Centre of Excellence in Data Science

Concept and Objectives

Modern scientific research is largely data-driven. Recent technology advances made it possible to collect increasingly large data sets with a high rate of acquisition, opening a host of new challenges and opportunities in diverse application areas. As the volume of data is increasing rapidly, it is now expected that by 2020 there will be more than 16 zettabytes of useful data (16 Trillion GB), implying a growth of 236% per year from 2013 to 2020, according to a study “Worldwide Big Data Technology and Services 2014-2018” by IDC, a US market intelligence provider. Data has truly become an asset - one that can create significant competitive advantages and drive innovation, increase competitiveness, and create tangible social impact.

Data Science addresses the problem of knowledge extraction from structured and unstructured data, including very large data sets often referred to as Big Data. Due to government and industry pressing need for Data Science expertise, the market demand for data scientists has soared in recent years. Currently, there are several high education institutions and research institutes in Croatia actively engaged in various theoretical and applicative segments of Data Science. Despite this, the existing research efforts are not at all coordinated and there is a serious lack of communication among the research groups - a state of affairs also attributable to heterogeneous and multidisciplinary nature of the Data Science field. As a result, there is a lack of critical mass to advance the scientific state of the art and to provide support to national industry and government.

The proposed national Centre of Excellence in Data Science (CEDS) will bring together the best research groups in Croatia involved in research on foundations and applications of Data Science to advance the state of the art in theory, technology, and systems. It is expected that the complementary nature of the research team and the focus on common challenges of interest will generate exchange of knowledge, experience, and ideas between researchers from various research fields, resulting in synergistic effects that reach far beyond individual groups’ achievements and capacities. Forming such an interdisciplinary, active group of researchers should generate a significant scientific impact and greatly enhance the international visibility. We expect that the establishment of the CEDS will result in an increased number of high-impact publications, a larger number of competitive research projects, organisation of most relevant international events, education of young researchers, and increased incoming and outgoing mobility.

CEDS will build partnerships with academic, government, and business partners in the areas of expertise covered by the CEDS research team, such as machine learning, data mining, complex networks and social networks, bioinformatics, natural language processing, text mining, business analytics, high-performance computing infrastructure, signal and image processing, and financial applications.

Mission. CEDS’es mission is to become the regional leader in research and applications of Data Science to improve the quality of life and support the economic growth in Croatia. To this end, CEDS will collaborate with other academic partners, industry, and government agencies.

Target groups. The target groups are the CEDS member institutions that will enhance their capacity for (1) research and (2) technology transfer in Data Science theory and applications.

Final beneficiaries. Final beneficiaries are the national scientific community, Croatian companies, national government, and the society as a whole. The resulting economic growth and development of society in general represents a benefit for the whole society.

Overall objectives. The overall objectives of the CEDS are:
- Research excellence: Advance the state of the art in Data Science and develop new technologies, while increasing the number of research grants at the national and international level;
- Transfer of technology: Encourage the transfer of technology from scientific realm to the industry and government, while supporting the growth of SMEs and large enterprises in Croatia;
- Consultations and resources: Provide the government and industry with an access to advanced computing facilities and expertise;
- Education and training: Increase the number of highly skilled data scientists to satisfy the needs of the national job market by supporting undergraduate and graduate programs.

Specific objectives. The overall objectives will be achieved through specific objectives of the CEDS:
- Conducting specific research activities in Data Science foundations and applications;
- Strengthening innovation and exploitation activities to increase the transfer of technology and to provide expertise to industry and government to support the economic growth in Croatia;
- Networking and dissemination activities aiming to create strategic partnerships with academic and industry partners at the national level, and to increase the pool of well-trained and highly skilled data scientists.
1.1 Expected Results

Research Activities. The expected research-related results of the CEDS are: (1) New methods and algorithms for specific Data Science problems, (2) Scientific publications published in top Data Science-related and specific application-related journals, and (3) Research project proposals prepared and submitted to national and international research funding agencies. We expect that, in the course of five years, the research team will propose a minimum of 20 project proposals to the Croatian Science Foundation and 40 applications for the H2020 programme, including the Future Emerging Technologies (FET) topics of the H2020. We plan to at least double the number of journal papers in the span of five years and participate at all major international conferences relevant for Data Science. With respect to visibility, our plan is to regularly (co)organize local/regional meetings related to Data Science, analytics, and big data technologies. In addition, we plan to host and organize one conference of a medium rank, such as Discovery Science, and one of the high profile conferences, such as ECML/PKDD (European Conference for Machine Learning and Knowledge Discovery from Databases).

Innovation and Exploitation. The expected results are: (1) Established mechanism for innovation screening, where transfer technology offices of CEDS member institutions will be in charge of conducting common innovation screening procedures, (2) Developed technology transfer and IP protection policies for CEDS, and (3) Strengthened collaboration with the industry by increasing applied research grants funded by national programs (e.g., HAMAG-BICRO) and further extended to companies and funding opportunities in the context of national competence centres.

Networking and Dissemination. The expected results are: (1) Intra-consortium networking activities such as member plenary meetings and Data Science summer schools aimed at young researchers; (2) Networking with external partners including academic, industrial, government, and society through popularisation events; (3) Dissemination activities in the form of scientific publishing, new Data Science curricula; and (4) Networking and dissemination in digital world, which includes the design of the visual identity of the CEDS and digital PR through social networks (Twitter profile, Youtube channel, Researchgate profiles of CEDS members).

1.2 Relevance to National and EU Strategies

Current status and future trends. According to a study “Exploring Data-Driven Innovation as a New Source of Growth” by the Organisation for Economic Co-operation and Development (OECD) from 2013, the potential value from Big Data in the EU Public administration could be between EUR 150 billion to EUR 300 billion in new value every year (Considering the governments of the 23 largest EU member states). According to the McKinsey Global Institute (2011), the potential value from Big Data in the Healthcare and Social Care may be more than EUR 90 billion in new value every year considering only the reduction of national healthcare expenditure in the EU. According to the OECD study, the potential value from transport and logistics may be more than EUR 500 billion in new value every year worldwide in the form of time and fuel savings, or 380 megatonnes of CO2 emissions saved. The Big Data value market measured by the revenue that vendors earn from sales of related hardware, software, and ICT services is a fast growing multibillion-euro business. According to the IDC, a US market intelligence provider, report from 2013, the Big Data market is growing six times faster than the overall ICT market. The compound annual growth rate of the Big Data market over the period 2013-2017 will be around 27%, reaching an overall total of $50 billion. The economy growing around Big Data could be a major opportunity for Europe to become a leader in this area. Publicly funded through the Horizon 2020 programme, the European Data Science Academy (EDSA) is an initiative aimed at education and training of new data scientists across Europe. The United States face a similar situation. According to the McKinsey Global Institute (2011), by 2018 the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills, as well as 1.5 million managers and analysts with the know-how in Big Data analysis mandatory for effective decisions. These facts clearly show the need and opportunity to establish a national CEDS in Croatia, capable of making a significant contribution to the economic growth in the region and well-being of Croatian citizens.

Relevance to EU strategic documents and policies. The CEDS proposal is perfectly aligned with the following EU strategies: (1) the EU Strategic Forum for Research Infrastructure (ESFRI) Roadmap and (2) the EU Operational Programme Competitiveness and Cohesion 2014-2020 (for Croatia). These strategic documents strongly emphasize the excellence, collaboration, knowledge transfer, innovation, and research capacity building in the field of fundamental science and engineering. Some of the main funding priorities to be tackled by public policy measures by the year 2020 and foreseen under the “Priority Axis 1” are: strengthening research excellence and performance; creating conditions for business sector to invest into R&D&I by directing public research organisations focus towards applied research projects, aligned with the needs of industry and economy; and supporting innovation and knowledge transfer.

Relevance to national strategic documents and policies. CEDS objectives are perfectly aligned with the following strategic documents: (1) Strategy for Science, Education and Technology of Republic of Croatia 2014 and (2) Croatian Research and Innovation Infrastructure Roadmap 2014. The latter lists the following priority areas for the development of research infrastructure in the Republic of Croatia: Biomedicine, Biotechnology, Natural sciences, Engineering, Social
Sciences and Humanities, and Interdisciplinary Science. Data Science is a fundamental prerequisite for all of the above priority areas and there is a clear and substantial potential to encourage the economic growth in Croatia through knowledge-based decision making based on the analytical approach of Data Science.

Relevance to national eInfrastructure. Croatia began the development of eInfrastructure through the CRO GRID project in 2003, which resulted in the establishment of the national grid infrastructure (CRO NGI), connecting all universities and scientific research institutions in the country. CRO NGI became a part of the European Grid Infrastructure (EGI). The EU structural funds will enable the upgrade and expansion of the existing eInfrastructure (CRO NGI with 1112, and computing cluster Isabella with 708 processing cores) in Croatia with 7000 additional processor cores at the University of Rijeka, and through the HR ZOO project with 22300 cores distributed across all universities. The investment is estimated at 25M EUR. In order to benefit from the existing and planned infrastructure, it is necessary to create the national CEDS as an additional scientific/application upper layer of eInfrastructure in Croatia.

1.3 CEDS Member Institutions and Coordinators

To fulfill the overall and specific objectives of the CEDS, and thus generate a significant impact on Croatian science and economy, we have formed a multidisciplinary research team by creating a nationwide partnership of the major higher education and research institutions in Croatia. The members of the research team are chosen based on their scientific and professional experience and research excellence in the fields relevant to Data Science. The research team includes some of the most productive and most cited Croatian scientists whose research relates to foundations and applications of Data Science.

The CEDS member institutions
1. Faculty of Electrical Engineering and Computing, University of Zagreb (UNIZG-FER),
2. Ruder Bošković Institute (RBI), Zagreb,
3. Faculty of Science, University of Zagreb (UNIZG-PMF),
4. Faculty of Transport and Traffic Sciences, University of Zagreb (UNIZG-FPZ),
5. Catholic University of Croatia (CUC),
6. Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split (UNIST-FESB),
7. University of Rijeka, Center for advanced computing and modelling
8. Faculty of Civil Engineering, University of Rijeka (UNIRI-GF),
9. Faculty of Electrical Engineering, University of Osijek (UNIOS-ETF),

1.4 Expressions of Support from Industry and Government Agencies

The initiative to establish a national Centre of Excellence in Data Science has been enthusiastically welcomed by SMEs and large enterprises in Croatia, as well as government agencies that are faced with the challenge of using Big Data or providing Big Data-related services in their everyday business activities. The support to this proposal is evident by fourteen letters of support by SMEs, large enterprises, and a government agency that express the need for and interest in such a centre, as well as an interest in collaboration with the CEDS once it is established. The letters of support are included in the supplementary documentation submitted as a part of this proposal.

1.5 Thematic Activities

A0: Project management
Task 1: Project management
Task 2: Project quality assurance
Task 3: Equipment procurement

A1: Strategic research area: Multimodal data processing, data handling and knowledge representation
Topic A1.1: Text and NLP processing techniques
Topic A1.2: Image, video and signal processing, blind source separation
Topic A1.3: Advanced data and knowledge oriented DB: (storage, querying, and knowledge presentation, NoSQL DB, Graph DB, RF-triple-stores, data warehousing, OLAP, spatio-temporal databases and data streams

A2: Strategic research area: Machine learning and data mining techniques
Topic A2.1: Versatile and scalable algorithms, Multilabel classification; active and online learning; semi-
supervised learning
Topic A2.2: Interpretative models: Multi-view clustering and re-description mining, Causality enabled data mining techniques
Topic A2.4: time-series & data streams mining
Topic A2.5: Machine learning techniques for mining complex systems and networks, Modelling & mining processes in multi-layer networks, Outlier detection in graphs
Topic A2.6: Visualization of data and models

A3: Strategic research area: Heterogeneous computing and advanced cloud services
Topic A3.1: Control/Data Flow Computing Architecture
Topic A3.2: Algorithms for Heterogeneous Computing
Topic A3.3: Energy aware Algorithms and Computing Architectures
Topic A3.4: Scalable Cloud and Fog Scientific Computing Services

A4: Strategic research area: Application use cases
Topic A4.1: Bio-sciences and healthcare
Topic A4.2: Business analytics and finance
Topic A4.3: Web & multimedia
Topic A4.4: Intelligent transport solutions
Topic A4.5: Aerospace

A5: Innovation and Exploitation
Task 1: Transfer of technology
Task 2: Intellectual property management

A6: Networking and dissemination
Task 1: Advanced data science courses and workshops
Task 2: Networking with industry
Task 3: Networking with EU organizations