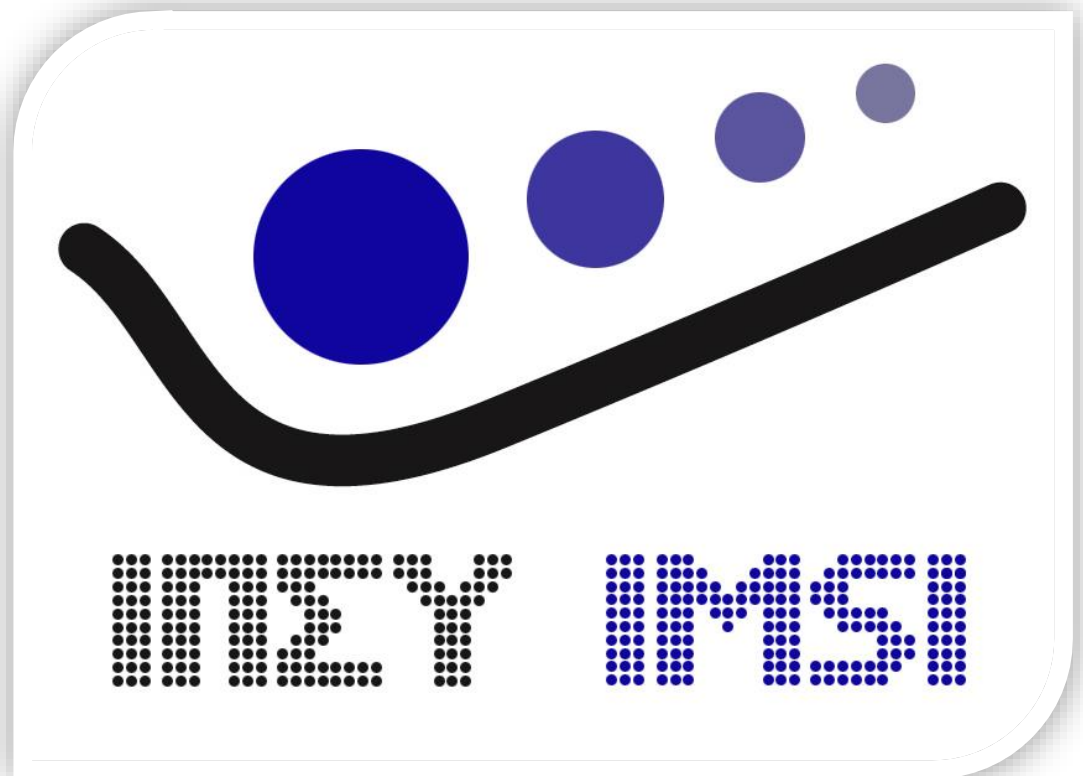
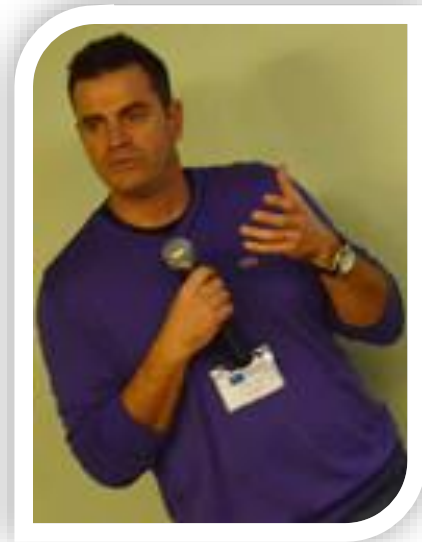


IMSI annual report 2018



Preface

Welcome to the Information Management Systems Institute (IMSI) of the ATHENA Research and Innovation Centre! Established in 2007, IMSI 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, IMSI researchers have been very successful in attracting and implementing numerous cutting-edge research & development projects, at both the national and international level; furthermore, IMSI 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 2018, IMSI 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, IMSI 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, IMSI 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, Information Management Systems Institute (IMSI)

ATHENA Research and Innovation Centre

Athens, Greece, 1/10/2019



Contents

| | |
|-------------------------------|----|
| Preface | 1 |
| Profile | 3 |
| Infrastructures..... | 6 |
| Highlights | 8 |
| Research Directions | 10 |
| Projects | 23 |
| Publications | 50 |
| Dissemination Activities..... | 54 |
| Systems | 56 |
| Education..... | 63 |
| Facts and Figures | 65 |
| Contact | 72 |



Profile

The Information Management Systems Institute (IMSI) is one of the research institutes of 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. IMSI was founded in 2007 with the mission to conduct research in the area of data management and large-scale information systems. In 2009, the Digital Curation Unit, which had been established in 2007 in "ATHENA" R.C, became a unit of IMSI.

The research at IMSI 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 IMSI is to conduct research, develop applications and products and to offer services in the areas of information management and large-scale information systems. IMSI 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 IMSI are:

- to participate and carry 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.
- to produce experimental and industrial prototypes, and to develop new innovative products in cooperation with industry.
- to transfer knowledge on research and development topics and to offer training and certifications on information technology and related processes for the management of information and information systems.

To attain these objectives, IMSI 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 IMSI. 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 Digital Curation Unit (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 IMSI departments are supported by the “ATHENA” R.C. Economic and Administration Office.

Infrastructures

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 the 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 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 territories in the world.



Highlights

Events

- The H2020 project City.Risks (<http://cityrisks.eu>) was successfully completed in April 30, 2018. The final review of the project took place in Rome on May 18, 2018. The review included a public event, organized by Roma Capitale, that was open to citizens and stakeholders interested in the themes of urban security and safety. Project Coordinator from IMSI: Dimitris Skoutas.



- A new project, titled “Visual Facts - Democratizing Visual Analytics, A Self-Service Platform for Big Data Exploration” funded by the 1st Call for Postdoc Research Projects - Hellenic Foundation for Research and Innovation (ELIDEK) has started in October 2018. Project Coordinator from IMSI: George Papastefanatos.
- A new collaboration between IMSI and the National Centre for Social Research (EKKE) has started in June 2018 under the project “Design and Development of the YouWeP platform”. The project aims at developing novel data services for social sciences. Project Coordinator from IMSI: George Papastefanatos.
- Organisation of the [first Open Science Symposium in Greece](#), entitled “Open Science in the Greek research ecosystem: research processes, research data and partnerships”. The Symposium called all major Greek academic and research stakeholders to discuss the current state of open access and research data management in Greece, identify challenges and propose pragmatic solutions. Moreover, the Symposium was the occasion for better communication and collaborations as well as preparations for the formulation of a [bottom-up national Open Science Task Force](#), lead by Athena Research Center. (Read more on the outcomes [here](#))

- [APOLLONIS](#), the Greek Infrastructure for Digital Arts, Humanities and Language Research and Innovation, the union of CLARIN-EL and DARIAH-GR, held its kickoff meeting at Athena R.C. in Athens in February 2018. Coordinator from IMSI: Panos Constantopoulos.
 - The EU (H2020-MSCA-ITN-2015) project ARCADES organized successfully the following two events during 2018:
 - a. The 2nd Learning Week on “Business Modelling” that took place at the INRIA research center at Sophia Antipolis, France on March 19th.
 - b. The 2nd Doctoral School & ESR Days that took place in Barcelona on September 3rd.
- Project scientific coordinator: Ioannis Z. Emiris

Awards

- **Prof. Minos Garofalakis, IMSI Director, has awarded the ACM Fellow title (2018)** “for contributions to data processing and analytics, particularly data streaming, approximation and uncertainty”. This is the ACM's most prestigious member grade that recognizes the top 1% of ACM members for their outstanding accomplishments in computing and information technology and/or outstanding service to ACM and the larger computing community.
- The publication “Nikos Bikakis, Stavros Maroulis, George Papastefanatos, and Panos Vassiliadis - RawVis: Visual Exploration over Raw Data” received the **best paper** award at the 22nd European Conference on Advances in Databases and Information Systems held in Budapest, Hungary, September 2–5, 2018.
- Georgia Koutrika, research director at IMSI, became an ACM Senior Member and a Distinguished Speaker of the ACM.



The Program Committee of the
22nd International Conference
on Advances in Databases and Information Systems
has granted the ADBIS 2018 Best Paper Award to
**Nikos Bikakis, Stavros Maroulis,
George Papastefanatos, and Panos Vassiliadis**
"RawVis: Visual Exploration over Raw Data"

As the publisher of the ADBIS 2018 Proceedings,
Springer furnishes this prize with a financial reward of Euro 100.-

Budapest, Hungary – September 2018

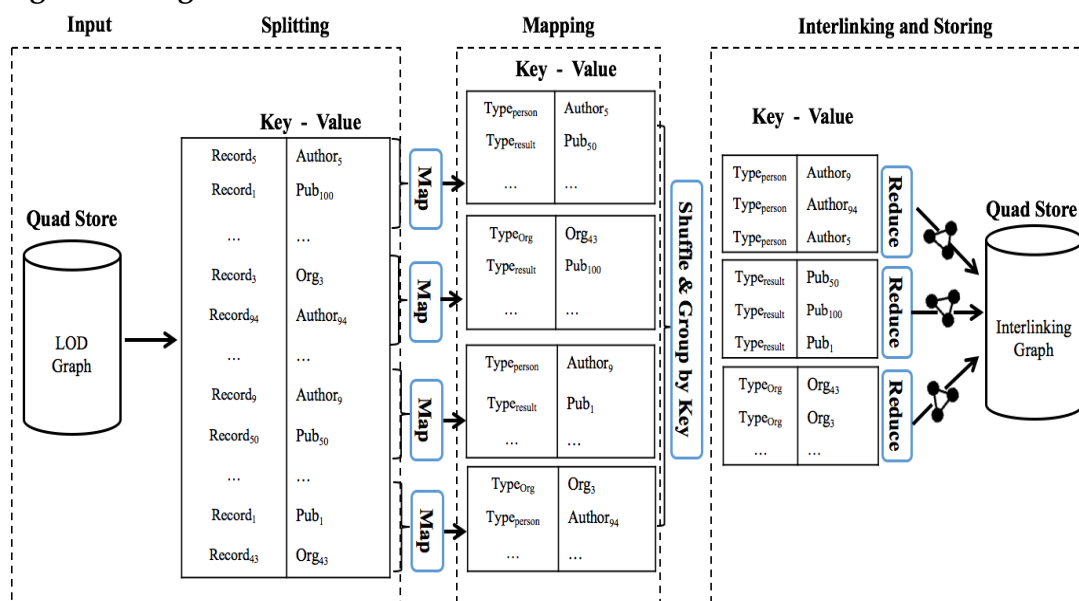
Research Directions

IMSI research activities fall into the following areas.

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, Exploration & 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. In addition, a parallel direction of our work aims at enabling efficient and interactive visual analysis of raw data files (e.g., csv, json, etc). We have developed in-memory data structures that address the visual needs and enable users to perform several visual exploration scenarios. These techniques have been integrated in the RawViz tool. 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 IMSI, 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.

- **Tweet Personalized Recommendations**

Twitter users get the latest tweets of their followees on their timeline. However, they are often overwhelmed by the large number of tweets, which makes it difficult for them to find interesting information among them. In this work, we focused on an efficient semantic recommendation method that helps users filter the Twitter stream for interesting content. The foundation of this method is a Knowledge Graph (KG) that can represent all user topics of interest as a variety of concepts, objects, events, persons, entities, locations and the relations between them. Our method uses the KG and graph theory algorithms not yet applied in social network analysis in order to construct user interest profiles by retrieving semantic information from tweets. Next, it produces ranked tweet recommendations. In addition, we use the KG to calculate interest similarity between users, and we present a followee recommender based on the same underlying principles. An important advantage of our method is that it reduces the effects of problems such as over-recommendation and over-specialization. As another advantage, our method is not impaired by the limitations posed by Twitter on the availability of the user graph data.

- **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 anonymization techniques on their data. 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. In the past, emphasis was given to genomic data related to miRNAs sequencing process and their analysis.



IMSI, Univ. of Thessaly

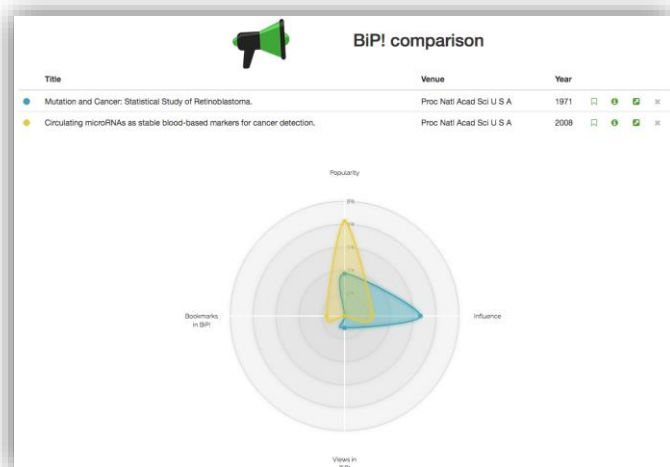
and GRNET have designed, implemented and maintained a research infrastructure for genomic data management, oriented to processing, analysis and visualization of computationally predicted miRNA targets (<http://diana.imis.athena-innovation.gr>). Part of the work is being integrated into the ELIXIR-GR 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. In addition, we are currently leading the design and implementation of the ELIXIR-GR Cloud Infrastructure, which is based on containerization technologies to support all computational needs of the ELIXIR-GR community. The relevant 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 for Scientific Publications**

Scientific publications retrieval and ranking based on their impact has been a long-established research topic, especially due to the massive number of papers being published annually.

We are working on technologies to identify high-impact research output, focusing on the study of impact aspects like long-term and short-term impact. Our approach is based on PageRank adaptations for citation

networks and machine learning on paper metadata. We also apply text mining techniques on scientific texts to automatically extract useful knowledge. Finally, we develop powerful information retrieval systems that are based on the previously described technologies to provide useful services to the research & science community (e.g., BIP! Finder and SciTo visualiser).



- **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 ELIXRI-GR. Note that IMSI, 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. IMSI 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 IMSI 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**

We have launched 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 minimised 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.

Semantic Web Technologies

Current (ongoing) research directions in this area include:

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

Another direction of research is related to the application of efficient 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 modelling). 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.

- **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. Some emerging directions, concern online entity resolution techniques, which aim at integrating the blocking\metablocking\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.

User-Data Interaction Systems

New (emerging) research directions in this area include:

- **Intelligent Data Interfaces to Databases**

The advent of computers drastically improved mankind's opportunities to information access. Unfortunately, the volume and complexity of available data preclude easy access to information for most people but the savviest technical experts. For the majority, standard keyword-search interfaces allow users to ask simple questions and receive answers to them, without however actively assisting their users in satisfying their information needs. The volume of data and the need for data democratization call for shift from the classical «query-based information access paradigm» popularized by existing systems to a novel information access paradigm, where the system takes upon a more conversational and active role in helping users effectively explore data of varying quality, complexity, and relevance. We are looking into several aspects of such systems:

(a) *conversational systems*: the system converses with the user not only returning results to her queries but offering advice, explanations and suggestions in a natural language interaction with the user

(b) *data exploration*: the system actively guides the user through their information access endeavor by offering just-in-time recommendations, exploration options, and help as needed

(c) *interaction in context*: the system keeps track of the context (e.g., goals, background information, earlier queries) of a user quest, and it can respond and adapt accordingly

(d) *proactive systems*: the system anticipates user needs proactively presenting information and suggestions without the user explicitly prompting the system, thus promoting continuous and easier learning.

- **Modern Recommender Systems**

Recommender systems provide advice on items that may be of interest to a user (e.g., movies, products, travel, and leisure activities) by learning user preferences and relationships between users and items, and they have become a salient part of Amazon, Google, Netflix and other web sites. Early recommender systems used simple computational or machine learning models to learn user preferences (content-based approaches) or to capture relationships between users or items (collaborative filtering approaches) to make recommendations. The landscape of recommender systems has drastically shifted from these early systems. Novel algorithmic approaches, including matrix factorization, multi-armed bandits and deep learning, show outstanding results and open up new opportunities and research paths. We are looking into this area from:

(a) *The data perspective*: where the objective is to explore non-conventional types of data and inputs to recommendations,

(b) *The problem perspective*: where the objective is to study recommendation problems beyond typical ones focusing on user consumption problems (selecting movies to watch, products to buy, applications to download, and so forth). Such problems include data exploration, query optimization, visualization, data integration, and workflow design among others, where the purpose is to select tuples, queries, views, exploration actions, query plans, visualization graphs etc.

(c) *The methods perspective*: where the objective is to build on the newest algorithmic developments in the field of recommender systems.

Geospatial Data Management

Current (ongoing) research directions in this area include:

- **Identifying Areas of Interest**

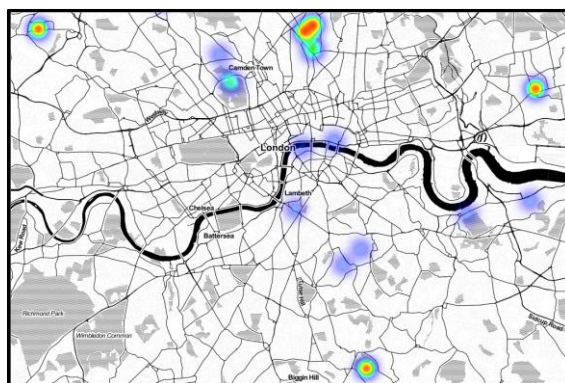
Large amounts of user-generated content are becoming available on the Web daily, with an increasing portion of it being associated with geospatial information. Typical examples include community-based mapping projects, such as OpenStreetMap, databases of Points of Interest (POIs), e.g., from Wikipedia or Foursquare, geotagged photos, e.g., from Flickr or Instagram, etc. This information constitutes a valuable

source for discovering and exploring locations and areas of interest, with numerous applications in location-based services, geomarketing, trip planning, and other domains.



Given a set of geospatial objects, our work has focused on the Best Region Search (BRS) problem, which aims at finding the optimal placement of a fixed-size rectangle so that the value of a user-defined utility function over the enclosed objects is maximized. We have introduced the top-k BRS problem and we have presented a method for

efficiently and progressively computing the next best result for any number k of results requested by the user. Our approach can also accommodate additional user-defined constraints. In particular, we support the requirement of computing the next best result that has no or little overlap with the already retrieved ones, which reduces repetition and redundancy in the result set. Our experiments have shown that our algorithms are efficient and scalable.



Moreover, we have applied spatial clustering and autocorrelation techniques to identify various types of hot spots (e.g., business, crime, etc.) in collections of geospatial data.

- **Spatio-Textual Queries and Analytics**

We have studied the problem of matching and clustering users based on their geolocated posts. First, individual posts are matched according to spatial distance and textual similarity thresholds. Then, we define user similarity as the ratio of posts that match between two users. Based on these criteria, we have introduced efficient algorithms for identifying pairs of matching users in large datasets. Further, we have developed a method for identifying spatio-textual user clusters, based on the Louvain method for community detection. Our algorithms operate on a user graph where edge weights represent spatio-textual user similarities. Since the exact user similarity graph can be prohibitively expensive to compute, we exploit our similarity join algorithms to derive efficient methods that reduce execution time both by avoiding to compute exact similarity scores and by reducing the number of similarity calculations performed. The proposed method provides a trade-off between computation time and quality of detected clusters.

- **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.

- **Spatio-textual Data Integration and Annotation**

Spatio-textual data integration is an open and quite challenging research problem with high value in several commercial applications, including geomarketing, navigation and social networks. It mainly comprises interlinking (deduplicating) entity descriptions from different sources/datasets. A complementary problem, annotation of spatio-textual entities, e.g. POIs, with categories is a respectively important problem for the aforementioned fields.

Spatio-textual data integration is significantly hindered by the large heterogeneity of entities: POIs, addresses, toponyms, land parcels are all types of spatio-textual entities, however, each type carries a different set of attributes, with varying importance in determining an entity's matching with another entity. The heterogeneity also expands to the namings of entities, resulting in quite diverse alternative descriptions regarding the same entity. Most methods in the literature do not delve in the specificities of the attributes (e.g. names, relations to other entities) of spatio-textual entities, rather than apply generic similarity measures and classification methods for interlinking and annotation. On commercial level, the shortcomings are even larger, since very simplistic similarity measures, rules and filters combined with largely manual processes are usually adopted to solve the problems.

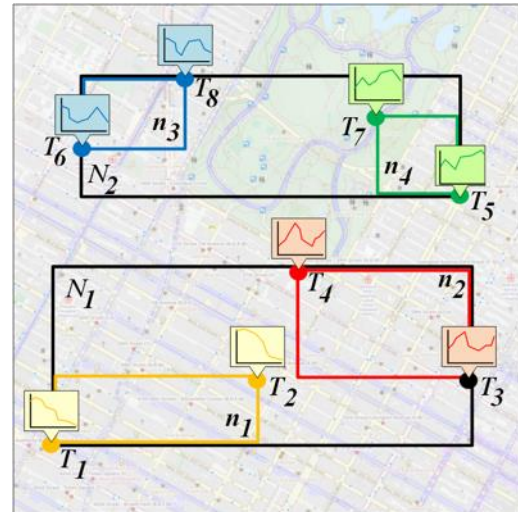
Our work emphasizes on analyzing the spatio-textual data at hand (toponyms, addresses, etc.) and deriving novel, domain-specific similarity measures, training features and Machine Learning methods for solving several real-world spatio-textual data integration problems. Currently, our methods are materialized as open source python libraries for solving four distinct problems via Machine Learning techniques: toponym interlinking, POI classification, fusion of geocoding sources and land parcel integration (<https://github.com/LinkGeoML>). Our goal is to build a large pool of training features and Machine Learning processes/pipelines for handling an extended set of spatio-textual data integration and annotation problems, applicable in real-world processes and applications of GIS companies.

New (emerging) research directions in this area include:

- **Similarity Joins on Geolocated Time Series**

Time series associated with specific locations, such as visitor check-ins or sensor readings, have increased in size and popularity in various domains. Although several works have focused on efficient time series similarity search, there has been limited attention to the inherent challenge that geolocated time series introduce for hybrid queries on both spatial proximity and time series similarity.

Our work addresses the problem of hybrid similarity joins over geolocated time series. We have proposed algorithms for performing such join operations using different types of indices, including spatial-only, time series-only, and hybrid indices. Such centralized indexing schemes can cope well with moderate data volumes but face scalability issues when the dataset size increases significantly. To overcome this problem, we have developed a MapReduce-based processing scheme with space-driven partitioning. Our algorithm leverages a hybrid index for geolocated time series to efficiently execute similarity joins locally within each partition and minimize the amount of data that needs to be shuffled between processing nodes.



- Extracting value from (underutilized) Big Geospatial Data via Deep Learning**
 Boosting the data economy and supporting the development of data value chains has been a major goal of the EC during the last years. In this spirit, our future work focuses on building Deep Learning and Transfer Learning methods for extracting and modelling knowledge from underutilized geospatial datasets and exploiting it to solve real-world problems. As an indicative example, cadastration companies have obtained, through years of manual efforts, big datasets of annotated aerial images on rural, semi-urban and urban areas. The volumes of these datasets allow the effective training of Deep Learning models for object detection, semantic segmentation and instance segmentation. Such models are directly applicable and can produce great value in several land management applications, including Cadastrals and Land Parcel Identification Systems. To this end, several research challenges need to be addressed, including the varying resolution and quality of aerial images, the low separability of certain object classes in rural areas and the appropriate/optimal application of Transfer Learning methodologies on the above processes. Another real-world application involves the existence of large open and proprietary databases of toponyms and addresses that can be utilized to build domain specific, distributed representations of spatio-textual entities, exploitable in various data integration and annotation tasks.

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 programme 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: modelling 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 modelling working practices and information behaviours in specific research and professional communities, especially in the humanities; user community modelling 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. At the national level, IMSI leads the APOLLONIS Greek Infrastructure for Digital Arts, Humanities and Language Research and Innovation (P. Constantopoulos, coordinator) resulting from the unification of CLARIN-EL and DARIAH-GR. At the european level, IMSI 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"). IMSI, through DCU, has been heavily involved in building the ARIADNE infrastructure for archaeology. It provides advanced aggregation services for the Europeana ecosystem. 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:

- **Digital storytelling**

The initiation of project SHARE3D marks a new line of work on issues of information delivery and publication as an integral part of the digital content lifecycle and establishing an efficient continuum from data curation to creative uses

and dissemination. The specific project focuses on enabling users to explore, create and experience 3D objects as stories.



Projects

EU R&D Projects



GRACIOUS - Grouping, Read-Across, Characterisation and classification framework for regulatory risk assessment of manufactured nanomaterials and Safer design of nano-enabled products

| | |
|------------------------|-------------------------------------------------------------------------|
| Project manager | Georgia Tsiliki |
| Coordinator | Heriot-Watt University |
| IMIS - funding | 124 037.5 Euros |
| Programme | H2020-NMBP-2017-two-stage |
| Start date | 1/1/2018 |
| Duration | 3.5 years |
| Website | https://www.h2020gracious.eu |

The GRACIOUS project will develop a highly innovative science-based framework that supports the assessment of risk posed by the ever increasing array of nanomaterials on the market and under development. The framework will streamline the process for assessing their risk by logically grouping nanomaterials thereby allowing extrapolation between (read-across) nanomaterials and reducing the need to assess exposure to and toxicity on a case by case basis. The project will work continuously with stakeholders in an iterative cycle of design, testing and refinement to ensure that the Framework effectively meets the needs of both regulators and industry. Application of the Framework will allow movement away from the case-by-case risk assessment paradigm, thereby improving the efficiency of risk analysis and decision making for safer design of quality nano-enabled products.



Europeana DSI-3

| | |
|----------------------------|-----------------------------------------------------------------------------------------------------------------|
| Project coordinator | Dimitris Gavrilis |
| Coordinator | STICHTING EUROPEANA(EF) |
| IMSI - funding | 58,000 euros |
| Programme | Service Contract-“Deployment and Maintenance of Europeana DSI Core Services-SMART-2016/1019” |
| Start date | 1/9/2017 |
| Duration | 12 months |
| Website | https://pro.europeana.eu/project/europeana-dsi-3 |

The Europeana DSI-3 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 and DSI-2 actions under CEF. It is the third 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. IMSI participates in DSI-3 both as a technical partner with the MORE platform (supporting the CARARE network as part of DSI) and also through Europeana Research.



NEP4DISSENT – New Exploratory Phase in Research on East European Cultures of Dissent. European Cooperation in Science and Technology

| | |
|----------------------------|-----------------------------------------------------------------------------------------------------|
| Project coordinator | Costis Dallas |
| Coordinator | Polish Academy of Sciences (PL) |
| IMSI - funding | About 50,000 euros (The funding of COST actions is adjusted yearly) |
| Programme | Open Call Collection OC-2016-2 |
| Start date | 16/10/2017 |
| Duration | 4 years |
| Website | http://www.cost.eu/COST_Actions/ca/CA16213 |

Resistance and dissent in former socialist Europe 1945-1989 constitutes a remarkable chapter of Europe’s recent past, which not only informs in a decisive

way the identities of post-socialist societies, but has also reshaped the continent as a whole and still provides an important reference for contemporary social movements worldwide. The proposers of this Action believe that, after a period of growth and consolidation, this field of study and the respective domain of cultural heritage have stalled and fell short of its true significance. This state of affairs results from (1) the inheritance of Cold War-era conceptual distinctions, (2) confinement of research within national silos and (3) neglecting the problem of access to original archival sources for digitally enabled research due to both their heterogeneity and uneven investment in research infrastructures.



Wider Impacts and Scenario Evaluation of Autonomous and Connected Transport

| | |
|----------------------------|---------------------------------------------------------|
| Project coordinator | George Papastefanatos |
| Coordinator | University of Greenwich |
| IMSI - funding | - |
| Programme | COST ACTION |
| Start date | 30/7/2015 |
| Duration | 4 years |
| Website | https://wise-act.eu/ |

Autonomous vehicle (AV) trials are currently taking place worldwide and Europe has a key role in the development of relevant technology. Yet, very limited research exists regarding the wider implications of the deployment of such vehicles on existing road infrastructure, since it is unclear if and when the transition period will start and conclude. It is anticipated that improved accessibility and road safety will constitute the primary benefits of the widespread use of AVs, whilst co-benefits may also include reduced energy consumption, improved air quality or better use of urban space. Therefore, the focus of this COST Action is on observed and anticipated future mobility trends and implications on travel behaviour, namely car sharing, travel time use or residential location choice to name a few. Other important issues to be explored under different deployment scenarios are social, ethical, institutional and business impacts. To achieve this, it is essential to culminate co-operation between a wide range of stakeholders at a local, national and international level, including academics and practitioners. Consequently, this COST Action will facilitate collaboration within Europe and beyond about this emerging topic of global

interest.



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

| | |
|----------------------------|------------------------------------------------------------------------------------------|
| Project coordinator | Prof. Phoebe Koundouri |
| Coordinator | CSIC, Spain |
| IMSI - 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.



SLIPO - Scalable Linking and Integration of Big POI Data

| | |
|----------------------------|-------------------------------------------------------|
| Project Coordinator | Spiros Athanasiou |
| Coordinator | IMSI |
| IMSI - funding | 748 KEuros |
| Programme | IA, H2020-ICT-14-2016 |
| Start date | 1/1/2017 |
| Duration | 3 years |
| Website | http://www.slipo.eu |

SLIPO (www.slipo.eu) is an Horizon 2020 Innovation Action (IA) developing technologies for the scalable and quality assured integration of Points of Interest (POI) Big Data assets. POI data are the cornerstone of any application, service, and product even remotely related to our physical surroundings. From navigation applications, to social networks, to tourism, and logistics, we use POIs to search, communicate, decide, and plan our actions. The creation, update, and provision of POIs consists a multi-billion cross-domain and cross-border industry, with a value chain natively incorporating most domains of our economy. The evolved POI value chain introduces opportunities for growth, but also complexity, intensifying the challenges relating to their quality-assured integration, enrichment, and data sharing. However, the integration of POIs using current approaches remains labor-intensive and scalable only for domain-specific or small-scale efforts. In SLIPO, we argue that linked data technologies can address the limitations, gaps and challenges of the current landscape in integrating, enriching, and sharing POI data. Our goal is to transfer the research output generated by our work in project GeoKnow, to the specific challenge of POI data, introducing validated and cost-effective innovations across their value chain. Specifically, SLIPO develops effective and scalable software and processes for: transforming conventional POI formats and schemas into RDF data; interlinking POI entities from different datasets; enriching POI entities with additional metadata, including temporal, thematic and semantic properties; fusing Linked POI data in order to produce more complete and accurate POI profiles; assessing the quality of the integrated POI data; offering value added services based on spatial aggregation, association extraction and spatiotemporal prediction.



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

| | |
|----------------------------|-------------------------------------------------------------------------|
| Project coordinator | Dimitrios Skoutas |
| Coordinator | SPACE HELLAS |
| IMSI - 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, visualises 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 coordinator | Theodore Dalamagas (Deputy: Stelios Sartzetakis) |
| Coordinator | European Molecular Biology Laboratory - EMBL |
| IMSI - 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 organisation 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.



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

| | |
|------------------------|--------------------|
| Project manager | Theodore Dalamagas |
| Coordinator | FRAUNHOFER IAIS |

| | |
|-----------------------|---------------------------------------------------------|
| IMSI - 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.



ARCADIES - Algebraic Representations in Computer-Aided Design for complex Shapes

| | |
|----------------------------|-------------------------------------------------------------------|
| Project coordinator | Ioannis Z. Emiris |
| Coordinator | IMSI |
| IMSI - 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.



SHARE-3D

| | |
|----------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| Project coordinator | Dimitris Gavrilis |
| Coordinator | Athena R.C. |
| IMSI - funding | 154,615 euros |
| Programme | CEF , Action No: 2017-EU-IA-0139 |
| Start date | 1/9/2018 |
| Duration | 18 months |
| Website | http://www.dcu.gr/en/share-3d-sharing-new-perspectives/ |



The Share-3D project is developing two applications, a Dashboard and a Storytelling tool, which will integrate APIs available from Europeana and Sketchfab to provide services for users to publish and re-use 3D and other digital materials. The aim is to use interest in 3D to encourage end-users to engage with content and make it accessible through Europeana in a creative and highly interactive way. In particular, the Action will create user friendly tools to upload 3D media and storytelling services that offer end-users interesting, informative and memorable experiences about European Culture. End-users will be able to browse cultural content, make links between related items, add descriptions and knowledge, and share the results with friends, colleagues, classmates, students, visitors and others. Lastly, the Action will showcase the potential of allowing users to both produce and consume cultural heritage content, and of giving users the opportunity to share their stories and share their discoveries.



ARKWORK - Archaeological practices and knowledge work in the digital environment

| | |
|----------------------------|-----------------------------------------------------------------------------|
| Project coordinator | Costis Dallas |
| Coordinator | Uppsala University (SE) |
| IMSI - 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.



OpenAIRE 2020

| | |
|------------------------|------------------------------------------------------|
| Project manager | Yannis Ioannidis |
| Coordinator | NKUA |
| IMIS - funding | 808.125€ |
| Programme | H2020-EINFRA-2014-1 (Grant Agreement no 643410) |
| Start date | 01/2015 |
| Duration | 30 months |
| Website | www.openaire.eu |

OpenAIRE2020 represents a pivotal phase in the long-term effort to implement and strengthen the impact of the Open Access (OA) policies of the European Commission (EC), building on the achievements of the OpenAIRE projects. OpenAIRE2020 expanded and leveraged its focus from (1) the agents and resources of scholarly communication to workflows and processes, (2) from publications to data, software, and other research outputs, and the links between them, and (3) strengthen the relationship of European OA infrastructures with other regions of the world, in particular Latin America and the U.S. OpenAIRE2020 continues and extends OpenAIRE's scholarly communication infrastructure to manage and monitor the outcomes of EC-funded research. It combines its substantial networking capacities and technical capabilities to deliver a robust infrastructure offering support for the Open Access policies in Horizon 2020, via a range of pan-European outreach activities and a suite of services for key stakeholders. It provides researcher support and services for the Open Data Pilot and investigates its legal ramifications. The project offers to national funders the ability to implement OpenAIRE services to monitor research output, whilst new impact measures for research are investigated. OpenAIRE2020 engages with innovative publishing and data initiatives via studies and pilots. By liaising with global infrastructures, it ensures international interoperability of repositories and their valuable OA contents.



OpenAIRE Advance

| | |
|------------------------|------------------------------------------------------|
| Project manager | Yannis Ioannidis |
| Coordinator | NKUA |
| IMIS - funding | 1.431.475€ |
| Programme | H2020-EINFRA-2017 (Grant Agreement no 777541) |
| Start date | 01/2018 |
| Duration | 24 months |
| Website | www.openaire.eu |

OpenAIRE-Advance continues the mission of OpenAIRE to support the Open Access/Open Data mandates in Europe. By sustaining the current successful infrastructure, comprised of a human network and robust technical services, it consolidates its achievements while working to shift the momentum among its communities to Open Science, aiming to be a trusted e-Infrastructure within the realms of the European Open Science Cloud. On the technical level OpenAIRE-Advance focuses on the operation and maintenance of the OpenAIRE technical TRL8/9 services, and radically improves the OpenAIRE services on offer by: a) optimizing their performance and scalability, b) refining their functionality based on end-user feedback, c) repackaging them into products, taking a professional marketing approach with well-defined KPIs, d) consolidating the range of services/products into a common e-Infra catalogue to enable a wider uptake.



OpenAIRE Connect

| | |
|------------------------|-------------------------------------------------------------------------|
| Project manager | Yannis Ioannidis |
| Coordinator | CNR |
| IMIS - funding | 480.500€ |
| Programme | H2020-EINFRA-2016-1 (Grant Agreement no 731011) |
| Start date | 01/2017 |
| Duration | 30 months |
| Website | https://connect.openaire.eu/ |

OpenAIRE - Connect is a project that aims to provide technological and social bridges, and deliver services enabling uniform exchange of research artefacts (literature, data, and methods), with semantic links between them, across research communities and content providers in scientific communication. It introduces and implements the concept of Open Science as a Service (OSaaS) on top of the existing OpenAIRE infrastructure, delivering out-of-the-box, ondemand deployable tools. OpenAIRE-Connect adopts an end-user driven approach (via the involvement of 5 prominent research communities), and enrich the portfolio of OpenAIRE infrastructure production services with a Research Community Dashboard Service and a Catch-All Notification Broker Service. The Research Community Dashboard Service serves research communities to (i) publish research artefacts (packages and links), and (ii) monitor their research impact. The Catch-All Notification Broker Service engages and mobilizes content providers, and serves them with services enabling notification-based exchange of research artefacts, to leverage their transition towards Open Science paradigms. Both services are served on-demand according to the OSaaS approach, hence are re-usable by different disciplines and providers, each with different practices and maturity levels, so as to favor a shift towards a uniform cross-community and cross-content provider scientific communication ecosystem.



EOSC pilot

| | |
|------------------------|-------------------------------------------------------------|
| Project manager | Yannis Ioannidis |
| Coordinator | STFC-UKRI |
| IMIS - funding | 500.650,00 |
| Programme | H2020-INFRADEV-2016-2 (Grant Agreement no 739563) |
| Start date | 01/2017 |
| Duration | 24 months (was given a 3 month extension) |
| Website | https://eoscipilot.eu/ |

The European Open Science Cloud will offer 1.7 million European researchers and 70 million professionals in science and technology a virtual environment with open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines by federating existing scientific data infrastructures, today scattered across disciplines and Member States.

The **EOSCpilot project** has been funded to support the first phase in the development of the European Open Science Cloud (EOSC). It will:

- Propose and trial governance frameworks for the EOSC and contribute to

- the development of European open science policy and best practice;
- Develop a number of demonstrators functioning as high-profile pilots that integrate services and infrastructures to show interoperability and its benefits in a number of scientific domains; and
- Engage with a broad range of stakeholders, crossing borders and communities, to build the trust and skills required for the adoption of an open approach to scientific research.



INSPIRED - The National Research Infrastructures on Integrated Structural Biology, Drug Screening Efforts and Drug target functional characterization

| | |
|------------------------|---------------------------------------------------------------------------------------------|
| Project manager | prof Yannis Ioannidis |
| Coordinator | NHRF |
| IMIS - funding | 140.000,00 Euros |
| Programme | 2014-2020 (EPAnEK) - Operational Programme competitiveness, Entrepreneurship and Innovation |
| Start date | 03/09/2018 |
| Duration | 4 years |
| Website | https://www.inspired-ris.gr/ |

INSPIRED is a national distributed research infrastructure unique in the field of Structural Biology that combines studies on bioactive (macro)molecules interactions and biomarkers identification. It offers services mainly in the field of biology, diagnostics and pharmacology, addressing the needs of the health sector with significant impact on agrofood and concerns a large number of organizations (potential users and collaborators). It bridges basic research with the Industry and SMEs supporting innovative actions by either providing services or in the frame of funded programmes. It comprises two complementary infrastructures: INSTRUCT-EL | UPAT-RISF. INSPIRED addresses the needs of the field of Health Sciences and Pharmaceuticals. The partners cover all the Greek Regions. It is a distributed Research Infrastructure across Greece.



FRESQO - Freshness REcording System for fish Quality Observation

| | |
|------------------------|----------------------------------|
| Project manager | prof Yannis Ioannidis |
| Coordinator | IMIS |
| IMIS - funding | 122.951,09 € |
| Programme | Fisheries and Maritime 2014-2020 |
| Start date | 04/05/2018 |
| Duration | 4,5 years |

Website

The main objective of the FRESQO project is the construction of an innovative product (hardware and software) which will allow the automatic recognition of the freshness of the most important commercial catches of the Greek market with possibility of expanding the detection facilities to an unlimited set of marketable and non-marketable fish. This product will essentially consist of a small (portable) spectral camera which, with a simple photograph of the fish, can provide specific indications about the freshness of a fish. The camera will communicate wired or wireless with a small-sized controller that will communicate with the specialized repository to collect and manage the relevant data as a supporting independent infrastructure (multiple capture images, chemical analysis data and organoleptic measurements as well as auxiliary data for export safer conclusions).



DARE- Delivering Agile Research Excellence

| | |
|------------------------|-----------------------|
| Project manager | prof Yannis Ioannidis |
| Coordinator | NCSR |
| IMIS - funding | 224.588,21 € |
| Programme | H2020-EINFRA-2017 |
| Start date | 01/01/2018 |

Duration 3 years

Website <http://project-dare.eu/>

DARE (Delivering Agile Research Excellence on European e-Infrastructures) aims to provide scientific communities with a unifying hyper-platform and development context to allow for user-friendly and reproducible carrying out of huge data-driven experiments, and rapid prototyping. DARE specifically addresses the requirements of innovating teams of research developers and scientists, who work on the intersection of software engineering and scientific domains, and on data, complexity and computing extremes. The size and complexity of scientific data, as well as the difficulty in formulating domain-specific solutions in reproducible and reusable ways, may often lead to throw-away, unsustainable end-user products, or long release cycles. This complexity increases exponentially with the size and diversity of input and produced data. Furthermore, widely used big-data technologies and analytics, while they are known to lead to increased productivity in commercial settings, they are often not taken advantage of in scientific. The requirement to deal with diverse exascale data resources dictates the need to ensure and increase productivity through the controlled disruption of the current modus operandi of European RIs. DARE aims to be the technological pivot for this transition, while providing transparent, traceable and developer-friendly bridges over existing infrastructures and services. Building on extensive experience in research e-infrastructures, semantification and the handling of metadata, and on big-data technologies and domain applications, DARE will equip teams of innovators with meaningful abstractions and tools allowing for rapid prototyping of reproducible and efficient research solutions. DARE will improve further and integrate tried and tested programmatic dataflow specification APIs, big-data technologies and provenance/datalineage solutions to address the requirements of European RIs, initially of EPOS, on Earth science, and IS/ENES2, on climate.



RDA 4.0- The European plug-in to the global Research Data Alliance

Project manager prof Yannis Ioannidis

Coordinator Trust-IT

IMIS - funding 100.525,00 €

Programme H2020-INFRA-SUPP-2017-1 (3rd Party)

Start date 01/03/2018

Duration 2.5 years

Website <https://www.rd-alliance.org/>

RDA Europe 4.0 addresses the INFRASUPP-02-2017 call targeting the area “European support to the Research Data Alliance, RDA” designing Europe’s contribution to implementation of an effective governance model and strategy in RDA global, while ensuring that RDA delivers on locally relevant issues. RDA Europe 4.0 focuses on the need for open and interoperable sharing of research data & on the need to build social, technical and cross-disciplinary links to enable such sharing on a global scale. It strives to do this with its community-driven and bottom-up approach launched since 2012. In fact, RDA Europe 4.0 directly builds on the current RDA Europe effort, by efficiently bringing in the organisations that implemented RDA Europe since 2012. The scope of RDA Europe 4.0 is to become the centrepiece for an EU Open Science Strategy through a consolidated European network of National Nodes, bringing forward an RDA legacy in Europe, providing skilled, voluntary resources from the EU investment to address DSM issues, by means also of an open cascading grant process. The ambitious, 27-month project is implemented by 5 beneficiaries (Trust-IT Services, Gottingen State University Library, the Digital Repository of Ireland at the Royal Irish Academy, the Digital Curation Centre and the RDA Foundation), skillfully supported by 9 National Nodes which carry out specific operational activities & act as national champions for their respective region. One of the specific goals of RDA Europe 4.0 is to complete a capillary European network by on-boarding additional 13 nodes by project end. Main Outputs: Expansion from 9 to 22 national nodes in EU; 7500+ individual members & 75 organisational members; 9 RDA recommendations as ICT technical specifications; consolidated programme for experts, early careers, ambassadors & adoption projects; 5 international RDA plenary meetings; integration with EOSC, ESFRI & other pan-EU initiatives; RDA Europe self-sustained after project completion.



MH-MD - My Health - My Data

Project manager prof Yannis Ioannidis

Coordinator LYNKEUS

IMIS - funding 419.375,00 €

Programme H2020-ICT-2016-1

Start date 01/11/2016

| | |
|-----------------|---------------------------------------------------------------------------|
| Duration | 3 years |
| Website | http://www.myhealthmydata.eu/ |

MyHealthMyData (MHMD) is a Horizon 2020 Research and Innovation Action which aims at fundamentally changing the way sensitive data are shared. MHMD is poised to be the first open biomedical information network centred on the connection between organisations and individuals, encouraging hospitals to start making anonymised data available for open research, while prompting citizens to become the ultimate owners and controllers of their health data. MHMD is intended to become a true information marketplace, based on new mechanisms of trust and direct, value-based relationships between EU citizens, hospitals, research centres and businesses.

National R&D Projects



VR-Park: Augmented reality platform for urban parks

| | |
|------------------------|----------------------------------------------------------------------------------------------------------------|
| Project manager | Yannis Stavrakas |
| Coordinator | H.A.O. "Demeter" |
| IMSI - funding | 204,000 euros |
| Programme | RESEARCH - CREATE - INNOVATE, Operational Programme Competitiveness, Entrepreneurship and Innovation 2014-2020 |
| Start date | 9/2018 |
| Duration | 30 months |
| Website | https://vr-park.org/ |

Urban parks and open green areas are important attractions of environmental interest to city residents and visitors. Careful and well-targeted promotion of these areas, not only enhances the importance of their existence in the urban space, but at the same time can assist in the development of alternative forms of "green tourism", and towards the direction of environmental awareness among citizens, which is particularly important nowadays and crucial for the future of the planet. New technologies are a key tool in enhancing the experience of touring urban parks, as they can make the tour much more attractive, highlighting interesting information about the flora and fauna of the park, as well as various other points of interest. At the same time, they can help guide the visitor inside relatively large

parks, and easily identify his/her paths, thus highlighting areas of the park that would otherwise be neglected. They can also assist park managers in organizing events, thus solving one of the key operating problems mainly of large-scale urban parks, which is the failure to exploit their entire site due to reduced or problematic accessibility. This project will build an integrated system that comprises an augmented reality mobile application for visitors of urban parks, and a corresponding park management web application for the managers of such park. Through the mobile app, an attractive, interactive touring environment will be created which will highlight the environmental and historical interest of those sites. At the same time, the web application will receive multimedia data from the users and will automatically collect anonymous data that may be useful to park managers to improve the visitors' touring experience and to better highlight the advantages of visiting such parks.



Design and Development of the YouWeP platform

Project manager George Papastefanatos

Coordinator IMSI - Athena RC

IMIS - funding 24,800 euros

Programme Contract

Start date 29/06/2018

Duration 2 years

Website <http://www.socioscope.gr/> , <http://www.youwho.gr>

Following a long-standing collaboration between IMSI-ATHENA R.C and the National Centre for Social Research (EKKE) the project aims at extending and providing new features to the Socioscope platform, a data platform developed by IMSI and used for hosting and visualizing data collected from social surveys by the National Centre for Social Research (EKKE).

Moreover, the project will implement the YouWeP platform, which will be a web platform for social and political scientists to conduct online surveys and visualize the results in interactive ways. The platform will offer new novel ways of conducting surveys based on chat-based questionnaires for interacting with users and getting their feedback. The platform will be evaluated during the implementation of a large-scale online survey for the identification of trends and behaviors in young people aged 18-29 years old.



Visual Facts - Democratizing Visual Analytics, A Self-Service Platform for Big Data Exploration

| | |
|------------------------|---------------------------------------------------------------------------------------------------|
| Project manager | George Papastefanatos |
| Coordinator | IMSI - Athena RC |
| IMIS - funding | 172,000 euros |
| Programme | 1st Call for Postdoc Research Projects - Hellenic Foundation for Research and Innovation (ELIDEK) |
| Start date | 16/10/2018 |
| Duration | 3 years |
| Website | https://visualfacts.imsi.athenarc.gr/ |

Self-service visual analytics is a new paradigm, widely promoted in modern corporate environments, in which business users are enabled and encouraged to directly manipulate (explore, blend, analyze) underlying data in rich visual ways, in order to derive insights from business information as quickly and efficiently as possible. Allowing less tech-savvy end users to make decisions based on their own queries and analyses, frees up the organization's business intelligence and information technology (IT) teams from the tedious work of data preparation.

The aim of VisualFacts is to develop a scalable platform for providing self-service visual analytic capabilities to a wide range of corporate and non-corporate users to access, explore, analyze open and privately-held data and collaborate on the analytic results of their work by sharing, annotating and reusing them in the form of visual facts.



LinkGeoML: Automatic and accurate interlinking of geospatial data using machine learning

| | |
|------------------------|------------------------------|
| Project manager | Giorgos Giannopoulos |
| Coordinator | IMSI |
| IMIS - funding | 199.300,00 Euros |
| Programme | RESEARCH – CREATE – INNOVATE |
| Start date | 9/7/2018 |

Duration 2 years

Website <https://linkgeoml.gr/>

LinkGeoML aims at researching, developing and extending machine learning methods, utilizing the vast amount of available, open geospatial data, in order to implement automated and highly accurate algorithms for interlinking geospatial entities. The proposed methods will implement novel training features, based on domain knowledge and on the analysis of open and proprietary geospatial datasets. Further, they will extend and specialize machine learning models on classification and similarity learning. The implemented technologies will be published as open source software and, also, will be integrated into existing, commercial applications for cadastration, geocoding and geomarketing, aiming at improving their functionality and increasing their commercial value and application domain. LinkGeoML comprizes a partnership between enterprises and research organizations, aiming to perform high quality, industrial research with a twofold purpose: Provide SMEs with useful geospatial data integration tools to solve real-world problems, and advance the state of the art on machine learning methods for geospatial data integration. To achieve these purposes, LinkGeoML identifies use cases based on real-world integration problems, elicited by its two industrial partners, and researches how machine learning-based, interlinking methods can be applied to these use cases and facilitate their handling, in real-world data. Check out our initially prescribed use cases, as well as our first results.



Big Data in Monitoring and Analyzing Sea Area Traffic: innovative ICT and analysis models

Project manager Theodore Dalamagas

Coordinator IMSI

IMIS - funding 150,000 euros

Programme RESEARCH - CREATE - INNOVATE, Operational Programme Competitiveness, Entrepreneurship and Innovation 2014-2020

Start date 7/2018

Duration 30 months

Website <https://www.imsi.athenarc.gr/el/projects/project/59>

The project will design and develop an innovative ICT platform for the collection and analysis of big traffic data, spatial data, environmental data and meteorological data, to support sea area monitoring and observation. The vision is to provide effective and efficient data integration, processing and analysis

technologies with the aim to deliver (a) a "Combined Real-time Operational Snapshot", and (b) a "Combined Historical Snapshot" of sea areas. The project will exploit state-of-the-art IT, and design and develop innovative IT based on in-memory database algorithms, models and methods for parallel computation, and methods for big data analytics.



ELIXIR-GR - Information Infrastructure for the Life Sciences

| | |
|------------------------|-------------------------------------------------------------------------|
| Project manager | Theodore Dalamagas (Deputy: Stelios Sartzetakis) |
| Coordinator | Biomedical Science Research Center Alexander Fleming |
| IMSI - funding | 92,600 euros |
| Programme | National Roadmap for Research Infrastructures |
| Start date | 9/2017 |
| Duration | 3 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.



MEDA 2: Moving from Big Data Management to Data Science

| | |
|------------------------|-------------------------------------------------------------------------------------------------------|
| Project manager | Theodore Dalamagas |
| Coordinator | IMSI |
| IMSI - funding | 800,000 euros |
| Programme | National Roadmap for Research Infrastructures |
| Start date | 9/2017 |
| Duration | 3 years |
| Website | http://web.imsi.athenarc.gr/projects/meda.2 |

The Information Management Systems Institute (IMSI) has already established a strong technical and scientific background on big data technologies. IMSI has recently completed MEDA, a national funded R&D project focusing on the efficient management of big data, in three important data categories: operational, scientific and social data. The project comprised a significant development action for IMSI, offering added value to existing research results and delivering novel algorithms, models and prototype tools for managing big data. MEDA.2 makes a step ahead towards data science, exploiting S&T results from MEDA. The objectives of MEDA.2 are: (a) study and experiment with data science tools and technologies, (b) setting up a data science infrastructure (tools and data collections), (c) delve into pilot cases relevant to national RIS3 smart specialization requirements, and identify related data science problems to deal with (d) design and develop innovative technologies and data services to deal with problems in (c).



APOLLONIS

Greek Infrastructure for Digital Arts, Humanities and Language Research and Innovation

| | |
|----------------------------|-----------------------|
| Project coordinator | Panos Constantopoulos |
| Coordinator | IMSI |
| IMSI - funding | 615,500 euros |

| | |
|-------------------|-----------|
| Programme | ESPA |
| Start date | 1/11/2017 |
| Duration | 3 years |
| Website | - |

APOLLONIS is the Greek Infrastructure for Digital Arts, Humanities and Language Research and Innovation, recently formed by the union of two existing ESFRI-related national research infrastructures: clarin:el, the CLARIN-related Greek network for language resources, technologies and services; and DARIAH-GR/DYAS, the DARIAH-related Greek network for digital research in the Humanities. The development of the APOLLONIS infrastructure advances the existing clarin:el and DARIAH-GR/DYAS services within a common framework that will ensure interoperability and reach to broader user communities, as well as promoting open science principles. Clarin:el will provide a permanent, stable infrastructure for accessing language resources and language processing web services, will support all kinds of language-related activities (regardless of subject), and collaborative workspace for application development environment. DARIAH-GR/DYAS will provide access to curated digital resources and services for the development, analysis and visualization of data, best practice guidelines, and dissemination and training activities on the use of digital methods and tools in the Humanities. A Digital Humanities Observatory will monitor the penetration of digital practices in the Humanities.



eLib GGDE - eLib of Independent Authority for Public Revenue

| | |
|----------------------------|-----------------------------------------------------------------------------------|
| Project coordinator | George Papastefanatos |
| Coordinator | Remaco S.A. |
| IMSI - funding | 51,200 euros |
| Programme | NSRF - Public Tender |
| Start date | 30/7/2015 |
| Duration | 4 years (ongoing) |
| 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 coordinator | Yannis Stavrakas |
| Coordinator | IMSI |
| Programme | Internal project |
| Start date | 12/2015 |
| Duration | 4 years and 1 month |

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.



Digital Curation Services

| | |
|----------------------------|-----------------------|
| Project coordinator | Panos Constantopoulos |
| Coordinator | IMSI |
| IMSI - 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 IMSI 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.



Europeana DSI-4

| | |
|----------------------------|----------------------------------------------------------------------------------------------|
| Project coordinator | Dimitris Gavrilis |
| Coordinator | STICHTING EUROPEANA(EF) |
| IMSI - funding | 97,372 euros |
| Programme | Service Contract-“Deployment and Maintenance of Europeana DSI Core Services-SMART-2017/1136” |
| Start date | 1/9/2018 |
| Duration | 24 months |

Website <https://pro.europeana.eu/project/europeana-dsi-4>

The Europeana DSI-4 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, DSI-2 and DSI-3 actions under CEF. It is the fourth 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. IMSI participates in DSI-4 both as a technical partner with the MORE platform (supporting the CARARE network as part of DSI) and also through Europeana Research.

Industry – Funded R&D Projects



Design and Development of big data solution and methods for stream analytics on network data

| | |
|----------------------------|-----------------------|
| Project coordinator | George Papastefanatos |
| Coordinator | Intracom Telecom S.A. |
| IMSI - funding | 269,080 euros |
| Programme | Contract |
| Start date | 16/10/2017 |
| Duration | 21 months (ongoing) |
| Website | - |

A new collaboration between **IMSI, Intracom Telecom and Ericsson** started in *October 2017*. IMSI has been contracted to design and develop a big data solution and methods for stream analytics on network data coming from IoT devices. The solution is based on well-established big data and cloud platforms, i.e., Microsoft Azure, Cloudera Ecosystem and SPARK streaming and enables the collection of vast amount of network measurement streams from connected devices, their scalable processing, analysis and storage and the visualization of several KPIs.

Publications

Journal Publications

- Danae Pla Karidi, Yannis Stavrakas, and Yannis Vassiliou. **Tweet and Followee Personalized Recommendations Based on Knowledge Graphs**. Journal of Ambient Intelligence and Humanized Computing, 9(6): 2035-2049, 2018.
- Nikos Giatrakos, Antonios Deligiannakis, Minos Garofalakis, Daniel Keren, Vasilis Samoladas. **Scalable Approximate Query Tracking over Highly Distributed Data Streams with Tunable Accuracy Guarantees**. Information Systems, Volume 76, Pages 59-87, July 2018.
- Nikos Giatrakos, Antonios Deligiannakis, Minos Garofalakis, Yannis Kotidis. **Omnibus Outlier Detection in Sensor Networks using Windowed Locality Sensitive Hashing**. Future Generation Computer Systems, 2018, accepted for publication.
- E. Anagnostopoulos, I.Z. Emiris, and I. Psarros. **Randomized embeddings with slack, and high-dimensional approximate nearest neighbor**. ACM Trans. on Algorithms, 14:18:1–18:21, 2018. <https://doi.org/10.1145/3178540>
- D. Karagkouni, M. D. Paraskevopoulou, S. Chatzopoulos, I. S. Vlachos, S. Tastsoglou, I. Kanellos, D. Papadimitriou, I. Kavakiotis, S. Maniou, G. Skoufos, T. Vergoulis, T. Dalamagas, A. G. Hatzigeorgiou: **DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions**. Nucleic Acids Research 46(Database-Issue): D239-D245.
- M Koniaris, G Papastefanatos, I Anagnostopoulos. **Solon: A Holistic Approach for Modelling, Managing and Mining Legal Sources**. Algorithms Journal, 11 (12), 196, doi:10.3390/a11120196 (2018).
- M. Meimaris, G. Papastefanatos, P. Vassiliadis, I. Anagnostopoulos. **Computational Methods and Optimizations for Containment and Complementarity in Web Data Cubes**, Inf. Syst. 75: 56-74 (2018).
- Alexandros Belesiotis, Dimitrios Skoutas, Christodoulos Efstathiades, Vassilis Kaffes, Dieter Pfoser: **Spatio-textual user matching and clustering based on set similarity joins**. VLDB J. 27(3): 297-320 (2018)
- Alexandros Belesiotis, George Papadakis, Dimitrios Skoutas: **Analyzing and Predicting Spatial Crime Distribution Using Crowdsourced and Open Data**. ACM Trans. Spatial Algorithms and Systems 3(4): 12:1-12:31 (2018)
- Laužikas, R., C. Dallas, S. Thomas, I. Kelpšienė, I. Huvila, P. Luengo, H. Nobre, M. Toumpouri, V. Vaitkevičius. **Archaeological Knowledge Production and Global Communities: Boundaries and Structure of the Field**. Open Archaeology, Vol. 4 (1), pp. 350–364, 2018, ISSN: 2300-6560

International Conference / Workshop Publications

- Evangelos Bartzos, Ioannis Z. Emiris, Jan Legerský, Elias Tsigaridas. **On the Maximal Number of Real Embeddings of Spatial Minimally Rigid Graphs**. In Proceedings of the 2018 ACM International Symposium on Symbolic and Algebraic Computation (ISSAC '18), Carlos Arreche (Ed.). ACM, New York, NY, USA, 55-62. <https://doi.org/10.1145/3208976.3208994>
- Pantelis Chronis, Giorgos Giannopoulos, Spiros Athanasiou, Spiros Skiadopoulos: **Unsupervised Disaggregation of Low Granularity Resource Consumption Time Series**. PAKDD 2018: 558-570.
- N. Bikakis, S. Maroulis, G. Papastefanatos, P. Vassiliadis: RawVis: Visual Exploration over Raw Data. In 22nd European Conference on Advances in Databases and Information Systems (ADBIS 2018), September 2-5, Budapest, Hungary.
- Manghi, Paolo, Manunta, Michele, Baglioni, Miriam, Bardi, Alessia, Casu, Francesco, De Luca, Claudio, Kokogiannaki, Argiro. **Enabling Open Science publishing for Research Infrastructures via OpenAIRE: The EPOS use-case**. 2018, <http://doi.org/10.5281/zenodo.1314672>
- Dimitrios Skoutas, Dimitris Sacharidis, Kostas Patroumpas. **Efficient progressive and diversified top-k best region search**. In Proc. of the 26th ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL), pp. 299-308, Seattle, WA, USA, November 06-09, 2018.
- Georgios Chatzigeorgakidis, Kostas Patroumpas, Dimitrios Skoutas, Spiros Athanasiou, Spiros Skiadopoulos. **Scalable hybrid similarity join over geolocated time series**. In Proc. of the 26th ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL), pp. 119-128, Seattle, WA, USA, November 06-09, 2018.
- Vassilis Kaffes, Alexandros Belesiotis, Dimitrios Skoutas, Spiros Skiadopoulos. **Finding shortest keyword covering routes in road networks**. In Proc. of the 30th International Conference on Scientific and Statistical Database Management (SSDBM), pp. 12:1-12:12, Bozen-Bolzano, Italy, July 09-11, 2018.
- Dimitris Sacharidis, Paras Mehta, Dimitrios Skoutas, Kostas Patroumpas, Agnès Voisard. **Selecting representative and diverse spatio-textual posts over sliding windows**. In Proc. of the 30th International Conference on Scientific and Statistical Database Management (SSDBM), pp. 17:1-17:12, Bozen-Bolzano, Italy, July 09-11, 2018.
- Georgios Chatzigeorgakidis, Dimitrios Skoutas, Kostas Patroumpas, Spiros Athanasiou, Spiros Skiadopoulos. **Map-Based Visual Exploration of Geolocated Time Series**. In Proc. of the EDBT Workshop on Big Data Visual Exploration and Analytics (BigVis), pp. 92-99, Vienna, Austria, March 26, 2018.
- Jeliaskova N, Ritchie P, Rashid S, Basei G, Zabeo A, Tsiliki G, Hristozov D. **Data curation, model development and integration: the GRACIOUS platform**. Scientific Workshop on grouping nanomaterials, 12-13 September 2018, Paris, France.

- Basei G, Hristozov D, Lamon L, Zabeo A, Jeliaskova N, Tsiliki G, Marcomini A, Torsello A. **Making use of available and emerging data to predict the hazards of engineered nanomaterials by means of *in silico* tools: a critical review**. 3rd Nanosafety Forum for Young Scientists, 10-11 September 2018, Valetta, Malta.
- Pertsas, V., P. Constantopoulos, I. Androutsopoulos. **Ontology-Driven Extraction of Research Processes**. Proc. ISWC 2018, 17th International Semantic Web Conference, Monterey, CA, USA, October 8–12, 2018, pp. 162-178. Springer, 2018, ISBN: 978-3-030-00670-9.
- Pertsas, V., and P. Constantopoulos. **Ontology-Driven Extraction of Research Processes**. Digital Libraries for Open Knowledge, Proc.TPDL 2018., pp. 241-253. Springer, 2018, ISBN: 978-3-030-00065-3.
- Dimitra Papadimitriou, Yannis Velegarakis, Georgia Koutrika. **Modeling and Exploiting Goal and Action Associations for Recommendations**. EDBT 2018: 409-420
- Georgia Koutrika. **Recent Advances in Recommender Systems: Matrices, Bandits, and Blenders**. EDBT 2018: 517-519
- Dimitra Papadimitriou, Georgia Koutrika, Yannis Velegarakis, John Mylopoulos. **Finding Related Forum Posts through Content Similarity over Intention-Based Segmentation** (Extended Abstract). ICDE 2018: 1805-1806
- Marialena Kyriakidi, Georgia Koutrika, Yannis E. Ioannidis. **Recommendations for Explorations based on Graphs**. ExploreDB@SIGMOD/PODS 2018: 1:1-1:4
- Georgia Koutrika. **Modern Recommender Systems: from Computing Matrices to Thinking with Neurons**. SIGMOD Conference 2018: 1651-1654

National Conference / Workshop Publications

- T. Vergoulis, I. Kanellos, S. Chatzopoulos, C. Tryfonopoulos, T. Dalamagas, Y. Vassiliou: **Pub Finder: Assisting the discovery of qualitative research**. In 16th Hellenic Data Management Symposium (HDMS'18), Larnaca, Cyprus, July 2018.

Other Publications / Technical Reports

- T. Vergoulis, K. Zagganas, A. Zissimos, S. Sartzetakis, T. Dalamagas: **Data-intensive computing for the ELIXIR-GR community**. ELIXIR All Hands 2018, Berlin, Germany, June 2018.
- T. Vergoulis, I. Kanellos, S. Chatzopoulos, N. Manola, T. Dalamagas: **Open Science Services for the ELIXIR-GR community**. ELIXIR All Hands 2018, Berlin, Germany, June 2018.
- Manola, Natalia, Papadopoulou, Elli, Robertson, Dale, Tsiavos, Prodromos, Cavalli, Valentino, Thorat, Rahul, & Thorley, Mark. **EOSCpilot Open Science and Open Scholarship: Supporting Document to D3.3 Draft Policy Recommendations**. Zenodo 2018. <https://zenodo.org/record/2176076#.XYjfbigzY2w>

- Michael Ball, Margreet Bloemers, David Carr, Valentino Cavalli, Maria Haglund, Vasso Kalaitzi, ... Karen Vandavelde. **A Vision for Open Science**. Zenodo, 2018. <http://doi.org/10.5281/zenodo.1491303>

Dissemination Activities

Invited / Keynote Talks

IMSI members participated in the following invited / keynote talks:

- Round table: "**Digital Humanities: Is Music part of them?**", 2nd conference of Music Libraries and Archives, Athens Conservatoire, 27/4/2018. Moderator: Agiatis Benardou.
- Presentation of the Apollonis infrastructure at the Kick off meeting of "**Culture, Digital Media and Entrepreneurship: From Digitization to Gamification**", 18/05/2018. Presenter: Agiatis Benardou.
- Presentation at the "**Digital History**" one-day event, Historical Archive of the University of Athens, 14/06/2018. Presenter: Agiatis Benardou.
- Presentation on "**Connecting archaeology and architecture data: updating the CARARE metadata schema**", CARARE Workshop "Archaeology and Architecture in Europeana", University of Lund, Sweden, 20/06/2018. Presenter: Dimitris Gavrilis (with Kate Fernie).
- Presentation on: "**Immersive technologies as a method for showcasing historically loaded sites of cultural heritage**", Workshop on the Promotion and Management of Monuments in Greece, 26/10/2018. Presenter: Agiatis Benardou.

Scientific Community Service

IMSI members have served in the **Program Committee** of more than 18 International Conferences and Workshops in 2018, including well-known Conferences like ICDE, SIGMOD, TPD, SIGSPATIAL, WWW, ISWC, DATA, LOD and more.

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

- **1st International Workshop on Big Data Visual Exploration and Analytics (BigVis2018)**. In conjunction with EDBT/ICDT 2018, Vienna, Austria, May 2018. George Papastefanatos.
- **ACM SIGMOD International Conference on Management of Data (SIGMOD) 2018, USA**. Georgia Koutrika, Demo PC co-chair.

IMSI members Manolis Terrovitis and Minos Garofalakis became members of the Sectoral Scientific Board on Mathematics and Computer Science (TES). TES is an advisory board for the formulation of the research strategy in Greece.

Other Dissemination Activities

- **Amnesia Webinar.** Webinar in the context of the project OpenAIRE, 24/4/2018.
- **Amensia Webinar.** Research Data Canada (RDC), 27/9/2018.

- **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, the 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 which 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 OAIS-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**

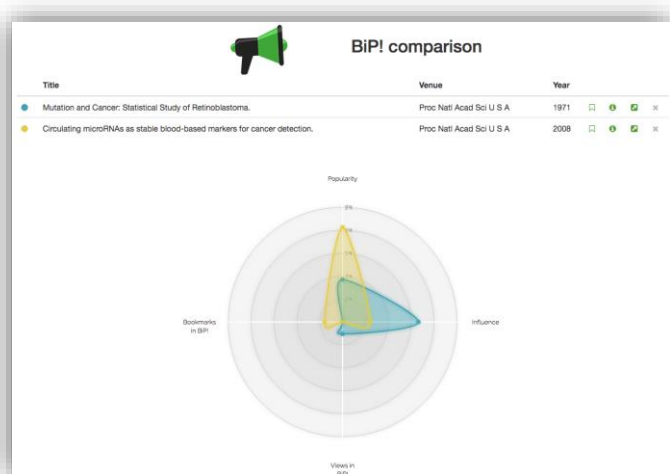
DAIAD is the first *integrated residential demand management* system for water. It applies Big Data and Machine Learning technologies to leverage smart water meter data, engaging and informing consumers to induce sustainable changes in consumption behaviour, as well as providing novel large-scale analytics 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 behaviour. Water utilities have access to several analysis services (segmentation, clustering, forecasting) enabling them to understand consumption behaviour 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 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 high-impact publications in the field of life sciences. This tool supports ranking and comparing of scientific articles based on different aspects of their impact



in their discipline, like their popularity (i.e., the current attention they receive) or influence (i.e., their long-term impact in the discipline). Furthermore, the tool provides useful features like intuitive infographics for each article and a mechanism of bookmarks.

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

- **BUFET**

BUFET is an open-source software under the GPL v.3 licence, designed to speed up Bleazard's unbiased miRNA enrichment analysis algorithm. BUFET generates an empirical distribution of genes targeted by miRNA and calculates p-values for related biological processes. Benjamini-Hochberg FDR correction produces a '*' or '**' for significance at 0.05 FDR and 0.01 FDR respectively.

<https://github.com/diwis/BUFET>

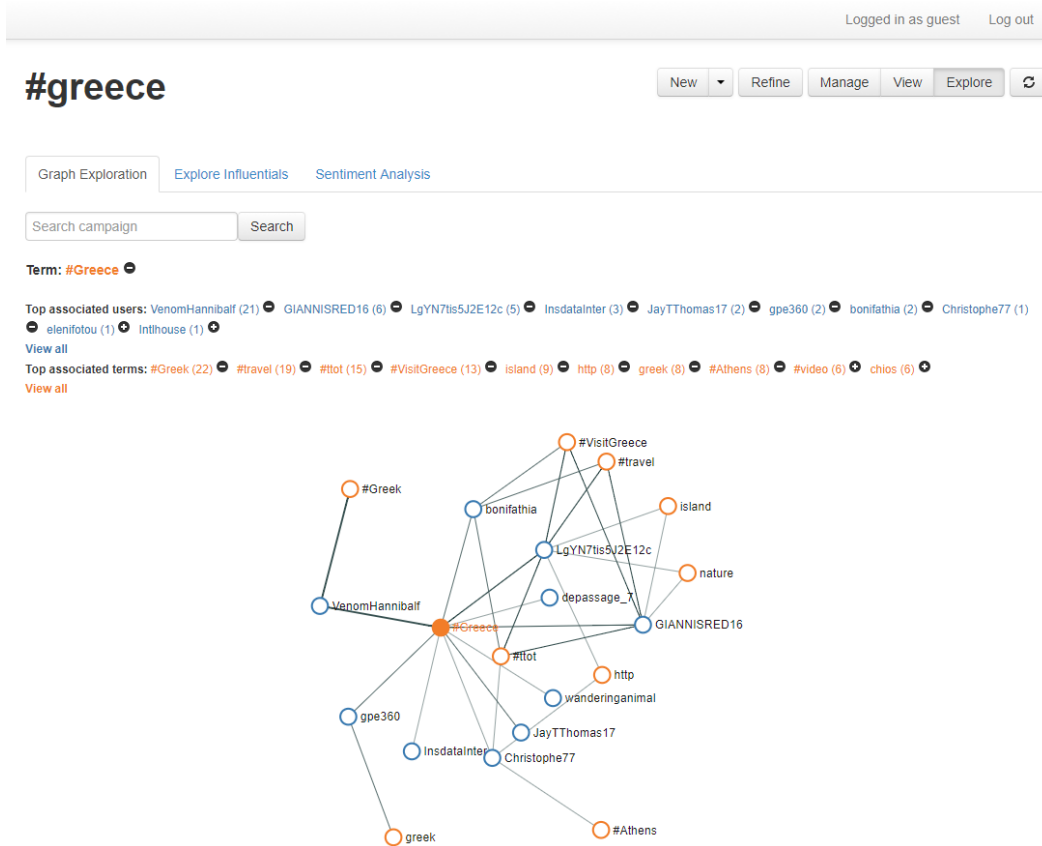
- **PaperRanking**

PaperRanking is an open source library containing implementations of paper ranking methods that have been proposed in the literature. Our implementations utilise a suite of MapReduce scripts and can be used either on a single machine, or a Hadoop cluster. All codes were developed in the context of a paper ranking survey that aimed to evaluate each method's strengths and weaknesses.

<https://github.com/diwis/PaperRanking>

- **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

The screenshot displays the 'Tract 110100' interface. On the left, a sidebar contains a list of filters: 'Points of Interest', 'Health - Birth rate', 'Social - Checkins', 'Social - Tweets', 'Socioeconomic - Crime', and 'Socioeconomic - Unemployment'. Below these filters are two barplots: 'Barplot' and 'Barplot 2'. The 'Barplot' shows a distribution of values with a mean of 12.62 and a selected tract of 9.00. The 'Barplot 2' shows a distribution of values with a mean of -3.62. The main area is a map of Chicago with tracts colored according to the selected filter. On the right, a panel shows the 'Points of Interest' filter selected, with a list of tracts and their percentages: 106 tracts (13.23%), 124 tracts (15.48%), 110 tracts (13.73%), 122 tracts (15.23%), 111 tracts (13.85%), 117 tracts (14.61%), and 111 tracts (13.85%). Below this is a histogram and a value range filter set to 5.99 to 17.18. At the bottom, a filter condition is set to 'No'.

- linkzoo

www.linkzoo.gr

- **Socioscope**

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 uses 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**

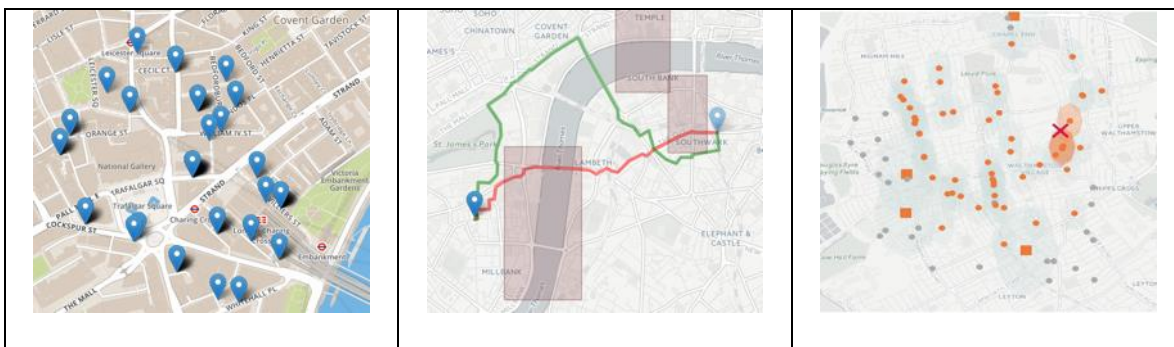
FAGI is a tool that allows the fusion of geospatial Linked Data. It is designed to retrieve data through SPARQL endpoints, and implements a wide range of fusion actions both on spatial properties of the entities and on non-spatial metadata. These include moving, rotating, scaling and aligning the geometries 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 functionalities: (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.imsi.athenarc.gr/>



Education

PhD / MSc / Diploma Thesis Co-supervision

IMSI members actively co-supervise several undergraduate and graduate university students, who often conduct their work in IMSI premises. As a result, in 2018 several MSc and Diploma thesis have been co-supervised by IMSI members, who often serve as members in the respective examination committees.

IMSI members also co-supervise PhD students. The following PhD dissertations have been completed in 2018:

- Marios Meimaris. **Managing, Querying and Analyzing Big Data on the Web.** University of Thessaly. Collaborating researchers: [George Papastefanatos](#).

The following PhD students collaborated closely with IMSI members in their research during 2018:

- Giorgos Alexiou. Topic: **Entity disambiguation and data interlinking.** 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 the Peloponnese. Collaborating researcher: [Dimitris Skoutas](#).
- Giorgos Chatziogeorgakidis. Topic: **Scalable Indexing and Exploration of Big Time Series Data.** Joint supervision with the University of Peloponnese. Collaborating researcher: [Dimitris Skoutas](#), [Spiros Athanasiou](#).
- 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](#).
- Konstantinos Theocharidis. Joint supervision with the University of the Peloponnese. Collaborating researcher: [Manolis Terrovitis](#).
- Ilias Kanellos. Joint supervision with the National Technical University of Athens. Collaborating researcher: Theodore Dalamagas, Thanasis Vergoulis
- Konstantinos Zagganas. Joint supervision with the University of the Peloponnese. Collaborating researcher: Theodore Dalamagas, Thanasis Vergoulis
- Serafeim Chatzopoulos. Joint supervision with the University of the Peloponnese. Collaborating researcher: Theodore Dalamagas, Thanasis Vergoulis

- Marialena Kyriakidi. Topic: **Recommendations over Graph Databases**. Joint supervision with the University of Athens. Collaborating researcher: Georgia Koutrika.

Other Educational Activities

Other educational activities involving IMIS members include the following.

- **Organization of the 2nd ACM Europe Summer School in Data Science** 12–18 July 2018, Athens, Greece. The summer school invites researchers around the world (i.e., young faculty, postdocs, PhD candidates, young engineers from industry and in some special cases MSc students or even senior undergraduate students) to attend a series of interesting and constructive lectures by distinguished professors. Lectures focus on several forms of analytics, machine learning and data mining techniques, big data systems, and the ethics of data science.
- Giorgos Giannopoulos taught the **Big Data Management** course of “Space Science Technologies and Applications” MSc program of the University of the Peloponnese.
- George Papastefanatos was Adjunct Lecturer in Spring 2018 for the course “Technologies for the Management of Cultural Information” (Undergraduate course) in University of Aegean, Department of Cultural Technology and Communication).
- Thanasis Vergoulis was a teaching associate at the University of the Peloponnese (undergraduate course “Artificial Intelligence”) and at the Technical Education Institute of Peloponnese (undergraduate course “System Programming”).
- Georgia Koutrika gave a 3-day Advanced Seminar on “**Machine Learning Methods for Recommendations**” at Harvard University, in Spring 2018.

Facts and Figures

Financial report

In 2018, IMSI continued its participation in EC and national funded research and development projects. The key economic indicators regarding the expenses and revenues in 2018 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, about 62%, comes from EC funded projects. It is important to note that the revenues from the activities of IMSI (EC projects, national funded projects, Product and Service Sales) are more than 15 times the public expenditure received by IMSI.

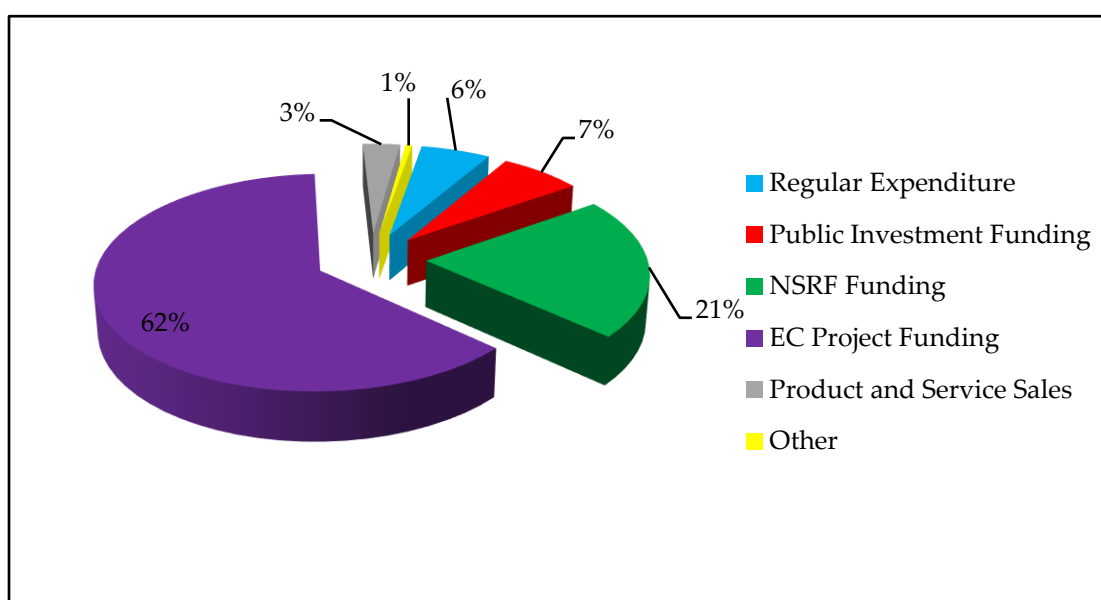


Figure 1. Distribution of revenues in 2018

Table 1. Expenses and Revenues for 2018

| Expenses 2018 | |
|-------------------|-------------------------|
| Travel Expenses | 192.472,62 |
| Operational Costs | 49.000,41 |
| Equipment | 91.024,82 |
| Other Expenses | 600.682,99 ¹ |

¹ This number included the transfer of 354.425,88€ 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)"

| | |
|-----------------------------------------------------|--------------|
| Personnel fees and payments to third parties | 2.544.504,65 |
| Total | 3.477.685,49 |
| Revenues 2018 | |
| Regular Expenditure | 324.794,26 |
| Public Investment Funding | 383.261,82 |
| NSRF Funding | 1.208.807,34 |
| EC Project Funding | 3.536.570,82 |
| Product and Service Sales | 173.080,02 |
| Other | 31.198,95 |
| Total | 5.657.713,21 |

Table 2 shows the revenues of IMSI since 2014 while a comparison of the revenues in the years 2014 - 2018 is illustrated in Figure 2. We can see that the revenues coming from participation in European projects reached 3.5M euros in 2018, an increase of almost 85% from 2014. As expected, a part of the NSRF Funding within the Partnership Agreement 2014-2020 was paid off in 2018, with the revenues from the participation of IMSI in national funded projects reaching 1,2M euros.

Table 2. Revenues from 2014 to 2018

| | Revenues | | | | |
|----------------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| | 2014 | 2015 | 2016 | 2017 | 2018 |
| Regular Expenditure | 76.226,10 | 82.307,98 | 164.154,00 | 290.802,02 | 324.794,26 |
| Public Investment Funding | 454.421,97 | 185.186,15 | - | 328.701,50 | 383.261,82 |
| NSRF Funding | 790.948,38 | 2.055.402,11 | 102.597,68 | 0,00 | 1.208.807,34 |
| EC Project Funding | 1.916.199,47 | 2.335.227,81 | 3.804.217,79 | 2.690.098,90 | 3.536.570,82 |
| Product and Service Sales | 291.943,42 | 368.338,97 | - | 6.384,60 | 173.080,02 |
| Other | 85.508,66 | 27.506,22 | 40.983,75 | 26.797,83 | 31.198,95 |
| Total | 3.615.248,00 | 5.053.969,24 | 4.111.953,22 | 3.342.784,85 | 5.657.713,21 |

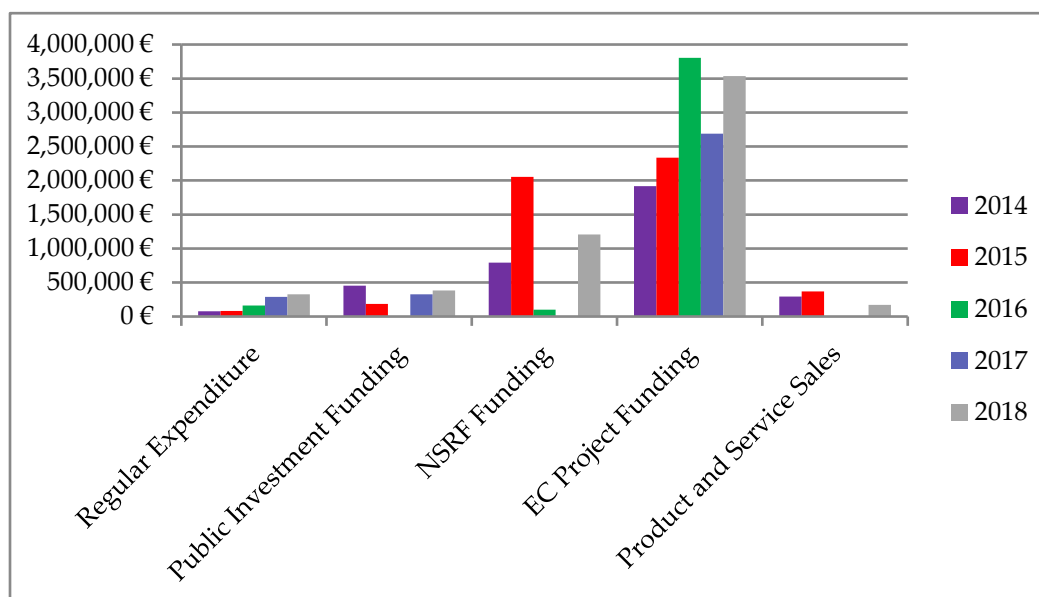


Figure 2. Comparison of revenues 2014 - 2018

Staff

Director & Researchers

- Garofalakis Minos, Professor, Technical University of Crete, Director
- Koutrika Georgia, Researcher A
- Sartzetakis Stylianos, Researcher A
- Simitsis Alkiviadis, Researcher A
- Dalamagas Theodore, Researcher B
- Stavrakas Yannis, Researcher B
- Skoutas Dimitris, Researcher C
- Terrovitis Manolis, Researcher C

Associate Researchers / Professors

- Androutopoulos Ion, Associate Professor, AUEB
- Chatziantoniou Damianos, Assistant Professor, AUEB
- Chatzigeorgiou Artemis, Professor, University of Thessaly
- Chatzopoulos Michael, Emeritus Professor, University of Athens
- Christofidis Vasilios, Professor, University of Crete
- Christopoulos Dionisios, Professor, Technical University of Crete
- Constantopoulos Panos, Professor, AUEB
- Dagklis Ioannis, Professor, University of Athens
- Dallas Costis, Assistant Professor, Toronto University
- Deligiannakis Antonis, Assistant Professor, Technical University of Crete
- Delis Alexios, Professor, University of Athens
- Dokoumetzidis Aristidis, Assistant Professor, University of Athens
- Emiris Ioannis, Professor, University of Athens
- Giaglis Georgios, Professor, AUEB

- Giaoutzi Maria, Professor, NTUA
- Katsiri Eli, Assistant Professor, Democritus University of Thrace
- Kavouras Marinos, Professor, NTUA
- Kavouras Pavlos, Professor, University of Athens
- Kokosis Antonis, Professor, NTUA
- Kontogiannis Constantinos, Assistant Professor, NTUA
- Kotidis Yannis, Associate Professor, AUEB
- Koubarakis Emmanouil, Professor, University of Athens
- Kountouri Phoebe, Assistant Professor, AUEB
- Koziris Nectarios, Professor, NTUA
- Mamoulis Nikos, Assistant Professor, University of Ioannina
- Papatheodorou Christos, Professor, Ionian University
- Patsakis Konstantinos, Assistant Professor, University of Piraeus
- Pittis Nikitas, Professor, University of Piraeus
- Prezerakos Georgios, Professor, Technological Education Institute of Piraeus
- Roditi Evgenia, Assistant Professor, University of Athens
- Samoladas Vasilis, Associate Professor, Technical University of Crete
- Sellis Timos, Professor, Swinburne University of Technology
- Sivridis Dimitrios, Professor, University of Athens
- Skiadopoulos Spiros, Assistant Professor, University of Peloponnese
- Stavdas Alexandros, Professor, University of Peloponnese
- Theodoratos Dimitrios, Associate Professor, NJIT USA
- Theodoridis Yannis, Professor, University of Piraeus
- Vasileiadis Panos, Assistant Professor, University of Ioannina
- Vassalos Vassilis, Associate Professor, AUEB
- Vassilopoulos Achilleas, Assistant Professor, University of Ioannina
- Vazirgiannis Michalis, Professor, AUEB
- Zarras Apostolos, Assistant Professor, University of Ioannina

Postdoctoral Researchers / Scientific Collaborators

- Arkoumanis Constantinos, Elec. & Comp. Eng. NTUA, PhD NTUA
- Benardou Agiatis, BA, PhD in Ancient History, King's College London
- Dimitropoulos Charalampos, Elec. & Elec. Eng. City University London, PhD University of London
- Gavrilis Dimitris, Elec. Comp. & Eng. University of Patras, PhD University of Patras
- Giannopoulos Giorgos, Elec. & Comp. Eng. NTUA, PhD NTUA
- Giatrakos Nikolaos, BSc in Informatics, University of Piraeus, MSc in Information Systems, AUEB, PhD, Department of Informatics, University of Piraeus
- Karagiannis Fotis, Elec. & Comp. Eng. NTUA, PhD NTUA
- Konaxis Christos, Dept. of Mathematics UoA, PhD . Dept Informatics & Telecoms, UoA
- Papastefanatos George, Elec. & Comp. Eng. NTUA, PhD NTUA
- Pierrakos Dimitris, Dept. of Physics UoA, PhD UoA
- Tsani Stella, Dept. Of Economics UoA, PhD University of Reading
- Tsiavos Prodromos, Dep. of Law UoA, Master of Laws (LL.M), Univ. of London, MSc Analysis Design & Management of Information Systems, LSE & Political Science, PhD LSE
- Tsiliki Georgia, BSc in Statistics, University of Piraeus, PhD in Statistics, Imperial College London
- Tzotsos Aggelos, BA in Rural and Surveying Engineering, NTUA, PhD NTUA
- Vergoulis Thanasis, Elec. Comp. & Eng. University of Patras, PhD NTUA

- Zacharia Eleni, Dept. of Inform. and Telecom. UoA, PhD UoA

Administrative and Technical Personnel

- Apostolopoulou Nelly, Mech. Eng. NTUA, Msc NTUA, Head of Administration
- Georgantas Panagiotis, Elec. & Comp. Eng. NTUA, Technician
- Paradissi Eva, Secretarial Support

Research Staff

- Aidonopoulos Orfeas, Comp. & Inform. Eng. Univ. of Patras, Msc Dept. of Inform. & Telecom, UoA
- Alexakis Michalis, Elec. & Comp. Eng. NTUA
- Alexiou George, University of Aegean, Msc AUTH
- Athanasiou Spyros, Elec. & Comp. Eng. NTUA
- Baltzi Sofia, Dept. of Inform. and Telecom. UoA
- Bartzos Vangelis, Dep. of MBG, DUTH, Dep. of Math., Univ. of Patras, Msc Ecole normale superieure, Lyon, Msc Universite 1, Lyon
- Chatzilenas Christos, Department of Informatics and Telematics, Harokopio University, Msc, Department of Informatics, AUEB
- Chatzopoulos Serafeim, BA, Msc, Dept. of Inform. and Telecom. UoA
- Chronis Pantelis, Comp. & Inform. Eng. University of Patras
- Costagiolas Nikolaos – Stefanos, , Dept. of Inform. and Telecom. UoA, Msc in Data Science, AUEB
- Diplas Ioannis, Dept. of Inform. and Telecom. UoA
- Emmanouilidi Sofia, Deree, Business Administration (Accounting and Finance)
- Foufoulas Ioannis Dept. of Inform. and Telecom. UoA, Msc UoA
- Galouni Constantina, Dept. of Inform. and Telecom. UoA
- Gentimis Thanassis, Elec. & Comp. Eng. NTUA, Msc UoA
- Giannakopoulos Anastasios, Dept. of Inform. and Telecom. UoA
- Giannopoulos Athanasios, BA, MSc, Dept. of Inform. and Telecom. UoA
- Iatropoulou Katerina, BA, Msc, Dept. of Inform. and Telecom. UoA
- Ilvanidou Maria, Faculty of History and Archaeology, UoA, Msc, University of Crete
- Kaffes Vassilis, Elec. & Comp. Eng. NTUA
- Kakalettris Georgios, BA, Msc, Dept. of Inform. and Telecom. UoA
- Kakalettris Konstantinos, Dept. of Electrical Engineers, TEI Piraeus, Msc, UoA
- Kalampokis Panagiotis, BA, Computer Science, University of Ioannina, Msc in Mobile and Web Computing, IHU
- Kanakakis Panagiotis, Dept. of Inform. and Telecom. UoA
- Kanellos Ilias, Elec. & Comp. Eng. NTUA
- Karagiannakis Nikos, Comp. & Inform. Eng. University of Patras
- Karamanolis Georgios, Dept. of Inform, AUEB
- Katrakazi Elissavet Olga, BA AUEB, Msc in Applied Environmental Economics, University of Kent
- Kokogiannaki Argiro, Dept. of Inform. and Telecom. UoA
- Koltsida Panayota, BA, Msc, Dept. of Inform. and Telecom. UoA
- Kostoulas Georgios, Dipl., MSc, Electr. & Comp. Eng, Technical University of Crete
- Koukouli Mirto, , Faculty of History and Archaeology, UoA
- Koulalis Antonios, BA Dept. of Inform. and Telecom. UoA
- Kouvaras Ioannis, Elec. & Comp. Eng. NTUA
- Kravvaritis Constantinos, Dept. of Inform. and Telecom. UoA, Msc, Dept. of Inform, AUEB

- Laroche Clement, Fundamental Mathematics Universite Paul Sabatier, Msc Universite Paul Sabatier, Mathematics for Research Universite Paul Sabatier
- Lebesis Antonis, BA, Msc, Dept. of Inform. and Telecom. UoA
- Leonidakis Nikiforos, Dept. of Inform. and Telecom. UoA
- Levantis Eleftherios, Dep. of Law, UoA, Msc Environmental Management EAEME
- Livisianos Tsabikos, Dept. of Inform. and Telecom. UoA
- Maili Evdokia, BA in Mathematics, University of Patras, Msc, Dept. of Inform. and Telecom. UoA
- Manola Natalia, Dept. of Physics, UOA
- Maroulis Stavros, Elec. & Comp. Eng. NTUA
- Martziou Stefania-Varvara, BA, Msc, Dept. of Inform. and Telecom. UoA
- Mavrogiannakis Nikos, BA, Msc in Physics, University of Manchester
- Meimaris Marios, Elec. & Comp. Eng. NTUA, Msc NTUA, University of Edinburgh
- Metaxas Omiros, BA, Msc, Dept. of Inform. and Telecom. UoA
- Mougiakos Ioannis, BA in Rural and Surveying Engineering, NTUA, Msc, University of Nottingham
- Mouris Dimitrios, BA Dept. of Inform. and Telecom. UoA
- Mpikas Georgios, Dept. of Electrical Engineers, TEI Piraeus
- Mpiliris Grigorios, BA in Rural and Surveying Engineering, NTUA, Msc, University College of London
- Nakos Charalampos, Elec. & Comp. Eng. NTUA
- Nasikas Christos, BSc in Informatics, University of Piraeus
- Nomikos Vangelis, Dept. of Inform. and Telecom. UoA
- Oikonomou Leonidas, Architecture, School of Architecture NTUA, MSc NTUA
- Papadaki Aikaterini, Dept. of Geography, University of the Aegean, Msc, Rural and Surveying Engineering, NTUA
- Papadopoulou Elli, BA Librarian & Informatics Scientist, Alexander Technological Educational Institute of Thessaloniki
- Papanikos Georgios, BA, MSc, Dept. of Inform. and Telecom. UoA
- Patroumpas Konstantinos, Comp. & Inform. Eng. University of Patras, Msc NTUA
- Pavlidou Androniki, BA Mathematics, AUTH, MSc on Science in Innovation Management, Univ. of Utrecht
- Petra Eleni, Comp. & Inform. Eng. University of Patras, BA University of Manchester
- Pla-Karidi Danae, Elec. & Comp. Eng. NTUA
- Pristouris Costis, Elec. & Comp. Eng. NTUA, MSc in Geoinformatics, NTUA
- Psallidas Michail, Comp. & Inform. Eng. University of Patras
- Sifakaki Ilektra, BA in Mathematics, UoA, Msc, Department of Informatics and Telematics, Harokopio University
- Soulis Athanasios, Dept. of Inform. and Telecom. UoA
- Spiliopoulou Panorea, School of Law, UoA, Master of Laws, University of Hamburg/Erasmus University/University of Vienna, Master of Laws, Bordeaux V-UoA
- Spiropoulos Dimitrios, Dept. of Inform. and Telecom. UoA
- Spirou Konstantinos, Dept. of Inform. and Telecom. UoA
- Theocharidis Constantinos, Dep. of Electrical and Computer Engineering, University of Thessaly
- Theodorakopoulos Efthimios, Dept. of Inform. and Telecom. UoA
- Tsitsigkos Dimitris, Dept. of Inform. and Telecom. UoA
- Tziotzios Diamantis, Elec. & Comp. Eng. DUTH
- Vagianou Maria, Dept. of Inform. University of Piraeus, Msc UoA

- Vagias Christos, BA in Rural and Surveying Engineering, NTUA, Msc, University of Nottingham
- Vos Aggeliki, BA, Department of Library Science & Information Systems, TEI of Athens, Msc in Information Systems, Linnaeus University, Sweden
- Xiros Panagiotis, Elec. & Comp. Eng. NTUA
- Zagganas Constantinos, Elec. & Comp. Eng. NTUA



Contact

| |
|--------------------------------------------------------|
| Information Management Systems Institute (IMSI) |
|--------------------------------------------------------|

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-imsi@athenarc.gr

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

<https://www.imsi.athenarc.gr/en>

