key components of an innovative information management platform for social networks. Such a platform fills the gap that exists between the restrictive and cumbersome interface offered by online social networks and the need of applications for easy collection and flexible sample selection. At IMIS, we are developing a novel platform, called *TwitHoard*, for managing information obtained from OSNs and we incorporating in it the results of relevant research activities. Moreover, the National (Synergasia) project *RealEstate2.0* develops an infrastructure for collecting and integrating comments about specific geographical areas from various social media platforms.

- **Influence Maximization in Social Networks**

Influence maximization in social networks focuses on how to distribute a message in a social network to reach the widest possible audience. While traditional methods investigate the choice of the most suitable users as broadcasters of the initial message, our work focuses on choosing the best features (i.e., interest topics) that can work as communication channels.

- **Privacy Preservation**

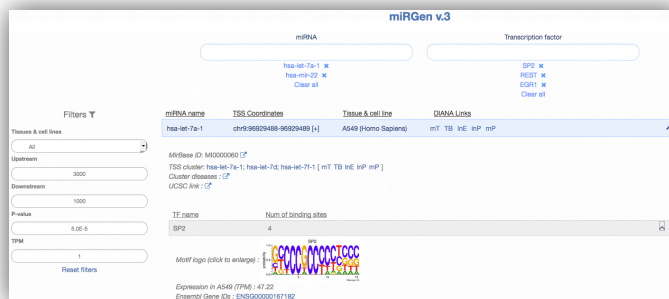
The protection of user privacy in data analytics is one of the major challenges faces by modern information systems. IMSI has worked extensively in the field of data anonymization, where personal data are transformed to anonymous where the identities of individuals are hidden and sensitive properties can no longer be attributes to them. IMSI has developed a series of anonymization algorithms for complex data (tree structured, RDF, set-values) and a tool, AMNESIA, that allows non-expert users to use the, The work on privacy preservation has resulted in a series of publication and it is used in the *OpenAIRE*, *My Health My Data* and *MEDA* projects.

- **Scientific databases and bioinformatics**

We work on models, architectures and methods to store, preserve, process and query genomic and life science data. The vision is to provide high-performance computing methods and tools to perform real-time analysis and processing of big and complex life science datasets. Emphasis is given to genomic data related to miRNAs sequencing process and their analysis. IMIS, Univ. of Thessaly and GRNET have designed, implemented and maintained a research infrastructure for genomic data management, oriented to processing, analysis and visualization of computationally predicted miRNA targets (<http://diana.imis.athena-innovation.gr>). Part of the work is being integrated into the ELIXIR-GR research infrastructure that develops the

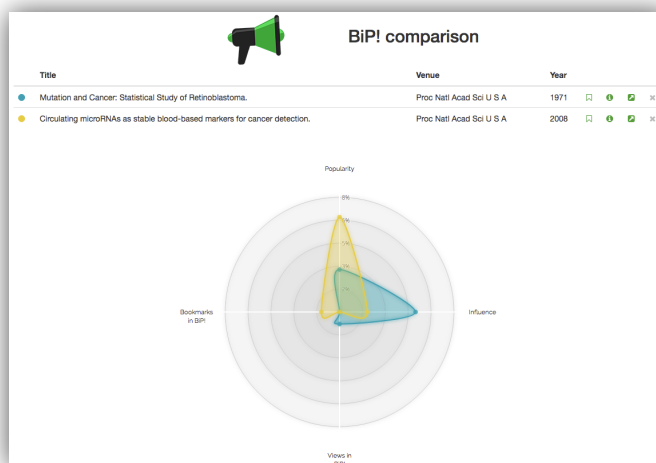


Greek Node of the ESFRI European RI ELIXIR (<https://www.elixir-europe.org>), a distributed e-Infrastructure aiming to build a sustainable European infrastructure for biological information. The research is supported by EXCELERATE, an H2020-INFRADEV project for fast-track ELIXIR implementation and ELIXIR-GR, a Greek project of the National Structural fund to support the Greek Roadmap for Research Infrastructures.



- **Text Mining and Information Retrieval**

Scientific paper retrieval and ranking based on qualitative aspects has been a long-established research topic, especially due to the massive number of papers being published annually. We are working on methods and algorithms to identify high-quality research output, focusing on the study of aspects like paper influence, hype, etc. Also, we work on PageRank-based models to capture citation trends and design ranking methods to promote high-quality papers that have not yet received sufficient citations.



- **Information and Communications Technologies for Big Data**

Specifically broadband networks and converged telecommunications and media services, cloud based large data processing, storage and network programming systems. The research is supported by EU and National projects related to scientific data infrastructures, currently ELIXIR-EXCELERATE and ELIXIR-GR. Note that IMIS, in collaboration with GRNET and Univ. of Thessaly, has designed, implemented and maintained an e-science platform for genomic data management, oriented to processing, analysis and visualization of computationally predicted miRNA targets (<http://diana.imis.athena-innovation.gr>). The platform is running on GRNET's cloud, providing reliable and high-performance data analysis and processing services to more than 2K active users.

New (emerging) research directions in this area include:

- **Big Data Analytics**

The attention Data Science receives recently is partly due to the availability of huge volumes of data and the opportunity to mine and extract useful knowledge. IMIS has already a significant activity and expertise around Big Data management, thus a natural next step is to focus on Data Science and Big Data Analysis. An interesting objective for IMIS is to examine how the machine learning techniques and statistical methods of Data Science can be combined with the Big Data need for scaling, particularly for non-conventional data types, like temporal, spatial, graph, stream, and scientific. We also plan to investigate the processing of deep learning operations (e.g., matrix factorization) in modern CPU architectures.

- **Data-driven Circular Economy**

Based on the successful results of AI4B (www.ai4b.gr/), a national R&D project that designed and developed innovative ICT for biomass supply chains, we plan to launch a multi-disciplinary research and innovation activity in data-driven approaches to unlock the circular economy potential. Contrary to the 'take, make, dispose' production model of a linear economy, in a circular economy, resource input, waste, emission, and energy leakage are minimized by narrowing material and energy loops through recycling, reuse, remanufacturing, repair, etc. Our aim is to build on state-of-the-art technologies and scientific solutions in data science and big data analytics, and provide novel methods, algorithms and tools to collect, process and analyze data to support key areas of circular economy, like smart waste management, automatic detection of resource/material flows in industrial networks, analysis and visualization of complex industrial symbiosis networks, etc.

- **Cloud Infrastructures**

Based on the experience from CONFES project that developed integrated wireless-wired transmission network infrastructure of ultra-high capacity optical technologies, project proposals are submitted to address challenges of the Future Internet especially in Network Function Virtualization - NFV / Software defined Networks –SDN for example by elaborating and constructing a dynamic Cloud infrastructure and in particular a converged Telco and IT node in access networks. One goal is to support advanced 5G/IoT services and applications with demanding QoS and edge analytics needs, with applications in many areas of great interest like in precision farming and 5G access and backhaul networks.

Web Data Management and Semantic Web Technologies

Current (ongoing) research directions in this area include:

- **Data Web, Semantic Annotation and Retrieval**



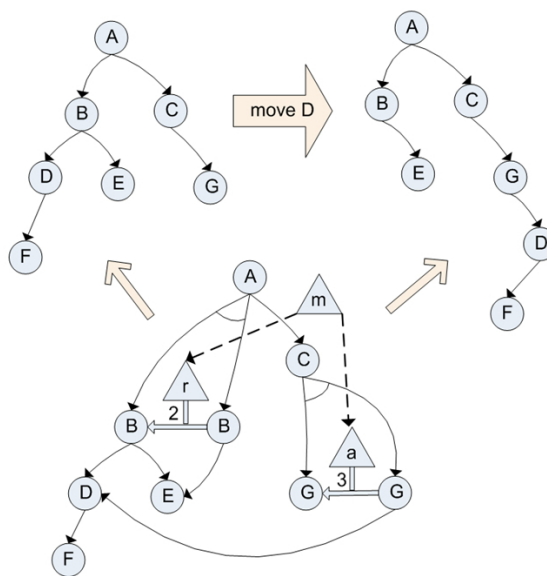
We study models and algorithms to support search facilities for content, structure, metadata, provenance and revision history of Web resources. Based on semantic enrichment technologies, we work on search methods that assist users to explore content without the need for an a-priori knowledge of schema information and query language syntax. Our approach can be used for keyword-based search on graph-structured/RDF data, where instead of providing answers directly from the RDF data graph, we generate automatically a set of candidate SPARQL queries that try to capture users information needs as expressed by the keywords used. Our approach is tailored to evolving RDF data. The research is supported by SLIDEWIKI, an H2020 IA project with large-scale pilots for collaborative OpenCourseWare authoring, multiplatform delivery and learning analytics.

- **OLAP analytics in the Web of Data**

Another direction of research is related to the application of efficient OLAP analytics on multidimensional RDF data, i.e., data usually treated under the OLAP prism, where they are represented as observations that are instantiated over pre-defined dimensions and measures (similar to traditional DW modeling). The increasing volume and diversity of these data (statistical authorities, academic institutes, financial organizations and pharmaceutical companies publish such data) pose the challenge of finding hidden relations between them in a most efficient and accurate way, with the aim to detect inconsistencies or infer new facts. We have addressed this problem, by introducing new relationships (e.g., containment and complementarity of data) between multidimensional RDF data, and new algorithms for efficient and scalable computation of these relationships.

- **Dynamics of the web of data**

In recent years, the Web of Data has been established as a vast source of data from diverse domains, such as biology, statistics, finance, science, and health. As it grows bigger and wider in range, new challenges arise related to the dynamicity, the heterogeneity, and the volume that characterize it. One of the research directions is related to the dynamics of the web of data. The management of evolving information in such a decentralized setting is quite challenging arising new problems related to the archiving and preservation of interlinked information, temporal modeling & evolution management (change detection and propagation) as well as benchmarking techniques in this area. The DIACHRON



project (ended in 2016) addressed many of the above issues, proposing a new model for representing and archiving evolving (diachronic) linked data, a new temporal query language for retrieving information and a benchmarking framework for assessing web data archives. In our approach, changes are discrete objects that have complex structure and retain their semantic and temporal characteristics, rather than being isolated low-level transformations on data.

- **Scalable Query Processing in the Web of Data**

Another line of work addresses the need for efficient processing of SPARQL queries over voluminous RDF stores. Many indexing approaches have been proposed in this area; still few of them take into account the inherent structure of RDF graph data and how this structure can be exploited for efficient processing and optimization of SPARQL queries. For that, we have developed a scalable approach for query processing of RDF stores, based on a novel indexing technique, called Extended Characteristic Sets, which builds on top of the characteristics of the triples in an RDF dataset and enables the efficient processing of complex multi-join queries.

New (emerging) research directions in this area include:

- **Distributed Storage and Parallel Query Execution**

Future interesting directions involve the distributed storage of the index and the parallel execution of such queries. Future directions include the extension of these techniques to cover more complex analytic techniques such as finding missing values, outliers and causalities on such data and make them scalable for very large volumes via parallelization. We are also working on online entity resolution techniques, which aim at integrating the blocking, metablocking, and entity matching tasks (in the form of query operators) in the query-processing phase. The goal is to enable users to seamlessly analyze on query-time heterogeneous (e.g., different schemas, disparate data sources) datasets that involve highly noisy data of different quality.

Geospatial Data Management

Current (ongoing) research directions in this area include:

- **Identifying Areas of Interest**

Large amounts of user-generated content are becoming available on the Web daily, with an increasing portion of it being associated with geospatial information. Typical examples include community-based mapping projects, such as OpenStreetMap, databases of Points of Interest (POIs), e.g., from Wikipedia or Foursquare, geotagged photos, e.g., from Flickr or Instagram, etc. This information constitutes a valuable resource for discovering and exploring locations and areas of interest, with numerous applications in location-based services, geomarketing, trip planning, and



other domains. Our work in this context has focused on identifying and describing streets of interest for a set of query keywords specified by the user. This is a particular type of Area of Interest, where the shape of the identified area is constrained by the underlying road network. More specifically, given a set of keywords or POI categories, we identify the top-k streets or street segments relevant to those. Moreover, we describe these street segments by computing a small and diversified set of geotagged photos located in their vicinity. The results can assist, for example, a tourist or more generally a visitor of a city to quickly identify the most relevant locations for a given need or task.

- **Spatio-Textual Similarity Search**

Nowadays, an increasing number of users accesses social networks via mobile devices, allowing location tracking services to geotag their posts. This information can be used to associate each user with a spatio-textual footprint, indicating the locations a user has visited and some respective textual content (e.g., tags, keywords) she has posted in those locations. Based on this, an application can identify pairs or clusters of similar users, which can subsequently be used, for example, for friend or content recommendations. Our work in this area has focused on similarity search for spatio-textual point sets. We build upon the state of the art for spatio-textual join queries over single points, generalizing this to sets of spatio-textual objects. We have proposed efficient algorithms that can identify all matching pairs with respect to a given similarity threshold or the top-k most similar users. We have also extended our methods to detecting clusters of similar users, which can be used, for example, for discovering spatio-textual communities in geosocial networks.

- **Anonymization of Spatial Data**

Our work has been directed towards the anonymization of trajectories. The main innovation of our proposal was the use of splitting a trajectory to smaller ones, as a data transformation operator in data anonymization.

New (emerging) research directions in this area include:

- **Querying and Analyzing Streaming Spatio-Textual Data**

Querying and analyzing streams of user-generated geotagged posts in social networks is important in a wide range of applications, such as event or topic detection, sentiment analysis, and opinion mining. Moreover, users often need to browse and navigate across content in microblogs to track and monitor the evolution of events and stories as they unfold in the dimensions of space and time. A core functionality for any such analysis and exploratory search is spatio-temporal keyword querying, i.e. queries including filters in each of these three dimensions: text, space and time. To that end, we have introduced and studied the top-k Coverage and Diversity aware Spatio-Temporal Keyword query. Specifying a spatial area, a time interval, and a set of keywords, this query return a ranked list of top-k posts where the ranking is driven by the spatio-temporal distribution of all available



posts satisfying those filters. We formally define the criteria of spatio-temporal coverage and diversity, and propose efficient algorithms for evaluating this type of query.

Digital Curation and Research Infrastructures

Current (ongoing) research directions in this area include:

- **Digital Curation**

Digital curation encompasses a set of activities aiming at the production of high quality, dependable digital assets; their organization, archiving and long-term preservation; and the generation of added value from digital assets by means of resource-based knowledge elicitation. In order to ensure adequate representation and long-term access to digital information as its context of use changes we adopt a lifecycle approach to the representation of curated information objects and a cross-disciplinary scope, so as to cater for differences between diverse scientific and functional contexts of use. Our work agenda includes conducting research, developing innovative systems and services, and promoting best practices for digital curation, with special focus on selected domains. Lines of research and development work include: modeling and managing digital curation and preservation processes; developing and maintaining knowledge resources and knowledge organization systems, repository systems, registries, aggregation services, curation micro-services, metadata quality assurance; ontologies, semantic search and process discovery; understanding and modeling working practices and information behaviors in specific research and professional communities, especially in the humanities; user community modeling and social tagging; streamlining curation and scientific communication; establishing and disseminating best practices with regard to data curation and management.

- **Digital research infrastructures**

A strategic action line is the development of digital research infrastructures for the humanities at the national and European levels. To this end, IMIS actively participates in the European Digital Research Infrastructure for the Arts and Humanities (DARIAH) since the preparatory phase, currently with leading roles in VCC2, DARIAH's Virtual Competence Centre for Research and Education (A. Benardou is VCC2 Co-Chair, C. Dallas chairs the Working Group "Digital Methods and Practices Observatory"); and is the prime technical partner in DARIAH-GR. IMIS is involved in building thematic infrastructures, such ARIADNE for archaeology, provides advanced aggregation services for the Europeana ecosystem, and actively contributes to the Research Data Alliance (RDA). It has developed the ESF NeDiMAH methods ontology (NeMO), now used to drive the automatic extraction of research processes. Collaborations with other research infrastructures, also in areas besides the humanities, are actively pursued.



New (emerging) research directions in this area include:

- **Electronic Records Management**

We plan to work on curation services for electronic records management. Records are important both for their content and as evidence of communications, decisions and actions. The development of electronic records management systems, especially tools for metadata management, classification, semantic annotation and preservation, constitutes a promising direction for applying our work.



Projects

EU R&D Projects



ARIADNE - Advanced Research Infrastructure for Archaeological Dataset Networking in Europe

Project manager	Costis Dallas/ Chistos Papatheodorou
Coordinator	PIN Soc. Cons. A R.L. – Servizi Didattici e Scientifici Per L' Universtita Di Firenze
IMIS - funding	537,440 euros
Programme	FT4-Infrastructures -2012-1
Start date	1/2/2013
Duration	4 years
Website	http://www.ariadne-infrastructure.eu/

The ARIADNE project is aimed to integrate the existing archaeological research data infrastructures so that researchers can use the various distributed datasets and new and powerful technologies as an integral component of the archaeological research methodology. DCU led Work Package WP3, which was responsible for: a metadata registry for archaeological resources, the design and implementation of long term preservation services, and the implementation of the interoperability framework of the ARIADNE infrastructure. It also organized two Summer Schools and Expert Forums on digital curation and digital research in archaeology.




Europeana Cloud

Project manager	Costis Dallas
Coordinator	The Europeana Foundation



IMIS - funding	154.533 euros
Programme	CEF-TC-2015-1-01
Start date	1/2/2013
Duration	4 years
Website	http://pro.europeana.eu/europeana-cloud

Europeana Cloud was a Best Practice Network coordinated by the Europeana Foundation, designed to establish a cloud-based system for Europeana and its aggregators. Europeana Cloud provided new content, new metadata, a new linked storage system, new tools and services for researchers and a new platform: Europeana Research, an initiative facilitating the use of Europeana Collections in research. In Europeana Cloud, DCU led the Work Package responsible for Assessing Researcher Needs in the Cloud and Ensuring Community Engagement, which was responsible for: a) The identification and definition of the Humanities and Social Sciences research communities that will be supported via the Europeana Cloud, b) The development of an effective research content strategy for Europeana, based on an evidence-based account of usefulness of Europeana and The European Library resources for research in the Humanities and Social Sciences, c) The improvement of the understanding of digital tools, research processes and content used in the Humanities and Social Sciences, thus informing the development of tools and aggregation of content in the Europeana Cloud, d) The active engagement of the Humanities and Social Sciences research communities in establishing user requirements for the development of Europeana Cloud.



DIACHRON - Managing the Evolution and Preservation of the Data Web

Project manager	Yannis Stavrakas
Coordinator	INTRASOFT
IMIS - funding	573,880 euros
Programme	FP7-ICT-2011-9
Start date	1/4/2013
Duration	3 years
Website	http://www.diachron-fp7.eu/



The Web has not only caused a revolution in communication; it also has completely changed the way we gather and use data. Open data -- data that is available to everyone -- is exponentially growing, and it has completely transformed the way we now conduct any kind of research or scholarship; it has changed the scientific method. The recent development of Linked Open Data has only increased the possibilities for exploiting public data. Given the value of open data how do we preserve it for future use? Currently, much of the data we use, e.g. demographic records, clinical statistics, personal and enterprise data as well as many scientific measurements cannot be reproduced. However, there is overwhelming evidence that we should keep such data where it is technically and economically feasible to do so. Until now this problem has been approached by keeping this information in fixed data sets and using extensions to the standard methods of disseminating and archiving traditional (paper) artifacts. Given the complexity, the interlinking and the dynamic nature of current data, especially Linked Open Data, radically new methods are needed. DIACHRON tackles this problem with a fundamental assumption: that the processes of publishing and preservation data are one and the same. Data are archived at the point of creation and archiving and dissemination are synonymous. DIACHRON takes on the challenges of evolution, archiving, provenance, annotation, citation, and data quality in the context of Linked Open Data and modern database systems. DIACHRON intends to automate the collection of metadata, provenance and all forms of contextual information so that data are accessible and usable at the point of creation and remain so indefinitely. The results of DIACHRON are evaluated in three large-scale use cases: open governmental data life-cycles, large enterprise data intranets and scientific data ecosystems in the life-sciences.



GLOBAQUA - Managing the effects of multiple stressors on aquatic ecosystems under water scarcity

Project manager	Prof. Phoebe Koundouri
Coordinator	CSIC, Spain
IMIS - funding	368,360 euros
Programme	EU's 7 th Programme for research, technological development and demonstration
Start date	2/2014
Duration	5 years
Website	www.globaqua-project.eu

The main aim of GLOBAQUA is to achieve a better understanding of how current water management practices and policies could be improved by identifying their main drawbacks and alternatives. Freshwater systems are under threat by a variety of stressors (organic and inorganic pollution, cover change, water abstraction, land use, etc.). The joint occurrence of many stressors (chemical, geomorphologic, biological) under water scarcity may produce novel and unfamiliar synergies and effects of unknown consequences. Therefore, it is crucial to understand deeply how water scarcity interacts with other stressors in freshwaters and to convey this information to managers, stakeholders and policymakers in order to minimize impacts, to adapt to oncoming changes, and to improve our management and policies. GLOBAQUA (Managing the effects of multiple stressors on aquatic ecosystems under water scarcity) has assembled a multidisciplinary consortium in order to study the interaction of multiple stressors within the frame of strong pressure on water resources. GLOBAQUA assesses the effects of water scarcity on aquatic ecosystems by focusing on six river basins (Adige, Anglian, Ebro, Evrotas, Sava and Souss Massa). These basins encompass a rich set of socio-ecological conditions and a wide geographic coverage, and focus on a specific set of stressors to illustrate different management scenarios.



DAIAD - Open Water Management

Project manager	Spiros Athanasiou
Coordinator	IMIS
IMIS - funding	1,06 MEuros
Programme	STREP, FP7-ICT-2013-11
Start date	1/3/2014
Duration	3.5 years
Website	http://www.daiad.eu

DAIAD is an EU-funded research project developing technologies for real-time monitoring, analysis, and understanding of water consumption data, aiming to induce sustainable changes in consumer behavior. Efficient water management is a challenging issue with the potential to affect the long-term well-being, economy and security of society. Policies for sustainable water management have been established in the EU. However, measures to support efficient water use for citizens are currently lacking. Consumers have limited means to accurately monitor their water consumption and thus stimuli to modify their behavior

towards a sustainable lifestyle. A potentially groundbreaking approach for efficient water use and reuse lies within the empowerment of consumers. The principles of open knowledge and participation have provided solutions and driven innovation in similar challenging and complex issues. We believe that a similar bottom up method, in which citizens can voluntarily adopt low cost water monitoring services, self-induce behavioral changes, and accordingly demand better services, can be a catalyst for large-scale changes in efficient water management. The DAIAD project constitutes an innovative approach for addressing the challenge of efficient water management through real-time knowledge of residential water consumption, bringing together leading members of the water and ICT domains. Our goal is to research and develop innovative low cost, inclusive technologies for real-time, high granularity water monitoring and knowledge extraction. We will devise multi-modal feedback interfaces, recommendation, and analysis services to communicate knowledge and incur behavioral changes to consumers in residential settings. We will apply big data management and analysis technologies to provide efficient management and analysis of real-time water consumption data, as well as multiple relevant data sources. This will enable water stakeholders to gain novel insight and explore the hidden correlations of the parameters that shape water demand strategies and water pricing, thus leading to more efficient water management.



DSI-1 - Access to digital resources of European Heritage ('Europeana')

Project manager	Dimitris Gavrilis
Coordinator	STICHTING EUROPEANA(EF)
IMIS - funding	180,260 euros
Programme	CEF-TC-2014-2-01
Start date	1/4/2015
Duration	14 months
Website	http://pro.europeana.eu/get-involved/projects/project-list/europeana-dsi

The Europeana Digital Service Infrastructure (DSI) comprises 'core service platforms' which enable trans-European connectivity and interoperability, and related 'generic services' which link national and sectorial infrastructures to the platforms. With funding from CEF, Europeana will develop into a widely recognised platform of services and resources, not only for metadata references,

but also for access to cultural content, tools and technologies, projects and other services. The ATHENA R.C. participates in DSI both as a technical partner with the MORE platform (supporting the CARARE network as part of DSI) and also through Europeana Research.



City.Risks - Avoiding and mitigating safety risks in urban environments

Project manager	Dimitrios Skoutas
Coordinator	SPACE HELLAS
IMIS - funding	452,125.00 euros
Programme	H2020-FCT-2014 - Research and Innovation action
Start date	1/5/2015
Duration	3 years
Website	http://project.cityrisks.eu/

The main objective of the City.Risks project is to increase the perception of security of citizens in big urban environments. This is achieved by activating in a more transparent and sustainable way their participation in communities, through which information and interventions can be provided both proactively, to protect citizens from falling victims to criminal activities, as well as reactively, to provide more timely and effective response and assistance. City.Risks leverages a set of innovative technologies, city infrastructures and available data sources. Moreover, it makes the citizens' smart phones the modern tool for increasing their personal and collective sense of security. The project designs and develops an innovative ecosystem of mobile services that transform the smart phone or the tablet of the citizen into a tool that collects, visualizes and shares safety-critical information with the appropriate authorities and communities. The project relies on a wide spectrum of available technologies to design and implement an interactive framework among authorities and citizens through mobile applications that allows in a collaborative way to prevent or mitigate the impact of crime incidents or other security threats. Thus, it contributes to an increase of the citizens' perception of security, which is measured and validated in real-life scenarios and conditions through the deployment and operation of pilot trials at several selected cities by the project partners.



ELIXIR-EXCELERATE - Fast-track ELIXIR implementation and drive early user exploitation across the life-sciences

Project manager	Theodore Dalamagas (Deputy: Stelios Sartzetakis)
Coordinator	European Molecular Biology Laboratory - EMBL
IMIS - funding	40,000 euros
Programme	EU, INFRADEV-3-2015
Start date	1/9/2015
Duration	4 years
Website	https://www.elixir-europe.org/excelerate

The precipitous drop in costs for high-throughput biology has enabled European research laboratories to produce a huge amount of complex and heterogeneous data. However, data will only generate long-term value if it is Findable, Accessible, Interoperable and Re-usable (FAIR). This requires a scalable infrastructure that connects local, national and European efforts and provides standards, tools and training for data stewardship. Motivated by this, ELIXIR, a distributed organization comprising national bioinformatics research infrastructures and the European Bioinformatics Institute, has developed ELIXIR-EXCELERATE. The project will fast-track ELIXIR's early implementation phase by i) coordinate and enhance existing resources into a world-leading data service for academia and industry, ii) grow bioinformatics capacity and competence across Europe, and iii) complete the management processes needed for a large distributed infrastructure. ELIXIR-EXCELERATE will enable cost-effective and sustainable management and re-use of data for millions of users across the globe and improve the competitiveness of European life science industries through accessible data and robust standards and tools.




RDA3 - RDA Europe – the European plug-in to the global Research Data Alliance

Project manager	Yannis Ioannidis
Coordinator	MAX PLANCK GESELLSCHAFT ZUR FOERDERUNG DER

WISSENSCHAFTEN E.V.(MPG)

IMIS - funding	232,292.50 euros
Programme	H2020-EINFRA-2014-2
Start date	1/9/2015
Duration	2.5 years
Website	https://www.rd-alliance.org/rda-europe-0

The Research Data Alliance (RDA) Europe project aims at facilitating European representation within the RDA Global alliance. IMIS participates in the main RDA Global governance body, namely the RDA secretariat with responsibilities in the RDA Organizational Assembly and Advisory Board. IMIS also participates in the European Policy Level engagement with emphasis on both research and industry engagement, as well as the outputs uptake work package which has a number of responsibilities: a) mapping the landscape of RDA outputs that are produced by the various working and interest groups, b) supporting collaboration projects, c) development and support of the “Atlas of Knowledge” knowledge base d) organization of datathon/hackathon events to promote the adoption and reuse of RDA outputs. Finally, IMIS is also involved in the dissemination work, focusing mainly on publications tracking, including tracking the RDA Working Group ones using the OpenAIRE Research Impact Service.



SLIDEWIKI - Large-scale pilots for collaborative OpenCourseWare authoring, multiplatform delivery and learning analytics

Project manager	Theodore Dalamagas
Coordinator	FRAUNHOFER IAIS
IMIS - funding	570,000 euros
Programme	EU, ICT-20-2015
Start date	14/11/2015
Duration	3 years
Website	https://slidewiki.eu

A major obstacle to increase the efficiency, effectiveness and quality of education in Europe is the lack of widely available, accessible, multilingual, timely, engaging and high-quality educational material (i.e. OpenCourseWare). The creation of comprehensive OpenCourseWare (OCW) is tedious, time-consuming and



expensive, with the effect that often courseware employed by teachers, instructors and professors is incomplete, outdated, inaccessible to those with disabilities and dull. Similarly to Wikipedia for encyclopaedic content, SlideWiki allows (1) to collaboratively create comprehensive, multilingual OCW (curricula, slide presentations, self-assessment tests, illustrations etc.) online in a crowdsourcing manner, (2) to support skills recognition and validation through learning analytics based on fine-grained content structuring and comprehensive learning self-assessment, and (3) set up an OCW community, where all relevant stakeholders from material provisioning, content production, translation, learning delivery and analytics are integrated into sustainable educational content value chains. SLIDEWIKI will perform four large-scale trials in secondary education, vocational and professional training, higher education and community-driven open-education. Each of these large-scale trials will be performed with hundreds of educators and thousands of learners in countries all over Europe.



ARCADES - Algebraic Representations in Computer-Aided Design for complEx Shapes

Project manager	Ioannis Z. Emiris
Coordinator	IMIS
IMIS - funding	485,000 euros
Programme	Marie Skłodowska-Curie Innovative Training Networks
Start date	1/1/2016
Duration	4 years
Website	http://arcades-network.eu

ARCADES aims at disrupting the traditional paradigm in Computer-Aided Design (CAD) by exploiting cutting-edge research in mathematics and algorithm design. Today, several approaches of the CAD industry are outdated, and 3D geometry processing is becoming increasingly the weak link. This is alarming in sectors where CAD faces new challenges arising from fast point acquisition, big data, and mobile computing, 3D printing, but also in robotics, simulation, animation, fabrication and manufacturing where CAD strives to address crucial societal and market needs. The challenge taken up by ARCADES is to build the next generation of CAD software based on strong foundations from algebraic geometry, differential geometry, scientific computing, and algorithm design. Our game-changing methods lead to real-time modelers for architectural geometry and visualisation, to isogeometric and design-through-analysis software for shape



optimisation, and marine design & hydrodynamics, and to tools for motion design, robot kinematics, path planning, and control of machining tools



DSI-2 - Access to digital resources of European heritage Europeana

Project manager	Dimitris Gavrilis
Coordinator	STICHTING EUROPEANA(EF)
IMIS - funding	153,010 euros
Programme	CEF-TC-2015-1-01
Start date	1/7/2016
Duration	14 months
Website	http://pro.europeana.eu/get-involved/projects/project-list/europeana-dsi-2

The Europeana DSI-2 project builds on and continues operation of the existing Europeana Digital Service Infrastructure's (DSI) Core Service Platform (CSP) following on from the Europeana DSI-1 action under CEF (CEF-TC-2014-2). It is the second project for the completion of Europeana's strategy 2015-2020. It is in line with the Connecting Europe Facility (CEF) Trans-European Telecommunications Network's work programme for 2015, delivering interoperability, connectivity and coordination for digital cultural heritage at a European level and providing efficient solutions for access and distribution of multilingual and multi-domain resources in digital European cultural heritage. IMIS participates in DSI-2 both as a technical partner with the MORE platform (supporting the CARARE network as part of DSI) and also through Europeana Research.

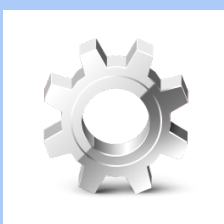


ARKWORK - Archaeological practices and knowledge work in the digital environment

Project manager	Costis Dallas
------------------------	---------------

Coordinator	Uppsala University (SE)
IMIS - funding	About 50,000 euros (The funding of COST actions is adjusted yearly)
Programme	COST European Cooperation in Science & Technology Programme, Action CA15201
Start date	6/10/2016
Duration	3 years
Website	https://www.arkwork.eu/

The COST action ARKWORK is aimed at forming an interdisciplinary network that will collect the knowledge and experiences of researchers and projects, national as well as European, in producing and exploiting archaeological knowledge. The network thus intends to contribute to the coordination of current efforts to study archaeological practices, and the potential for social and economic exploitation of archaeological knowledge in the digital environment. The network seeks to establish the state of the art in promoting archaeological knowledge and its application for the benefit of society, and to provide expert guidance in the production, validation, reproduction, management and use of archaeological knowledge. Apart from networking and information activities, the project collects and charts basic information and approaches concerning archaeological practices in field research, collections, and the production and validation of knowledge, through a combination of approaches from archaeological theory and methodology, information science, computer science, and social studies of science and technology.



DARIAH Reference Curriculum / DariahTeach

Project manager	Panos Constantopoulos
Coordinator	Maynooth University, Ireland
IMIS - funding	34,591 euros
Programme	Erasmus+
Start date	1/1/2015
Duration	30 months
Website	http://dariah.eu/teach/

In the DariahTeach project a consortium of 7 universities and research centres is developing online course material in the broader area of Digital Humanities, to be delivered openly and asynchronously. IMIS contributes to the design and evaluation of educational material, as well as by developing the unit *Introduction to conceptual modeling*.

National R&D Projects



ELIXIR-GR - Information Infrastructure for the Life Sciences

Project manager	Theodore Dalamagas (Deputy: Stelios Sartzetakis)
Coordinator	Biomedical Science Research Center Alexander Fleming
IMIS - funding	95,000 euros
Programme	National Roadmap for Research Infrastructures
Start date	1/2014
Duration	7 years
Website	http://www.elixir-greece.org

ELIXIR (<http://www.elixir-europe.org/>) is one of the ESFRI's pan-European research infrastructures (RIs) with key importance for the research and innovation in the life sciences. The aim of ELIXIR is to provide the facilities necessary for Europe's life scientists to manage and safeguard the massive amounts of data being generated every day by publicly funded research. Greece has recently joined ELIXIR, which has already entered its construction phase. ELIXIR operates a distributed research infrastructure based on ELIXIR nodes (at national level) which are connected to the ELIXIR Hub in order to provide infrastructure for data, tools, standards and training, as well as support for other ESFRI biological and medical science infrastructures. BSRC A. Fleming is joining forces with ATHENA RC and GRNET to lead the construction of the Greek component, ELIXIR-EL. Setting up ELIXIR- EL will be a dynamic process, taking into account user needs, available tools and datasets and technological advancements. The RI will be designed and setup using GRNET's Okeanos cloud IaaS service (<http://okeanos.grnet.gr/>) and ViMa service (<http://vima.grnet.gr>). The proposed RI aims to develop ELIXIR-EL RIs at the following levels: computing resources, data-intensive programming components, data resources, and tools.





CitySense - Retrieving, Visualizing and Combining Datasets on Urban Areas

Project manager	Yannis Stavrakas
Coordinator	IMIS
IMIS - funding	95,000 euros
Programme	Data value chain in industrial and urban environments - SIEMENS
Start date	1/4/2015
Duration	2 years
Website	http://web.imis.athena-innovation.gr/projects/CitySense/

Social networks, available open data and massive online APIs provide huge amounts of data about our surrounding location, especially for cities and urban areas. Unfortunately, most applications and research usually focus on one kind of data over the other, thus presenting a biased and partial view of each location in question, hence limiting the benefits of such approaches. To remedy this, the CitySense framework combines data from administrative sources (e.g., public agencies), massive Point of Interest APIs (Google Places, Foursquare) and social microblogs (Twitter) to provide a unified view of all available information about an urban area, in an intuitive and easy to use web-application platform. Our use case shows how different sources of information can be combined to provide an accurate and diverse visualization for the urban area of Chicago, USA.



eLib GGDE - eLib of Independent Authority for Public Revenue

Project manager	George Papastefanatos
Coordinator	Remaco S.A.
IMIS - funding	49,000 euros
Programme	NSRF - Pubic Tender
Start date	30/7/2015

Duration 2 years and 3 months

Website <http://www.publicrevenue.gr/elib/>

The project eLib aims at developing a digital library for the Independent Authority of Public Revenue (I. A.P.R). eLib analyzes and provides information to public servants and citizens on the legislation concerning the areas of IAPR competence (taxation, public revenues, etc). Legal documents are automatically harvested from institutional sources (National Printing Office, Di@ygeia), their content and metadata are extracted, semantically analyzed and structured according to the AKOMA NTOSO, and ELI specifications. Moreover, content is indexed for full-text search, browsing and faceted filtering. Users can search and navigate the legal content and create personal collections with references to whole documents or parts of them.

Other Projects



TwitHoard - retrieving, modeling and analyzing Social Data

Project manager Yannis Stavrakas

Coordinator IMIS

Programme Internal project

Start date 12/2015

Duration 2 years and 10 months

An increasing number of innovative applications use data from online social networks. In many cases data analysis tasks, like opinion mining processes, are applied on platforms such as Twitter, in order to discover what people think about various issues. In our view, selecting the proper data set is paramount for the analysis tasks to produce credible results. This direction, however, has not yet received a lot of attention. TwitHoard is a platform for supporting processes such as opinion mining on Twitter data, with emphasis on the selection of the proper data set. The key point of our approach is the representation of term associations, user associations, and related attributes in a single model that also takes into account their evolution through time. This model enables flexible queries that combine complex conditions on time, terms, users, and their associations.



ARK4

Project manager	Costis Dallas
Coordinator	NTNU University Library
IMIS - funding	6,426.75 euros
Programme	Norwegian University of Science and Technology (NTNU)
Start date	1/2/2016
Duration	9 months
Website	http://mubil.no/ark-4/

ARK4 was a cooperation project between four Trondheim institutions: NTNU University Library; NTNU Computer and Information Technology (IDI); Regional State Archive, Trondheim and Norsk Døve Museum, while IMIS was the international partner of the project. ARK4 was seeking to create a virtual dialogue between the public and institutions that hold information. By exploring innovative technology, it aimed to disseminate knowledge from the past to a wider, younger audience. Within ARK4, IMIS applied user-centred methods and collected new narratives to be shared by the national and regional community of libraries, schools and archives through surveys and interviews. Moreover, in the context of ARK4, IMIS re-used Europeana material to build a series of online quizzes, such as Children's Literature, a joyful topic that aimed to travel its users back to their childhood and test what they have been reading and remember from then. Through the questions, the users could have a look at beautiful visual records lying currently within Europeana.



Digital Curation Services

Project manager	Panos Constantopoulos
Coordinator	IMIS
IMIS - funding	48,786 euros

Programme	Internal project
Start date	1/2/2016
Duration	2 years
Website	-

The internal project Digital Curation Services supports the advancement of research and development efforts in the research directions and work programme of IMIS by using chiefly funds remaining from the execution of service provision projects. In view of parallel externally funded projects, this project on one hand ensures the continuity of our work programme, while on the other it enables actions of topical interest for which external funds may be insufficient at the time. It also allows better integrating and employing results of completed projects in sustained and new digital curation services, thus supporting the capitalization of those results.



AMNESIA

Project manager	Manolis Terrovitis
Coordinator	IMIS
Programme	Internal project
Start date	1/1/2014
Duration	3 years
Website	amnesia.imis.athena-innovation.gr

Amnesia allows users to easily anonymize data (tabular and set-valued data) and focuses on techniques that allow the users to guide the anonymization process and reduce the information loss. It employs visualization tools and supportive mechanisms that are able to handle data with missing values, to re-use existing anonymization scenarios etc. The tool has been used in several projects, including OpenAIRE (www.openaire.eu), MEDA (<http://meda.imis.athena-innovation.gr>), and MHMD (www.myhealthmydata.eu).

Publications

Book chapters

- George Papastefanatos and S. Maroulis: **Socioscope - A visual analytics platform for open socioeconomic data**. Book chapter in “Transformations of Protest in Greece”, Stathopoulou, T (ed.), Papazisis publishers & EKKE. (in print).
- Lorna Hughes, Panos Constantopoulos, Costis Dallas: **Digital Methods in the Humanities: Understanding and Describing their Use across the Disciplines**. Book chapter in Susan Schreibman, Ray Siemens, John Unsworth (Eds.): A New Companion to Digital Humanities, Chichester, United Kingdom: John Wiley & Sons, Ltd 150-170 (2016).

Journal Publications

- Ioannis S. Vlachos, Thanasis Vergoulis, Maria D. Paraskevopoulou, Filipoimin Lykokanellos, Georgios Georgakilas, Penny Georgiou, Serafeim Chatzopoulos, Dimitra Karagkouni, Foteini Christodoulou, Theodore Dalamagas, Artemis G. Hatzigeorgiou: **DIANA-mirExTra v2.0: Uncovering microRNAs and transcription factors with crucial roles in NGS expression data**. Nucleic Acids Research 44(Webserver-Issue): W128-W134 (2016)
- Georgios Georgakilas, Ioannis S. Vlachos, Konstantinos Zagganas, Thanasis Vergoulis, Maria D. Paraskevopoulou, Ilias Kanellos, Panayiotis Tsanakas, Dimitris Dellis, Athanasios Fevgas, Theodore Dalamagas, Artemis G. Hatzigeorgiou: **DIANA-miRGen v3.0: accurate characterization of microRNA promoters and their regulators**. Nucleic Acids Research 44(Database-Issue): 190-195 (2016)
- Maria D. Paraskevopoulou, Ioannis S. Vlachos, Dimitra Karagkouni, Georgios Georgakilas, Ilias Kanellos, Thanasis Vergoulis, Konstantinos Zagganas, Panayiotis Tsanakas, Evangelos Floros, Theodore Dalamagas, Artemis G. Hatzigeorgiou: **DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts**. Nucleic Acids Research 44(Database-Issue): 231-238 (2016)
- George Papadakis, George Papastefanatos, Themis Palpanas, Manolis Koubarakis: **Boosting the Efficiency of Large-Scale Entity Resolution with Enhanced Meta-Blocking**. Big Data Research 6: 43-63 (2016)
- Marios Meimaris, George Papastefanatos, Stratis Viglas, Yannis Stavrakas, Christos Pateritsas, Ioannis Anagnostopoulos: **A query language for multi-version data web archives**. Expert Systems 33(4): 383-404 (2016)
- Panagiotis Bouros, Nikos Mamoulis, Shen Ge, Manolis Terrovitis: **Set containment join revisited**. Knowl. Inf. Syst. 49(1): 375-402 (2016)



- Costis Dallas: **Digital curation beyond the “wild frontier”: a pragmatic approach.** Archival Science 16(4): 421–457 (2016)
- Costis Dallas: **Jean-Claude Gardin on archaeological data, representation and knowledge: implications for digital archaeology.** Journal of Archaeological Method and Theory 23(1): 305–330 (2016)
- Vayianos Pertsas, Panos Constantopoulos: **Scholarly Ontology: modelling scholarly practices.** Int J Digital Libraries 1-18 (2016, Online first)
- P. Koundouri, P. Ker Rault, V. Pergamalis, V. Skianis, I. Souliotis: **Development of an integrated methodology for the sustainable environmental and socio-economic management of river ecosystems.** Science of The Total Environment, Volume 540, 1 January 2016, Pages 90-100
- P. Manousis, P. Vassiliadis, A. Zarras and G. Papastefanatos: **Schema Evolution for Databases and Data Warehouses**, Lecture Notes in Business Information Processing (253), Springer 2016, ISBN 978-3-319-39242-4.
- Panagiotis Bouros, Nikos Mamoulis, Shen Ge, Manolis Terrovitis: **Set containment join revisited.** Knowl. Inf. Syst. 49(1): 375-402 (2016).

International Conference / Workshop Publications

- Pantelis Chronis, Giorgos Giannopoulos, Spiros Athanasiou: **Open Issues and Challenges on Time Series Forecasting for Water Consumption.** EDBT/ICDT Workshops 2016
- Giorgos Giannopoulos, Nikos Karagiannakis, Dimitrios Skoutas, Spiros Athanasiou: **Learning to Classify Spatiotextual Entities in Maps.** ESWC 2016: 539-555
- Olga Gkountouna, Manolis Terrovitis: **Anonymizing collections of tree-structured data.** ICDE 2016: 1520-1521
- Marios Meimaris, George Papastefanatos: **The EvoGen Benchmark Suite for Evolving RDF Data.** MEPDaW/LDQ@ESWC 2016: 20-35
- Konstantina Papanikolaou, Haris Papageorgiou, Nikos Papasarantopoulos, Theoni Stathopoulou, George Papastefanatos: **"Just the Facts" with PALOMAR: Detecting Protest Events in Media Outlets and Twitter.** SMN@ICWSM 2016
- Giorgos Alexiou, Marios Meimaris, George Papastefanatos: **Enabling persistent identification of groups of duplicates in data aggregators.** ICDE Workshops 2016: 124-126
- Nikos Bikakis, John Liagouris, Maria Krommyda, George Papastefanatos, Timos K. Sellis: **graphVizdb: A scalable platform for interactive large graph visualization.** ICDE 2016: 1342-1345
- Marios Meimaris, George Papastefanatos: **Double Chain-Star: an RDF indexing scheme for fast processing of SPARQL joins.** EDBT 2016: 668-669



- Marios Meimaris, George Papastefanatos, Panos Vassiliadis, Ioannis Anagnostopoulos: **Efficient computation of containment and complementarity in RDF data cubes**. EDBT 2016: 281-292
- George Papadakis, George Papastefanatos, Themis Palpanas, Manolis Koubarakis: **Scaling entity resolution to large, heterogeneous data with enhanced meta-blocking**. EDBT 2016: 221-232
- Dimitrios Skoutas, Dimitris Sacharidis, Kostas Stamatoukos: **Identifying and describing streets of interest**. EDBT 2016: 437-448
- Christodoulos Efstathiades, Alexandros Belesiotis, Dimitrios Skoutas, Dieter Pfoser: **Similarity search on spatio-textual point sets**. EDBT 2016: 329-340
- Paras Mehta, Dimitrios Skoutas, Dimitris Sacharidis, Agnès Voisard: **Coverage and diversity aware top-k query for spatio-temporal posts**. SIGSPATIAL/GIS 2016: 37:1-37:10
- Paras Mehta, Dimitris Sacharidis, Dimitrios Skoutas, Agnès Voisard: **Keyword-based retrieval of frequent location sets in geotagged photo trails**. WebSci 2016: 348-349
- Theodora Galani, George Papastefanatos, Yannis Stavarakas: **A language for defining and detecting interrelated complex changes on RDF(S) knowledge bases**. ICEIS (1) 2016: 472-481
- Danae Pla Karidi, Yannis Stavarakas, Yannis Vassiliou: **A personalized tweet recommendation approach based on concept graphs**. UIC/ATC/ScalCom/CBDCCom/IoP/SmartWorld 2016: 253-260
- Olga Gkountouna, Manolis Terrovitis: **Anonymizing collections of tree-structured data**. ICDE 2016: 1520-1521
- Dimitris Gavrilis, Vangelis Nomikos, Konstantinos Kravvaritis, Stavros Angelis, Christos Papatheodorou, Panos Constantopoulos: **MORE: A micro-service oriented aggregator, metadata and semantic research**. In Proceedings 10th Int Conf Communications in Computer and Information Science -CCIS 2016 15-26 Göttingen: Springer-Verlag.
- Alexandra Angeletaki, Agiatis Benardou, Nephelie Chatzidiakou, Eliza Papaki: **Playing with cultural heritage through digital gaming: The new narrative of the ARK4 project**. Digital Humanities 2016: 730
- Agiatis Benardou, Valentine Charles, Nephelie Chatzidiakou, Panos Constantopoulos, Costis Dallas, Ana Isabel González Sáez, Sergiu Gordea, Lorna M. Hughes, Themistoklis Karavellas, Gregory Marcus, Leonidas Papachristopoulos, Vayianos Pertsas: **Scholarly research activities and digital tools: when NeMO met FLOSS**. Digital Humanities 2016: 740-741
- Agiatis Benardou, Alastair Dunning, Stefan Ekman, Vicky Garnett, Caspar Jordan, Ilze Lace, Eliza Papaki: **Reflecting on and refracting user needs through case studies in the light of Europeana research**. Digital Humanities 2016: 742-743
- Agiatis Benardou, Vicky Garnett, Eliza Papaki: **"Do you remember the first time?": Case studies on digital content reuse in the context of Europeana Cloud**. 2016

- Panos Constantopoulos, Lorna M. Hughes, Costis Dallas, Vayianos Pertsas, Leonidas Papachristopoulos, Timoleon Christodoulou: **Contextualized integration of digital humanities research: Using the NeMO ontology of digital humanities methods**. Digital Humanities 2016: 161-163
- Dimitris Gavrilis, Eleni Afiontzi, Johan Fihn, Olof Olsson, Sebastian Cuy, Achille Felicetti, Franco Niccolucci: **A data integration infrastructure for archaeology**. CAA 2016
- Franca Debole, Nicola Aloia, Christos Papatheodorou, Dimitris Gavrilis, Carlo Meghini: **A catalog for archaeological resources**. CAA 2016
- Dimitris Gavrilis, Johan Fihn, Sebastian Cuy, Achille Felicetti: **Integrating data for archaeology**. EAA 22nd Annual Meeting 2016: 339-340
- Eliza Papaki, James O'Sullivan, Antonio Rojas: **Curating Community: Building a communications strategy for the European Association for Digital Humanities**. Digital Humanities 2016: 279
- Susan Schreibman, Agiatis Benardou, Claire Clivaz, Matej Durco, Marianne Huang, Eliza Papaki, Stef Scagliola, Toma Tasovac, Tanja Wissik: **#dariahTeach: online teaching, MOOCs and beyond**. Digital Humanities 2016 354-356
- Vinayak Mathur, Yannis Stavrakas, Sanjay Singh: **Intelligence analysis of Tay Twitter Bot**. IEEE 2nd International Conference on Contemporary Computing and Informatics (IC3I 2016), Noida-India, December 2016.
- Giorgos Alexiou, Sahar Vahdati, Christoph Lange, George Papastefanatos, Steffen Lohmann: **OpenAIRE LOD services: Scholarly communication data as linked data**. In 2nd International Workshop on Semantics, Analytics, Visualisation: Enhancing Scholarly Data (SAVE-SD16) co-located with WWW16, Montreal, Canada, April, 2016.

National Conference / Workshop Publications

- Marios Koniaris, George Papastefanatos, Yannis Vassiliou: **Towards automatic structuring and semantic indexing of legal documents**. PCI 2016: 4
- Marios Meimaris, George Papastefanatos, Panos Vassiliadis, Ioannis Anagnostopoulos: **Efficient computation of containment and complementarity in linked data cubes**. In 14th Hellenic Data Management Symposium (HDMS'16), Athens, July, 2016.
- George Papadakis, George Alexiou, George Papastefanatos, Georgia Koutrika: **Schema-agnostic vs schema-based configurations for blocking methods on homogeneous data**. In 14th Hellenic Data Management Symposium (HDMS'16), Athens, July, 2016.
- Yannis Roussakis, Ioannis Chrysakis, Kostas Stefanidis, Giorgos Flouris and Yannis Stavrakas: **A flexible framework for understanding the dynamics of evolving RDF**



datasets (same as ISWC 2015). The 14th Hellenic Data Management Symposium (HDMS 2016), Athens-Greece, July 28-29, 2016.

- John Liagouris and Manolis Terrovitis. **Efficient Identification of Implicit Facts in Incomplete OWL2-EL Knowledge Base.** In 14th Hellenic Data Management Symposium (HDMS'16), Athens, July, 2016.

Other Publications / Technical Reports

- Costis Dallas, Nephelie Chatzidiakou, Beat Immenhauser, Ingrida Kelpšienė, Maciej Maryl, Schneider, G., Zebec, T. (2016). **European survey on scholarly practices and digital needs in the arts and humanities.** Highlights Report. Athens. DOI: 10.5281/zenodo.260101
- G. Lock, A. Benardou, C. Dallas, P. Reilly & J. Huggett. **Unstable futures/potential pasts: Scenarios for digital archaeology 2020.** Computer Applications and Quantitative Methods in Archaeology (CAA2016), Oslo, 29 March – 2 April 2016.
- C. Dallas, A. Degraeve, K. Fernie. **Towards digital infrastructures for open, participatory, public archaeology.** In 22nd Annual Meeting of the European Association of Archaeologists (EAA2016) (p. 401), Vilnius, 2016



Dissemination Activities

Invited / Keynote Talks

- 12/2016, "**Preparing for change**", George Papastefanatos, panel speaker at the International Conference and Meeting "Acting on Change: New Approaches and Future Practices in Long term Digital Preservation" organized by Digital Preservation Coalition and PERICLES FP7 project, London UK.
- 12/2016, "**Proactive data management – where are we now and where are we going?**", George Papastefanatos, keynote speaker at the International Conference and Meeting "Acting on Change: New Approaches and Future Practices in Long term Digital Preservation" organized by Digital Preservation Coalition and PERICLES FP7 project, London UK.
- 3/2016, "**Challenges and opportunities on data evolution and preservation**", George Papastefanatos, invited speaker at the DIACHRON project Dissemination Workshop on "Data evolution and preservation", 24th of March 2016 Blank! Urban space Athens.
- 9/2016, "**Making Use of the Linked Open Data Services for OpenAIRE: Querying Data about Research Results, Persons, Projects and Organizations**", S. Vahdati, G. Alexiou, G. Papastefanatos, C. Lange. Tutorial in the *Digital Infrastructures for Research* event, Krakow-Poland.
- 12/2016, "**Big Data analysis and visualization: research and application**", Yannis Stavrakas, invited speaker at the Big Data and Statistics Information Day, Hellenic Statistical Authority (ELSTAT), Athens, Greece, December 9, 2016.
- 04/2016, "**New generation of digital tools in education**", Theodore Dalamagas, invited speaker at the Athens Science Festival 2016.
- 04/2016, "**Europeana Research in the Next 5 Years: Challenges and Opportunities**", N. van Schaverbeke, A. Benardou (moderators), C. Dallas, L. Isaksen, T. Tasovac, M. Žic–Fuchs, Using digital heritage for research Policy challenges and opportunities, Europeana & National Library of the Netherlands, The Hague.
- 04/2016, "**Culture, Museums and Social Networking Media**", D. Catapoti (moderator), C. Dallas, Th. Kavoura, M. Sideri, M. Leni, N. Papadimitriou, Frontiers and Boundaries: Archaeological Dialogues 2016, University of the Aegean, Mytilini.

Scientific Community Service

IMIS members have served in the **Program Committee** of more than 17 International Conferences and Workshops in 2016, including well-known Conferences, including IEEE ICDE, ESWC, CIKM, TPD, ODBASE, ECIR, and more.



IMIS members have participated in the **organization** or co-organization of the following events:

- **7th International Workshop on Data Engineering meets the Semantic Web (DESWeb)**. In conjunction with ICDE 2016, Helsinki, Finland, May 16, 2016 – Theodore Dalamagas, Yannis Stavarakas.
- **Ontology-Based Recording and Discovery of Research Patterns in the Humanities**. Pre-conference workshop, Digital Humanities 2016. Kraków – P. Constantopoulos, C. Dallas.
- **2nd International Workshop on Preservation of Evolving Big Data (DIACHRON 2016)** jointly with EDBT 2016 Workshop (15th of March 2016, Bordeaux, France) – George Papastefanatos.
- **DIACHRON project Dissemination Workshop on “Data evolution and preservation”**, 24th of March 2016 Blank! Urban space Athens – Yannis Stavarakas, George Papastefanatos, Christos Pateritsas.

Other Dissemination Activities

- **RDA Wheat Datathon**. (13-14 July 2016). The “Innovating the wheat community through the RDA services and outputs” Datathon took place at the ATHENA Research Centre in Athens on the 13th and 14th July 2016. The event focused on getting researchers and other stakeholders acquainted with the data formats, services, protocols and needs of the wheat community. The event was organized by the ATHENA Research Centre in the context of the RDA Europe 3 project in cooperation with the RDA Wheat Data Interoperability WG.
<http://rda-wheat.imis.athena-innovation.gr/>
- **Researchers Night 2016**: IMIS participated in the event presenting the results of many research and development activities.
- **Athens Science Festival 2016**: IMIS participated in the event presenting the results of many research and development activities.
- **Befinnovative bussiness acceleration program**: IMIS member Spyros Athanasiou participated to the program as a mentor.
- **Open Water Days**: IMIS organized, coordinated, and participated in a new series of open participatory events celebrating innovation for water efficiency. The *OpenWaterDays* (Athens, Bamberg, Alicante, London, Madrid) allowed us to tap into stakeholders and citizens, provided hands-on experimentation with the DAIAD project technologies, informed about water challenges and ICT solutions, and engaged water utilities and decision makers into evaluating the novel water monitoring technologies we have developed on a production setting
<http://daiad.eu/?p=3526>



Systems

- **Mopseus**

MOPSEUS is a scalable, curation-aware repository system designed to be metadata schema agnostic. It can support any complex data model either at the digital resource level or at the collection level. This means that content can be organized using any structure ranging from simple hierarchies to complex graphs. The entire structure both at the digital resource level and at the collection level is stored and represented in RDF and can be accessed through a SPARQL endpoint. MOPSEUS employs an expressive data model that supports both intra- and inter- object relations thus allowing arbitrary organization of objects. In particular, all entities in MOPSEUS are digital objects each of which may include an unlimited number of metadata and/or data files and may be associated with multiple metadata schemas. A special class of digital objects (containers) are used in order to organize information (digital evidence). Containers can be interconnected using semantic links thus giving rise to semantic graph structures of arbitrary complexity. Metadata can be represented either as XML or RDF triples. MOPSEUS gives special focus on interoperability and digital preservation and is compliant with the PREMIS standard ensuring that the entire lifecycle of each digital resource is stored and semantically annotated. In order to access the MOPSEUS services a modular architecture is employed whereby the user interacts with the system through a set of Web-based interfaces that allow one to define metadata schemas and thesauri and to manage the entire content stored within the system. A SKOS editor supports maintaining term thesauri, while a linked data approach is adopted in associating terms with relevant data objects. Moreover, semantic relationships between objects are supported and can be defined through the GUI. Finally, all operations are organized in workflows that are also defined via the GUI. Supported site installations of the Mopseus repository system include: “Digital Academy” – Repository of the digital collections of the Academy of Athens, and “Pyxida” – Academic repository and digital library, Athens University of Economics and Business.

- **MORe**

The Metadata and Object Repository (MORe) is a metadata aggregator designed to: harvest content (metadata records) from different sources and providers; enrich/curate; map to a target schema (e.g. EDM); and deliver the metadata using the OAI-PMH protocol to other systems, such as the Europeana library. MORe focuses on enriching / curating the aggregated content. This is accomplished through a set of micro-services that are streamlined in a workflow. These micro-services perform various curation



actions like normalizing, associating records (e.g. those in close proximity to each other), transforming spatial coordinates to a given coordinate system (e.g. WGS84), creating elements like place labels, etc. MORE is OAI-compliant and preserves the whole lifecycle of each digital object. All ingestion and curation actions create new versions of metadata streams that are stored and semantically annotated, thus allowing to view the entire history of changes associated with each digital resource. MORE is tuned to support massive imports in the order of about 0.8 million records per hour. Online services with the MORE system are provided to Europeana and the related communities formed and supported through the projects CARARE, 3-D ICONS, ARIADNE, LoCloud and CEF Europeana. A total of approximately 70 content providers in over 20 countries, are being regularly served by the MORE aggregation service to aggregate, enrich and deliver content to Europeana, with approximately 10 million heritage asset records processed so far.

- **DAIAD system**

DAIAD is the first *integrated residential demand management* system for water. It applies Big Data and Machine Learning technologies to leverage smart water meter data, engaging and informing consumers to induce sustainable changes in consumption behavior, as well as providing novel large-scale analytics to improve short-, medium-, and long-term demand management for water utilities. DAIAD provides personalized pricing and non-pricing interventions to consumers through mobile and web applications, adapted to their profile, individual determinant sensitivity, and consumption behavior. Water utilities have access to several analysis services (segmentation, clustering, forecasting) enabling them to understand consumption behavior at the household level, target specific consumer groups, and anticipate demand under various time scales. The average sustainable total water savings in residential water consumption achieved by the DAIAD system is -12%. DAIAD is available as an open-source software system under the Apache License.

<https://github.com/DAIAD>

<https://www.youtube.com/watch?v=YuLU9nitlss>

- **BiP! Finder**

BiP! finder is a tool that assists the discovery of qualitative publications in the field of life sciences. This tool supports ranking and comparing of scientific articles based on their popularity or influence, while it provides useful features like intuitive infographics for each article and a mechanism of bookmarks. Article quality metrics. The quality of papers has several aspects. For example, researchers usually search for popular papers, i.e., papers in the cutting edge of their field. However, a young researcher may also be interested in influential articles, i.e., those with a strong research impact. BiP! finder

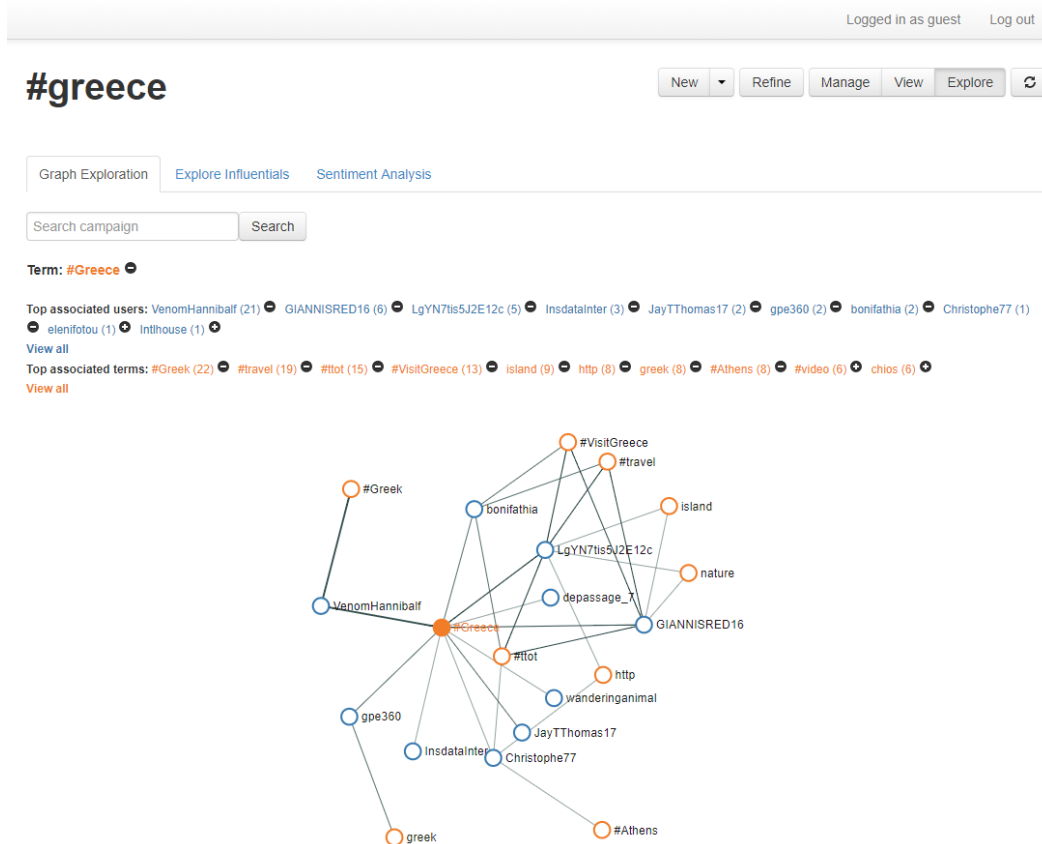


ranks and compares biomedical papers based on either their popularity or their influence. To do so, link analysis techniques are utilized.

<http://bip.imis.athena-innovation.gr/>

• **TwitHoard**

An increasing number of innovative applications use data from online social networks. In many cases data analysis tasks, like opinion mining processes, are applied on platforms such as Twitter, in order to discover what people think about various issues. In our view, selecting the proper data set is paramount for the analysis tasks to produce credible results. This direction, however, has not yet received a lot of attention. TwitHoard is a platform for supporting processes such as opinion mining on Twitter data, with emphasis on the selection of the proper data set. The key point of our approach is the representation of term associations, user associations, and related attributes in a single model that also takes into account their evolution through time. This model enables flexible queries that combine complex conditions on time, terms, users, and their associations.



<http://twithoard.imis.athena-innovation.gr:8080/twithoard>

• **eLib**

The project eLib aims at developing a digital library for the Independent Authority for Public Revenue. eLib analyzes and provides information to public servants and citizens

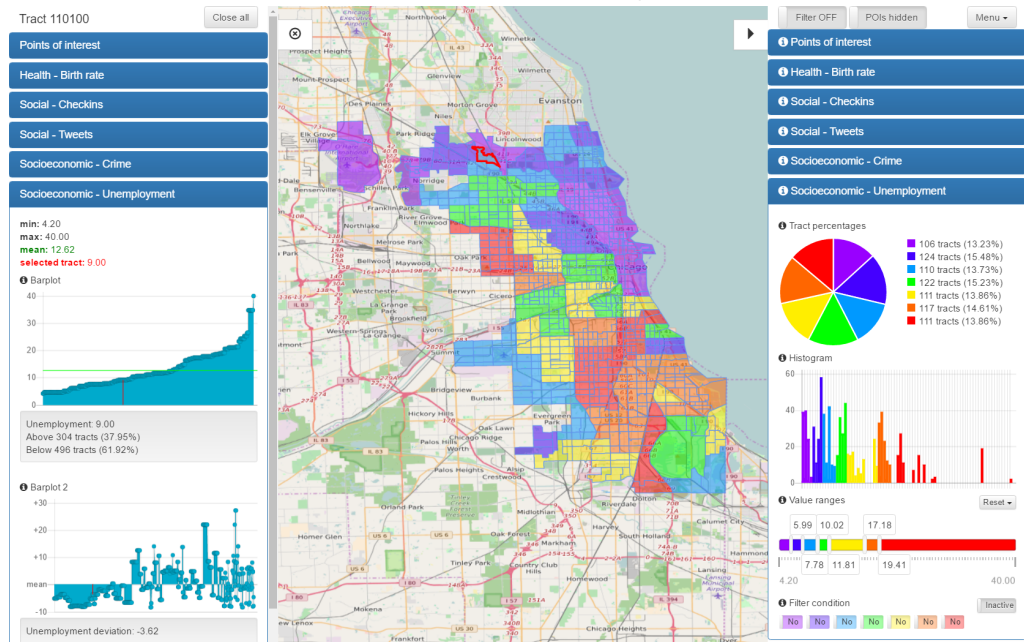


on the legislation concerning the areas of IAPR competence (taxation, public revenues, etc). Legal documents are automatically harvested from institutional sources (National Printing Office, Di@ygeia), their content and metadata are extracted, semantically analyzed and structured according to the AKOMA NTOSO, and ELI) specifications. Moreover, content is indexed for full-text search, browsing and faceted filtering. Users can search and navigate the legal content and create personal collections with references to whole documents or parts of them.

<http://www.publicrevenue.gr/elib>

- **GeoProfiler**

Social networks, available open data and massive online APIs provide huge amounts of data about our surrounding location, especially for cities and urban areas. Unfortunately, most applications and research usually focus on one kind of data over the other, thus presenting a biased and partial view of each location in question, hence limiting the benefits of such approaches. To remedy this, the GeoProfiler framework, developed within the CitySense project, combines data from administrative sources (e.g., public agencies), massive Point of Interest APIs (Google Places, Foursquare) and social microblogs (Twitter) to provide a unified view of all available information about an urban area, in an intuitive and easy to use web-application platform. Our use case shows how different sources of information can be combined to provide an accurate and diverse visualization for the urban area of Chicago, USA.



City of Chicago Demo: <http://geoprofiler.imis.athena-innovation.gr/>

- **linkzoo**

Linkzoo is a web-based, linked data enabled tool that supports collaborative management of information resources. www.linkzoo.gr enables users to create and

manage diverse types of web resources into common data spaces such as files, web documents, people, datasets and calendar events.

www.linkzoo.gr

- **Socioscope**

A visual analysis tool, used for visualization and exploration of social and political data (www.socioscope.gr) seeks to deliver a visual analytics platform for the social scientist to explore and analyze social facts through a user-friendly visual interface. The Socioscope platform offers a variety of interactive visualizations for each different type of data: charts and histograms, pies and tacked diagrams for numerical data; timelines for indices; and choropleth and point maps for geographical data. The platform is based on a multidimensional modeling approach and offers several visual operations for data exploration and analysis, such as filtering through faceted browsing, hierarchical representation of coded lists in charts, free keyword search of literal values, and capabilities for combining different datasets along common dimensions. Moreover, it makes knowledge reusable by making all data available for download in various formats including Linked Open Data.

www.socioscope.gr

- **TripleGeo**

TripleGeo is an ETL utility that can extract geospatial features from various sources (e.g. shapefiles, spatial DBMSs) and transform them into Basic Geo or GeoSPARQL compatible RDF triples, in several serialization formats. It copes with most common spatial data types, like points, linestrings and multi-linestrings, polygons and multi-polygons and supports on-the-fly transformations between different coordinate reference systems. Also, TripleGeo supports the transformation of INSPIRE-aligned spatial data and metadata into RDF, using XSL stylesheets, for selected INSPIRE data themes.

<https://github.com/SLIPO-EU/TripleGeo>

- **OSMRec**

OSMRec is a tool that trains on a set of spatial entities annotated with categories and provides category recommendations for new geospatial entities. OSMRec's goal is to exploit the richness of available geospatial datasets than contain entities already annotated with several categories (e.g., OpenStreetMap), to enrich new geospatial entities. OSMRec supports two modes of deployment: a generic command line, and a JOSM plugin, which allows the real-time recommendation of OSM categories for geospatial entities created within the JOSM user interface.

<https://github.com/SLIPO-EU/OSMRec>

<http://wiki.openstreetmap.org/wiki/JOSM/Plugins/OSMRec>



- **Amnesia anonymization tool**

Amnesia transforms a dataset with direct identifiers and quasi identifiers to an anonymized dataset, where formal privacy guaranties hold. Amnesia allows the use to customize the anonymization process, to choose the trade-off between data utility and privacy protection. Moreover, it allows users who are not IT experts to visually explore the data and the impact of different anonymization settings on them. It helps use to create supportive material to the anonymization process, like generalization hierarchies. Amnesia offers k-anonymity and km-anonymity and a parallel scalable anonymization algorithm.

- **FAGI**

FAGI is a tool that allows the fusion of geospatial Linked Data. It is designed to retrieve data through SPARQL endpoints, and implements a wide range of fusion actions both on spatial properties of the entities and on non-spatial metadata. These include moving, rotating, scaling and aligning the geometries of the entities, combining multiple, semantically related properties, maintaining both descriptions of a property of two linked geospatial entities, etc. It also supports batch fusion actions, automatic classification of fused entities using OSM categories, and provides a map-based UI.

<https://github.com/SLIPO-EU/FAGI>

- **CityRisks Crime Mapper**

This is a set of applications developed in the context of the EU funded City.Risks project. Specifically, it comprises the following: (a) descriptive and predictive analytics over crime data with respect to a variety of location-based features, including demographics, Points of Interest, transportation data, geotagged photos, etc.; (b) a routing service that incorporates safety-related criteria during route computation; (c) a simulation engine for simulating and visualizing various scenarios for stolen item tracking via participatory sensing.

<http://cityrisks.imis.athena-innovation.gr>



Education

PhD / MSc / Diploma Thesis Co-supervision

IMIS members actively co-supervise several undergraduate and graduate university students, who often conduct their work in IMIS premises. As a result, in 2016 several MSc and Diploma thesis have been co-supervised by IMIS members, who often serve as members in the respective examination committees.

IMIS members also co-supervise PhD students. The following PhD dissertations have been completed in 2016:

- George Lambrianidis. **Advanced methods and algorithms for collecting and processing spatiotemporal data from social networks**. National Technical University of Athens. Collaborating researchers: [Dieter Pfoser](#), [Timos Sellis](#).
- Olga Gountouna. **Privacy protection in publishing non-relational data**. National Technical University of Athens. Collaborating researcher: [Manolis Terrovitis](#).
- Nikos Bikakis. **Personalized, Semantic and Exploratory Data Analysis**. PhD Thesis, National Technical University of Athens, 2016. Collaborating researchers: [Timos Sellis](#).

The following PhD students collaborated closely with IMIS members in their research during 2016:

- Theodora Galani. Topic: **Modeling and querying the data evolution and provenance**. Joint supervision with the National Technical University of Athens. Collaborating researchers: [Yannis Stavrakas](#), [George Papastefanatos](#).
- Danai Pla-Karidi. Topic: **Recommendation models using Social Networks**. Joint supervision with the National Technical University of Athens. Collaborating researcher: [Yannis Stavrakas](#).
- Giorgos Chatziogeorgakidis. Topic: **Big Data Management and Analysis**. Joint supervision with the University of Peloponnese. Collaborating researcher: [Spiros Athanasiou](#).
- Pantelis Chronis. Topic: **Development of Data Mining Methods for Temporal Databases**. Joint supervision with the University of Peloponnese. Collaborating researcher: [Spiros Athanasiou](#).
- Marios Meimaris. Topic: **Evolution and Query processing of RDF data**. Joint supervision with the University of Thessaly. Collaborating researcher: [George Papastefanatos](#).



- Giorgos Alexiou. Topic: **Entity disambiguation and data interlinking**. Joint supervision with the National Technical University of Athens. Collaborating researcher: George Papastefanatos.
- S. Maroulis. Topic: **Data visualization techniques**. Joint supervision with the National Technical University of Athens. Collaborating researcher: George Papastefanatos.
- Vassilis Kaffes. Topic: **Keyword queries on road networks**. Joint supervision with the University of Peloponnese. Collaborating researcher: Dimitris Skoutas.
- Konstantinos Zagganas. Topic: **Efficient techniques for data intensive analysis and processing in life sciences**. Joint supervision with the University of Peloponnese. Collaborating researchers: Theodore Dalamagas, Thanasis Vergoulis.
- Ilias Kanellos. Topic: **Data extraction from scientific publications and research analytics**. Joint supervision with the National Technical University of Athens. Collaborating researchers: Theodore Dalamagas, Thanasis Vergoulis.

Other Educational Activities

Other educational activities involving IMIS members include the following.

- Organization of the **CARARE Training Workshop**, Athens, 28-29 Jan 2016.
- Organization of the **ARIADNE summer school - Digital curation of archaeological knowledge**, Athens, 12-17 June 2016.
- IAESTE student exchange program – supervision. Vinayak Mathur. **Experimenting with Tweet Analysis Techniques on Tay Chatter Bot**. Collaborating researcher: Yannis Stavrakas.
- IAESTE student exchange program – supervision. Evanthia Tiggiri. **Designing and developing information extraction technologies for scientific publications related to miRNA molecules**. Collaborating researchers: Theodore Dalamagas, Thanasis Vergoulis.
- Thanasis Vergoulis taught the **Databases** course of the Department of Mathematics of the University of Patras.
- Manolis Terrovitis co-taught the **Spatial Databases** course of the GeoInformatics MSc program of NTUA.



Facts and Figures

Financial report

In 2016, IMIS continued its participation in EC funded research and development projects. The key economic indicators regarding the expenses and revenues in 2016 are shown in Table 1 and their distribution in categories is illustrated in Figure 1. We can see that the highest percentage of the revenue stream, more than 90%, comes from EC funded projects. It is important to note that the revenues from the activities of IMIS (EC and national funded projects) are more than 23 times the public expenditure received by IMIS.

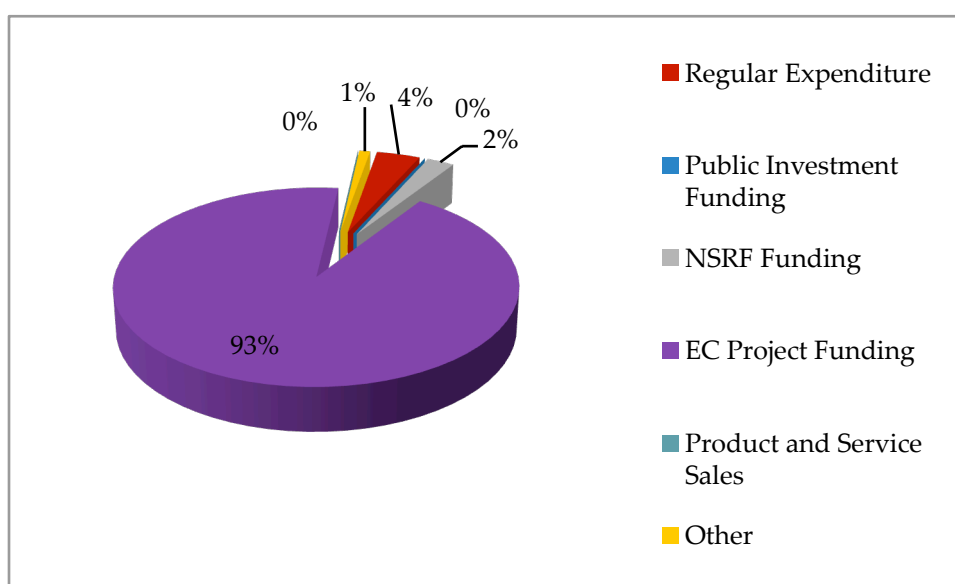


Figure 1. Distribution of revenues in 2016

Table 1. Expenses and Revenues for 2016

Expenses 2016	
Travel Expenses	182.017,24
Operational Costs	47.521,17
Equipment	41.212,18
Other Expenses	1.331.822,28 ¹
Personnel fees and payments to third parties	2.066.189,95
Total	3.668.762,82

¹This number included the transfer of 1.118.970€ for the expenses of publications arising from FP7 projects in open access journals within the WP "Gold OA: FP7 post grant APCs Pilot" of the EC project "OpenAIRE2020 (GA No 643410)"

Revenues 2016	
Regular Expenditure	164.154,00
Public Investment Funding	
NSRF Funding	102.597,68
EC Project Funding	3.804.217,79 ²
Product and Service Sales	
Other	40.983,75
Total	4.111.953,22

Table 2 shows the revenues of IMIS since 2013. Overall we can see that the revenues from EC funded projects have been increasing every year, reaching more than 3.8M euros in 2016, an increase of 62% from 2015.

Table 2. Revenues from 2013 to 2016

Revenues				
	2013	2014	2015	2016
Regular Expenditure	96.028,65	76.226,10	82.307,98	164.154,00
Public Investment Funding	149.497,86	454.421,97	185.186,15	-
NSRF Funding	595.941,35	790.948,38	2.055.402,11	102.597,68
EC Project Funding	1.258.297,00	1.916.199,47	2.335.227,81	3.804.217,79
Product and Service Sales	42.039,72	291.943,42	368.338,97	-
Other	6.933,61	85.508,66	27.506,22	40.983,75
Total	2.148.738,19	3.615.248,00	5.053.969,24	4.111.953,22

Figure 2 shows a comparison of the revenues in 2015 and 2016, where we can see that the revenues coming from participation in European projects has increased significantly. Due to the fact that a big part of the NSRF Funding within the Partnership Agreement 2007-2013 was paid off in 2015, the revenues from NSRF Funding in 2015 are much higher than the ones in 2016. Increased NSRF Funding is expected the years to come from the participation of IMIS in national funded projects within the new Partnership Agreement 2014-2020.

² This number included the funding of 1.110.773€ to cover the expenses of publications arising from FP7 projects in open access journals within the WP "Gold OA: FP7 post grant APCs Pilot" of the EC project "OpenAIRE2020 (GA No 643410)"

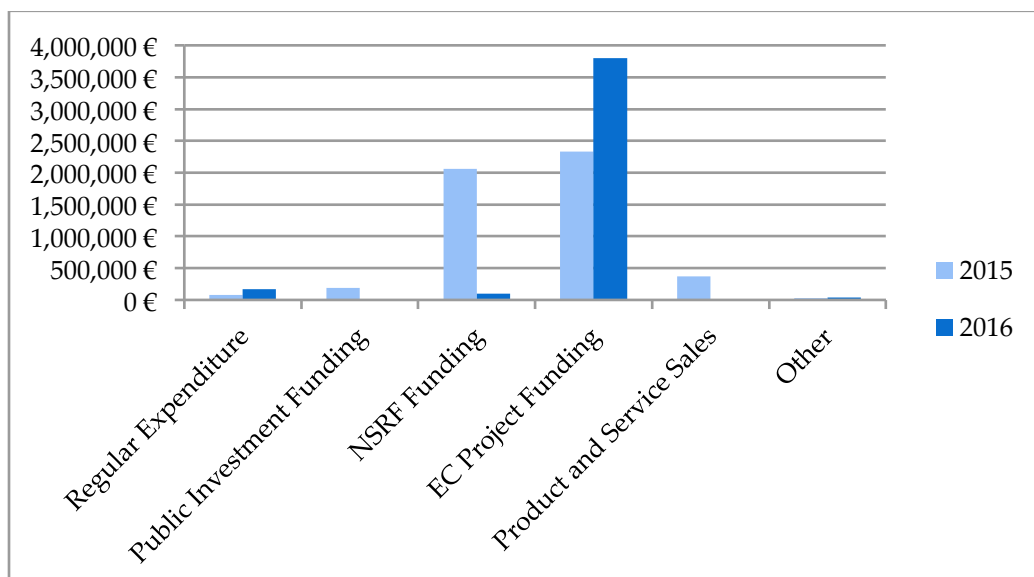


Figure 2. Comparison of revenues in 2015 and 2016

Staff

Director & Researchers

- Yannis Ioannidis, Professor, University of Athens, General Director of ATHENA RC
- Stylianos Sartzetakis, Researcher A
- Theodore Dalamagas, Researcher B
- Yannis Stavrakas, Researcher B

Associate Researchers / Professors

- Michael Chatzopoulos, Professor, University of Athens
- Panos Constantopoulos, Professor, AUEB
- Marinos Kavouras, Professor, NTUA
- Michalis Vazirgiannis, Professor, AUEB
- Yannis Kotidis, Associate Professor, AUEB
- Nectarios Koziris, Professor, NTUA
- Christos Papatheodorou, Professor, Ionian University
- Vassilis Vassalos, Associate Professor, AUEB
- Ion Androutsopoulos, Assistant Professor, AUEB
- Costis Dallas, Assistant Professor, Toronto University
- Antonis Deligiannakis, Assistant Professor, Technical University of Crete
- Panos Vasileiadis, Assistant Professor, University of Ioannina
- Phoebe Kountouri, Assistant Professor, AUEB
- Timos Sellis, Professor, Swinburne University of Technology
- Spiros Skiadopoulos, Assistant Professor, University of Peloponnese
- Vasilios Christofidis, Professor, University of Crete
- Eli Katsiri, Assistant Professor, Democritus University of Thrace
- Pavlos Kavouras, Professor, University of Athens
- Evgenia Roditi, Assistant Professor, University of Athens
- Georgios Giaglis, Professor, AUEB

- Damianos Chatziantoniou, Assistant Professor, AUEB
- Dimitrios Theodoratos, Associate Professor, NJIT USA
- Alexios Delis, Professor, University of Athens
- Ioannis Emiris, Professor, University of Athens
- Emmanouil Koubarakis, Professor, University of Athens
- Georgios Prezerakos, Professor, Technological Education Institute of Piraeus
- Maria Giaoutzi, Professor, NTUA
- Antonis Kokosis, Professor, NTUA
- Constantinos Kontogiannis, Assistant Professor, NTUA
- Artemis Chatzigeorgiou, Professor, University of Thessaly
- Ioannis Dagklis, Professor, University of Athens
- Apostolos Zarras, Assistant Professor, University of Ioannina
- Dionisios Christopoulos, Professor, Technical University of Crete
- Nikos Mamoulis, Assistant Professor, University of Ioannina
- Aristidis Dokoumetzidis, Assistant Professor, University of Athens
- Nikitas Pittis, Professor, University of Piraeus
- Yannis Theodoridis, Professor, University of Piraeus

Postdoctoral Researchers / Scientific Collaborators

- Dimitris Gavrilis, Elec. Comp. & Eng. University of Patras, PhD University of Patras
- Agiatis Benardou, BA, PhD in Ancient History, King's College London
- Fotis Karagiannis, Elec. & Comp. Eng. NTUA, PhD NTUA
- George Papastefanatos, Elec. & Comp. Eng. NTUA, PhD NTUA
- Dimitris Skoutas, Elec. & Comp. Eng. NTUA, PhD NTUA
- Manolis Terrovitis, Elec. & Comp. Eng. NTUA, PhD NTUA
- George Giannopoulos, Elec. & Comp. Eng. NTUA, PhD NTUA
- Thanasis Vergoulis, Elec. Comp. & Eng. University of Patras, PhD NTUA
- Constantinos Arkoumanis, Elec. & Comp. Eng. NTUA, PhD NTUA
- Alexandros Efentakis, Dept. of Physics UoA, PhD NTUA
- Charalampos Dimitropoulos, Elec. & Elec. Eng. City University London, PhD University of London
- Eleni Zacharia, Dept. of Inform. and Telecom. UoA, PhD UoA
- Alexandros Belesiotis, Dept. of Inform. University of Piraeus, PhD University of Edinburgh
- Christos Konaxis, Dept. of Mathematics UoA, PhD UoA
- Dimitris Pierrakos, Dept. of Physics UoA, PhD UoA
- Stella Tsani, Dept. Of Economics UoA, PhD University of Reading
- Christos Pateritsas, Elec. & Comp. Eng. NTUA, PhD NTUA
- Constantinos Nikas, Elec. & Comp. Eng. NTUA, PhD University of Manchester
- Sofia Karagiorgou, Computer Science, UoC, PhD NTUA

Administrative and Technical Personnel

- Nelly Apostolopoulou, Mech. Eng. NTUA, Msc NTUA, Head of Administration
- Eva Paradissi, Secretarial Support
- Panagiotis Georgantas, Elec. & Comp. Eng. NTUA, Technician
- Maria Chrysovitsioti, Dept. of Comm., Media and Culture, Administrative Support

Research Staff

- Stavros Aggelis, Dept. of Archives, Libr. Science & Museum Studies, Ionian University
- Spyros Athanasiou, Elec. & Comp. Eng. NTUA



- Michalis Alexakis, Elec. & Comp. Eng. NTUA
- Georgios Kakalettris, BA, Msc, Dept. of Inform. and Telecom. UoA
- Panayota Koltsida, BA, Msc, Dept. of Inform. and Telecom. UoA
- Ioannis Kouvaras, Elec. & Comp. Eng. NTUA
- Natalia Manola, Dept. of Physics, UOA
- Stefania-Varvara Martziou, BA, Msc, Dept. of Inform. and Telecom. UoA
- Konstantinos Patroumpas, Comp. & Inform. Eng. University of Patras, Msc NTUA
- Eleni Petra, Comp. & Inform. Eng. University of Patras, BA University of Manchester
- Theodoros Stratiotis, Comp. & Inform. Eng. University of Patras
- Thanassis Gentimis, Elec. & Comp. Eng. NTUA, Msc UoA
- Stelios Manousopoulos, Elec. & Comp. Eng. NTUA
- Vassilis Kaffes, Elec. & Comp. Eng. NTUA
- Georgios Karamanolis, Dept. of Inform, AUEB
- Nikos Georgomanolis, Comp. & Inform. Eng. University of Patras
- Loukas Renieris, Dept. of Inform, AUEB, MSc University of Piraeus
- George Chatzigeorgakidis Elec. & Comp. Eng. T.U.C., MSc D.T.U.
- Nikos Karagiannakis, Comp. & Inform. Eng. University of Patras
- Pantelis Chronis, Comp. & Inform. Eng. University of Patras
- Nikos Vitsas, Dept. of Inform, AUEB
- Stavros Maroulis, Elec. & Comp. Eng. NTUA
- Dimitris Tsitsigkos, Dept. of Inform. and Telecom. UoA
- Ilias Kanellos, Elec. & Comp. Eng. NTUA
- Antonis Korkofigkas, Elec. & Comp. Eng. NTUA
- Marios Meimaris, Elec. & Comp. Eng. NTUA, Msc NTUA, University of Edinburgh
- George Alexiou, University of Aegean, Msc AUTH
- Serafeim Chatzopoulos, BA, Msc, Dept. of Inform. and Telecom. UoA
- Eleni Afiontzi, Elec. & Comp. Eng AUTH, Msc, Dept. of Inform, AUEB
- Nefeli Chatzidiakou, BA in French, UoA, Msc, University of Aegean
- Leonidas Papachristopoulos, BA, Msc, Dep. of Archives, Libr. Science&Museology, Ion.University
- Constantinos Kravvaritis, Dept. of Inform. and Telecom. UoA, Msc, Dept. of Inform, AUEB
- Vangelis Nomikos, Dept. of Inform. and Telecom. UoA
- Efthimia Paza, BA, Msc, Department of Archives, Library Science & Museology, Ionian University
- Eliza Papaki, History and Archaeology, University of Cyprus, Msc Trinity College Dublin
- Omiros Metaxas, BA, Msc, Dept. of Inform. and Telecom. UoA
- Stamatios Christoforidis, Dept. of Inform. and Telecom. UoA
- Argiro Kokogiannaki, Dept. of Inform. and Telecom. UoA
- Antonis Lebesis, BA, Msc, Dept. of Inform. and Telecom. UoA
- Katerina Iatropoulou, BA, Msc, Dept. of Inform. and Telecom. UoA
- Orfeas Aidonopoulos, Comp. & Inform. Eng. Univ. of Patras, Msc Dept. of Inform. & Telecom, UoA
- Eleftherios Levantis, Dep. of Law, UoA, Msc Environmental Management EAEME
- Ioannis Souliotis, Dep. of Intern. & European Econ. Studies, AUEB, Msc University of Copenhagen
- Constantinos Zagganas, Elec. & Comp. Eng. NTUA
- Constantinos Theocharidis, Dep. of Electrical and Computer Engineering, University of Thessaly
- Anastasios Giannakopoulos, Dept. of Inform. and Telecom. UoA



- Danae Pla-Karidi, Elec. & Comp. Eng. NTUA
- Vangelis Bartzos Dep. of MBG, DUTH, Dep. of Math., Univ. of Patras, Msc École normale supérieure, Lyon, Msc Université 1, Lyon
- Eleni Koulocheri, Dep. of Math UoA, Msc, Hellenic Open University
- Sofia Baltzi, Dept. of Inform. and Telecom. UoA
- Angelos Tzotsos, School of Rural and Surveying Engineering, NTUA
- Maria Vagianou, Dept. of Inform. University of Piraeus, Msc UoA
- Ioannis Foufoulas Dept. of Inform. and Telecom. UoA, Msc UoA
- Clement Laroche, Fundamental Mathematics Universite Paul Sabatier, Msc Universite Paul Sabatier, Mathematics for Research Universite Paul Sabatier
- Diamantis Tziotzios, Dep. of Electrical and Computer Engineering, DUTH
- Ioannis Diplas, Dept. of Inform. and Telecom. UoA
- Dimitra-Nefeli Makri, Applied Informatics, UoM, Msc, Dept. of Inform, AUEB
- Nikon Gasparis, BA, Msc, Dept. of Inform. and Telecom. UoA
- Glikeria Katsari, BA, Msc, Dept. of Inform. and Telecom. UoA
- Anna Gogolou, BA, Dept. of Inform. and Telecom. UoA, Msc UoA
- Panagiota Georgiou, Elec. & Comp. Eng. NTUA, Msc Erasmus University of Rotterdam
- Theodora Galani, Elec. & Comp. Eng. NTUA
- Stefanos Geraggelos, Elec. & Comp. Eng. NTUA
- Ioannis Nikolentzos, Comp. & Inform. Eng. University of Patras
- Evangelos Aggelou, BA Dept. of Inform AUEB, Msc University of Amsterdam, Msc NTUA
- Nikolaos Panagiotou , Dept. of Inform. & Telecom. UoA, Msc Artif. Intelligence Univ. of Edinburgh
- Georgios Sidiropoulos, Dept. of Inform, AUEB
- Tsabikos Livisianos, Dept. of Inform. and Telecom. UoA
- Constantina Galouni, Dept. of Inform. and Telecom. UoA
- Nikolaos Zigouras, Comp. & Inform. Eng. Univ. of Patras, Msc Artif. Intelligence Univ. of Edinburgh
- Vassiliki Anastasopoulou, Dep. of Law, AUTH
- Michail Psallidas, Comp. & Inform. Eng. University of Patras
- Gioulan Ametoglou, Dept. of Inform. and Telecom. UoA



Contact

Institute for the Management of Information Systems (IMIS)

ATHENA Research and Innovation Center in Information, Communication and Knowledge Technologies

Postal Address:

Artemidos 6 & Epidavrou
15125, Maroussi, Athens
Greece

Phone: +30 210 6875403

Fax: +30 210 6856804

e-mail: contact@imis.athena-innovation.gr

Most current information about activities and contacts for the staff are available on the IMIS website:

www.imis.athena-innovation.gr

