

Centre of Excellence in Computer Science (EXCS)

Description of Work*

1 Motivation

Computer science (CS) studies the theoretical foundations of information and computation and their implementation and application in computer systems. In Estonia, research in this discipline has long traditions, going back to at least 1960 when the Institute of Cybernetics was founded in Tallinn. More recently, based on this experience and the competence that developed, a flourishing software industry has taken root here. The same competence has enabled extensive and efficient use of information technology (IT) in all sectors of the economy.

CS research in today's Estonia is of high quality and internationally well visible, especially against the small number of people involved countrywide and the very tough period of the 1990s when it lost many of its best talents to universities abroad and industry, as research funding was poor while other opportunities abounded. The people involved today have a long tradition of extensive collaboration, which has given them a considerable edge. But the full potential of the community is nevertheless hampered by the fragmentation into groups by different institutions, with both overlaps and complementarity of competencies and research activities. This calls for coordination that would help consolidate the research on the different lines pursued, avoiding duplication of efforts and enabling maximal use of the complementary competencies, as well as make the groups stronger players in the international research community, help them reach the potential users of their technologies more effectively etc.

EXCS is a coordination body of exactly the desired quality. It is based on a constellation of groups that has a strong record of prior successful collaboration. It capitalizes on the experience of Centre for Dependable Computing (CDC), the Estonian CoE 2002-2007 in computer science and engineering and the related FP5 IST project eVikings II (2002-2005) of which it is a reconfiguration. The project consortium is confident about its capability to achieve its objectives.

2 Consortium

EXCS will involve three institutions:

- **Institute of Cybernetics (IoC)**, a research institute at the **Tallinn University of Technology (TUT)**,
- **Cybernetica AS (CybAS)** (represented by its Information Security Institute),
- **Institute of Computer Science (ICS)** of the **University of Tartu (UT)**.

Its staff will consist of the research staff of the following target-financed themes (TFTs):

- **0322709s06 Dependable Software and Human Language Technologies (2006-2011) (TFT-U)**: This TFT at IoC is led by Tarmo Uustalu and covers the research staff of its Software department and Lab for phonetics and speech technology. The software dept. (Tarmo Uustalu, Varmo Vene, Hellis Tamm, Enn Tyugu et al.) has expertise in logic, programming language semantics, automata theory, ontologies, automated software engineering, computer-aided engineering. The phonetics lab (Einar Meister, Tanel Alumäe et al.) is the sole unit in Estonia working on computer models of spoken Estonian and has produced a number of speech synthesis and recognition systems.

The software dept. is very active internationally. It participated in the FP5 IST project APPSEM II, is participating in FP6 IST projects TYPES and MOBIUS, COST action IC0701, organized Baltic DB&IS 2002, APPSEM 2004, AFP 2004, ADBIS 2005, TFP/ICFP/GPCE 2005, MPC/AMAST 2006, JCKBSE 2006, is preparing Baltic DB&IS 2008. It also coordinated the CDC Estonian CoE and the FP5 IST project eVikings II (discussed in more detail below). The department is engaged in R&D projects for the Ministry of Defence. The phonetics lab contributes to the NordForsk network VISPP, participates actively in the Estonian national programme for Language-technological support for the Estonian language (EKKTT), organized HLT 2005, VISPP 2005, is preparing VISPP 2008.

*The electronic version (pdf) of this document is hypertext, with many links to the web. It is available online at <http://cs.ioc.ee/excs/>. The Description of Work is on the first 8 pages, the remainder is additional information.

- **0012708s06 *Theoretical and Practical Security of Heterogenous Information Systems (2006-2011) (TFT-B)***: This TFT, led by Ahto Buldas and involving Helger Lipmaa, Jan Willemson et al., has CyBAS as its base institution and covers the research staff of its Information Security Research Institute. The group has ten years of experience in various sub-areas of cryptography and information security, including PKI (particularly timestamping), multiparty computation, privacy-preserving data mining, economics of IT security.

The group played a key role in the design of the Estonian digital signature and e-election systems. The group participated in the FP5 IST projects OpenEvidence and RESET, is participating in the FP6 IST projects AEOLUS, and BalticTime and the FP7 ICT project VirtualLife. It organized NordSec 2005.

- **0182712s06 *The Methods, Environments, and Applications for Solving Large and Complex Computational Problems (TFT-V) (2006-2011)***: The leader of this TFT is Jaak Vilo. Hosted by ICS/UT, the group comprises its teams working on bioinformatics (prof. Jaak Vilo), distributed systems (prof. Eero Vainikko), software engineering (prof. Marlon Dumas), programming language semantics (prof. Varmo Vene, collaboration with IoC), data security (senior researcher Peeter Laud) and educational software (assoc. prof. Rein Prank). The department carries on the traditions of CS research and teaching at UT.

J. Vilo participated in the FP6 LSH project ATD (via EBC), is participating in FP6 LSH projects ENFIN (via EGeen) and COBRED and in the FP7 HEALTH project ESNATS (via Quretec). TFT-V organized the 2006 and 2007 Estonian summer schools in computer and system engineering (ESSCaSS) and is preparing ESSCaSS 2008. M. Dumas is a recently appointed professor (position funded by a pledge from Hansapank) and has extensive previous experience in managing research projects (both public and industry-funded) from his previous positions in Australia.

- **0180078s08 *Development and Implementation of Formalisms and Efficient Algorithms of Natural Language Processing for the Estonian Language (2008-2013) (TFT-K)***: This TFT is based on the language technology chair of ICS/UT, led by prof. Mare Koit, but involves also researchers from the general linguistics chair of UT. It continues an earlier TFT that was led by prof. em. Haldur Õim of the general linguistics chair. The group has twenty years of experience in natural language processing, esp. the Estonian language. It has collected large text corpora of Estonian and has developed various software for processing written Estonian.

The group is running several EKKTT projects, organized NODALIDA 2007.

Most of the key persons in the consortium are young researchers. Three of them, Marlon Dumas, Varmo Vene, Jaak Vilo became professors at UT in 2007. Several have PhDs from abroad (Eero Vainikko, Tarmo Uustalu, Peeter Laud, Jaak Vilo, Hellis Tamm, Sven Laur), have postdoc/researcher experience from abroad (Eero Vainikko, Tarmo Uustalu, Varmo Vene, Jaak Vilo, Helger Lipmaa, Kaarel Kaljurand), several are from abroad (Ulrich Norbistrath, Kristiina Jokinen, Marlon Dumas, Phaedra Agius).

Prior collaboration between the partners The partners have collaborated in formalized ways for about ten years in various configurations. Most importantly for this proposal, TFT-U, TFT-B and TFT-V were the main actors in 3 working groups out of the total 4 in the Centre for Dependable Computing (CDC), the Estonian CoE in computer science and engineering 2002-2007. The same TFTs were also the main contributors to WP2 (Strengthening R&D in software technology & trust and confidence) of the related FP5 IST project eVikings II while TFT-K and TFT-U had this role in WP3 (Strengthening R&D in language technology).

Further, TFT-U, TFT-B and TFT-V have since 1996 jointly organized the annual international Estonian winter school in computer science (EWSCS) and since 2002 the biannual Tallinn-Tartu theory days. Recently, TFT-U, TFT-B and TFT-V have been working together in the Doctoral School of Information and Communication Technologies (IKTDC) (2005-2008) of Measure 1.1 of the Estonian NDP for the implementation of the EU Structural Funds whereas TFT-K and TFT-U have been collaborating within the Doctoral School of Linguistics and Language Technology (KTTDK) of the same programme. TFT-K and TFT-U are collaborating also in the NGSALT NordForsk doctoral school.

Infrastructure All partners have adequate infrastructure for carrying out the proposed research. This concerns both general IT-infrastructure, specific equipment and software (for scientific computing, phonetics/speech technology), access to scientific literature via physical and digital libraries and lab/office space. Through the subscriptions of the ELNET consortium, the researchers have access to the digital libraries of virtually all relevant professional societies and publishers of journals and conference records. The language technology groups have collected large text and speech corpora.

3 Organization and Management

The research activities EXCS will be organized around working groups (WGs), aligned along the competencies and interests of the researchers and crossing, wherever appropriate, the TFT and institution boundaries.

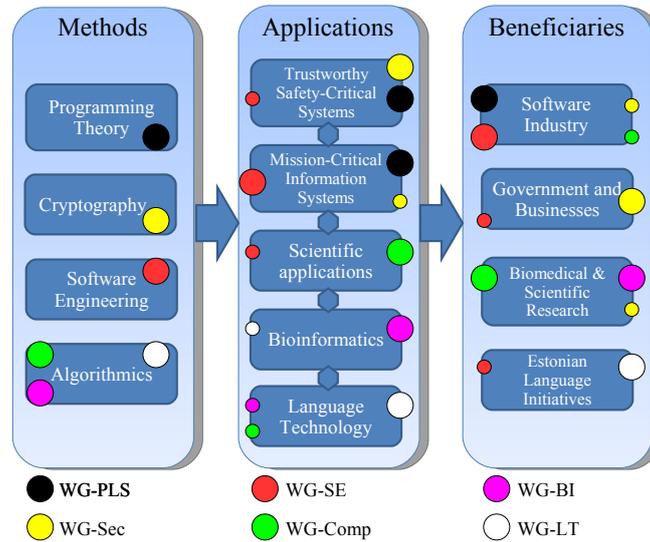


Figure 1: Activities of working groups

A management committee (MC) will coordinate the research activity and support actions of the centre, keeping, in particular, a close watch on potential synergies between the WGs. In addition, the centre will have a general assembly (GA) representing the partners involved and an international advisory board (IAB) to consult the MC.

3.1 Working Groups

EXCS will have six working groups (WGs) with well-defined research areas, corresponding to the competencies of the TFT involved and the strength of the Estonian CS. A WG will typically comprise researchers from several TFTs and institutions. Fig. 1 depicts the interests of the WGs in various theoretical and practical areas in and around CS.

The six WGs will be these:

- **Programming language and systems (WG-PLS):** The topics of study will be functional languages, categorical semantics, program logics and type systems for reasoning about correctness and safety of imperative code and code transformations, certified code, static analysis of multithreaded code, algebraic and categorical theory of automata and transducers. This WG will involve researchers from TFT-U and TFT-V.
- **Security (WG-Sec):** The topics will be design of protocols for various tasks (oblivious transfer, e-voting etc.) and formalization of methods for determining the security of complex primitives and protocols. This WG will be based on the research staff of TFT-B (protocol design) and TFT-V (reasoning methods).
- **Software engineering (WG-SE):** The topics will be enterprise software engineering, model-driven architecture, service-oriented architectures, composition of web-services, the semantic web etc. The WG will involve people of TFT-V and TFT-U.
- **Scientific and engineering computing (WG-Comp):** The topics will be parallel numerical methods, grid and peer-to-peer computing, systems for computer-aided engineering. The WG will be based on TFT-V (high-performance computing, grids), but will also include researchers from TFT-U (computer-aided engineering).
- **Bioinformatics (WG-BI):** The topics will be string algorithm design, adaptation of machine learning and data mining approaches to interpretation of biological data. The WG will mostly be based on TFT-V, but also involve TFT-B.
- **Human language technology (WG-LT):** This WG will develop written and spoken Estonian language technology, studying linguistic changes of Estonian, discourse and dialogue modelling, machine translation to/from Estonian, phonetical models and synthesis and recognition of spontaneous and emotional Estonian speech, Russian accent. This WG will pool the resources of TFT-K (specializing in written language technology) and part of TFT-U (specializing in speech technology).

The last two WGs are interdisciplinary overlapping with genetics and linguistics.

Each WG will nominate a WG leader from among the researchers in this WG. This person will be responsible for coordinating the inter-institutional activities of the WG and for representing it in the management committee.

3.2 Management Committee

The management committee (MC) will consist of the TFT and WG leaders. The MC will meet regularly (at least once per three months) to coordinate cross-WG research activities and the joint actions of the centre. The MC will have the power to form and appoint members to subcommittees tasked with the coordination of a particular activity.

3.3 General Assembly

The general assembly (GA) will consist of administration representatives of the institutions involved (one person from each of the 3 institutions) and the leaders of the TFTs. It will meet at least once a year to decide on administrative matters. The MC will report to the GA.

Both the MC and the GA will be chaired by the leader of the centre. Decisions should be reached through consensus. In case of a serious disagreement, the leader of EXCS will make the final decision.

3.4 International Advisory Board

The international advisory board (IAB) will consist of 6..7 internationally renowned researchers from the research areas of the centre. The IAB will advise the MC on the research plans of the centre, review the annual progress reports of the centre and meet at least once a year.

The members of the IAB of the CDC CoE (Reino Kurki-Suonio, Tampere UT, Kim Larsen, Aalborg Univ., José Oliveira, Univ. do Minho, Reinhard Wilhelm, Univ. des Saarlandes) have agreed to continue. We will add members to cover the research areas that were not present in CDC.

4 Implementation Plan

The centre's activities will fall into two categories: research and support actions.

4.1 Research Activities

The research activities of EXCS will be carried out foremostly through the six WGs listed above, progressing in parallel. We expect all WGs to be active from the beginning to the end of EXCS. One of the main benefits of EXCS will be synergies from the pooling of resources from several TFTs into a WG.

EXCS will also pursue cross-WG topics. For this purpose, the MC may form additional working groups, with possibly a shorter life-time.

4.1.1 Research within Individual WGs

WG-PLS This WG will pursue theoretical/language design issues concerning (co)inductive types and structured (co)recursion, effects and context-dependence in functional and type-theoretic languages. Of particular interest are language designs that support these idioms in a truly practical way, radically better than today's functional languages or proof assistants do.

The WG will continue working on program logics/type systems for certifying imperative (in particular unstructured low-level) code and code transformations (for languages like Java, C#) for functional correctness, and to verify non-functional properties such as optimal resource utilization and security. In particular, we will look at safety mechanisms for scripting languages combining in practical ways the flexibility of weak, dynamic typing and the safety that comes cheaply with strong, static typing. We will demonstrate the feasibility of automatic certification for functional correctness of code generated from DSL descriptions, in particular, simulation code generated from system models.

The WG will also study static program analysis of multithreaded programs. The particular focus will be on developing accurate and efficient static analysis methods aimed at guaranteeing the absence of resource access violations in concurrent safety critical programs.

It will work on the algebraic and categorical theory of (word and tree) automata and transducers, relating on the one hand to questions of combining context-dependence with functional computation and on the other hand to applications in document typing and transformation.

WG-Sec The WG will continue its work on efficient secure communication protocols for various tasks, including oblivious transfer, e-voting, privacy-preserving information retrieval and other tasks related to privacy-preserving data mining (PPDM) and secure function evaluation in general, time-stamping and other protocols for PKIs with an emphasis on availability issues. The WG will also continue the work on the formalization of methods for determining the (level of) security of various systems on various levels of abstraction and integrated defense against attackers with various goals. A system may mean a (complex) cryptographic primitive built up from simpler primitives. It may also mean a protocol or set of protocols to achieve a distributed computation task, or an entire organization, of which the information systems

are just a part. The WG's primary aims will be to devise: PPDM protocols and implementations that are efficient enough to be used in practice for processing sensitive data; efficient and simple-to-understand e-voting protocols; a scalable and tool-supported framework to formally establish the security of mechanisms of various granularities (from the primitive level to the organizational or global level), including the protocols mentioned above.

WG-SE This WG will devise and evaluate software engineering methods for large-scale distributed systems. Emphasis will be placed on methods that draw upon the principles of Service-Oriented Architectures (SOA). The WG will develop methods for: (i) rapidly aggregating semantically-described services according to user-defined goals; and (ii) connecting service aggregations to end-user interfaces. In this line, the WG will develop methods for semantics extraction, including induction of ontologies from annotated texts, database schemas, web data sources, etc. The resulting methods will be applied to business and government system integration as well as scientific applications (in collaboration with the other WGs). A second stream of research will apply data mining techniques in conjunction with concepts from software metrics, to devise and evaluate methods for measuring architectural attributes of service-oriented systems. This research will lead to: (i) guidelines to help designers of services to make informed architectural decisions; and (ii) tools to support the automated measurement of architectural attributes of service-oriented systems. These methods will be evaluated in various application domains, e.g. enterprise applications and grid computing.

The WG will also work on ontology learning, including induction of ontologies from annotated texts, (schemas of) relational databases etc, web mining.

WG-Comp The WG will devise new parallel algorithms for solving large and complex computational problems efficiently on modern distributed computing architectures. This involves parallel implementation of the algorithms within their own parallel black box solver DOUG (Domain Decomposition on Unstructured Grids) for solution of large systems of linear equations. DOUG and other similar computational packages will be adapted for computational Grids and novel Peer-to-peer (P2P) architectures, like the Friend-to-friend (F2F) computing framework. This work will be carried out in close collaboration with other workgroups like software engineering effort for SOA research.

The group will work on their own lightweight Grid middleware, the F2F computational architecture, based on the ease of setup through instant messaging systems, allowing fast aggregation of distributed computational resources into P2P structures for solving large parallel tasks for scientific and engineering computations. Many interesting research problems arise with such environments including the security issues. The group will work towards the ease of porting existing applications to the F2F framework and devising application development environments enabling simple creation of parallel code for F2F computing framework.

The WG will also work on extensions to Cocovila, a visual tool featuring a method of automated program synthesis, and apply it to engineering computations and modelling and simulation.

WG-BI The WG will focus on providing value-added services for biological data interpretation through applying advanced algorithmic designs, developing novel analysis methodologies and tools, applying machine learning and data mining approaches, visualization, techniques and hypothesis prioritization. Most importantly we continue to contribute research to dissecting the gene regulatory mechanisms, reconstruction and analysis of genetic networks, and gene expression data mining. These techniques are applied to understanding the risks and mechanisms of cancer, developmental and embryonic stem cell biology, novel drug screening strategies, and mining heterogeneous health and lifestyle related data in conjunction with genetic markers potentially related to diseases. The expected high impact results from bioinformatics WG will continue to be related to two scientific communities — the algorithm design oriented computer science and discovering new biological knowledge.

WG-LT In the written language technology research, pursued in TFT-K, the linguistic changes of the Estonian language will be analysed and modelled on the lexical level to guarantee that the tools of the computational processing of the language cope with the actual language use. The formalisms for representing the deep structure of the sentence will be studied in order to move from surface analysis to deep analysis. Various conceptual and formal means for the semantic representation of individual sentences, as well as the discourse, will be assessed as to their applicability to the Estonian language. The theoretical problems of dialogue modelling will also be studied. The strategies and formalisms of machine translation best suited in the case of a source and/or target language with free word order and rich morphology will be determined.

The spoken language technology research of TFT-U will focus on studies of speech recognition methods for Estonian. This includes the development of morpho-syntactic language model and adapting methods for semantic analysis in statistical language model. Adequate methods for coping with different types of speech signal variability will also be developed. Sizeable audio-visual corpora of spoken speech will be collected, together with the development of necessary storage solutions.

The WG will also develop man-machine dialogue systems where components of speech recognition and synthesis will be provided by TFT-U and the models for dialogue management by TFT-K.

4.1.2 Cross-WG Research Efforts

The following are some cross-fertilization possibilities, where problems rising in the area of one WG can be elegantly solved using the methods and tools of different ones. This list is bound to increase as work progresses.

- Several cooperation possibilities exist between WG-PLS and WG-Sec. To design reasoning methods for cryptographic protocols (one of the goals of WG-Sec), a suitable semantics has to be chosen for the language these protocols are expressed in. On the other hand, to handle cryptographic primitives in the arguments on or analyses of programs in a sound manner (a task of WG-PLS), one has to refer to the precise complexity-theoretical security definitions of those primitives.
- By applying knowledge from theoretical aspects of cryptography (WG-Sec) to business process modeling (WG-SE) and distributed algorithms (WG-Comp) we will seek to develop practical secure bioinformatics scientific workflow methodologies (WG-BI).
- Similar methods — text algorithm design, hidden Markov models — have been used for the (similar) tasks of mining DNA sequences for “hidden” signals and for information retrieval from text/speech in natural languages. It makes sense to try to apply any method devised for one of the tasks also on the other problem. This is a collaboration possibility between WG-BI and WG-LT.
- The research in WG-LT may make use of large language corpora, requiring a lot of computing power to process. Such power can be gathered in the most straightforward way by employing distributed computing, thus also requiring distributed algorithms that WG-Comp can provide.

The MC is responsible for locating such opportunities and acting on them.

4.1.3 Dissemination

The primary means of dissemination of EXCS will be **scientific publications** at venues recognized as high-level by the worldwide CS community. This means both journals (ISI WoS and other), which in CS are foremostly the venue for full archival versions of completed cycles of work, as well as pre-refereed conference proceedings, which are often very competitive. Some of the high-level proceedings series are Springer’s *Lect. Notes in Comput. Sci.* (from 2007 indexed only in ISI Proceedings), publications of ACM, Elsevier’s *Electron. Notes in Theor. Comput. Sci.* (not indexed even in ISI Proceedings).

The MC will monitor the impact of the centre’s publications relying on online databases and services accepted as de facto standard in CS. This includes bibliographic/citation databases such as Google Scholar and CiteSeer^x and the various analytic tools such as Harzing’s PoP using the former as data sources. These open-access databases index articles from digital libraries of the publishers, from public eprint archives etc., yielding a much more relevant choice of publication venues and these days also increasingly higher precision than what the commercial ISI databases can offer.

In addition to dissemination via publications, the researchers will also give **tutorials** at high-level international events, present **intensive courses and seminar talks** at foreign universities etc.

The **web portal** of the centre will be a repository of the research results for the centre, with pages for individual WGs and smaller projects, links to online versions of the publications etc.

4.2 Coordinated Support Actions

Coordinated support actions that the individual TFTs would not be able to undertake on their own, will be another major source of added value besides the joint research efforts of the TFTs.

4.2.1 Developing the Centre

Most of the Structural Funds support will be spent on strengthening the centre, making it a sustainable **thriving research environment**, capable of attracting and keeping talent in the fierce competition of institutions performing CS research and the software industry.

- **Human resources:** Human resources are the most valuable and expensive asset of the centre and strategically the most critical. Correspondingly, development of the human resources constitutes is the most important action.
 - **Positions for postdocs, PhD students, technical personnel:** EXCS will spend the largest share of its funding on positions for postdocs and PhD students, aiming to strengthen the centre by best available talent (which will in many cases likely be from abroad). Some of these funds will also be used to hire technical personnel to carry out larger-scale code development projects for technology demonstration and experiments (esp. in scientific computing, bioinformatics, text/speech technology), technical/popular writers, personnel to administer the centre etc.

- **Training:** EXCS will train its staff through internal seminars, workshops (in particular, Tallinn-Tartu theory days) and by sending its PhD students to international summer schools.
- **Equipment:** EXCS will also carefully develop its physical research environment. It will invest into storage for large data volumes, high-performance computing (clusters), various phonetic and speech-technological lab equipment of the TFTs. It will also ensure that the general personal computing equipment and central IT-infrastructure of the TFTs are up-to-date at any given time. The same applies to licences for the necessary general and special-purpose commercial software.

Both personnel hire and equipment purchase will be carefully planned by the MC to ensure that all WGs will at all times be adequately staffed and equipped.

4.2.2 Increasing the Impact

Impact actions will generally be directed at enhancing the centre's **visibility**, i.e., raising the awareness of the relevant target groups of the centre's research results. This will help attracting talent and material resources to the centre, thus contributing to its further development, while the society will receive the maximum return from the investment.

- **International cooperation:** The centre will be well visible in the international research community. This will in particular help making the centre attractive as a workplace for foreign researchers.
 - **Organization of high-level scientific events:** With the TFTs involved having a strong track record in organizing high-level international scientific events in Estonia, EXCS will support this activity. A major conference will be organized by the centre once every two years; with 1..2 smaller workshops organized each year. The organization of the annual international EWSCS and ESSCaSS schools with distinguished lecturers will continue as well.
 - **International cooperation projects:** The TFTs will participate in the preparation and implementation of EU FP and COST projects as well as bilateral projects etc. Each TFT will at any given time be active in 1..2 such projects.
- **Technology transfer:** The centre will go a long way to make its technologies known and taken up by the envisaged end-users.
 - **Contact events:** The centre will organize 1..2 contact days yearly for potential industrial partners from Estonia. The symposia on innovative software engineering will be continued.
 - **Industrial cooperation projects:** Not compromising its primary dedication to basic research, EXCS will engage in both international and industry cooperation projects (EUREKA, EAS), projects with individual governmental departments, businesses etc. This includes projects with the Ministry of Defence, collaboration with SIVAK on the NATO centre of competence, with Skype etc.
- **Contribution to policy-making:** EXCS will assume a proactive role in shaping the government policies in R&D, higher education and IT related areas. It will lobby for improvements in the current system of PhD studies, for a national programme for IT R&D. It will consult the relevant governmental departments on issues such as data interchange between state information systems, data protection, privacy protection, cyber defence etc. Likewise it will be involved in technology roadmapping, foresighting on both domestically and at the European level, involving participation at the European strategy forum on research infrastructures (ESFRI - CLARIN (language technology), ELIXIR (bioinformatics)).
- **Popularization:** EXCS will extensively popularize its research results, but also the subject of computer science in general. One aim is to attract high-school students to study computing.
 - **Media coverage:** The centre will regularly inform the general public about its research activities and results in printed media (*Horisont, Arvutimaailm* etc.), radio & TV (*Labor, Tehnokratt* etc.) and via its web portal.
 - **Popular books:** EXCS will publish some original popular books in Estonian about some of its areas of strength. It will also publish some important translations (e.g., Harel, *Computers Limited*, Papadimitriou, *Turing*).
 - **Events for the general public:** EXCS will also organize one “open-door” event yearly at one of the institutions involved in the centre for the general public to popularize CS.

In all of the above impact actions EXCS will collaborate with the groups of Raimund Ubar (DCE/TUT), Jüri Vain (DCS/TUT) and Leo Mõtus (DCC/TUT), representing computer engineering and systems engineering. These groups were part of the earlier centre CDC and the eVikings II project and are now in other CoE consortia because of the conditions of the new programme (composition of CoEs from TFTs, size of a CoE < 40 senior staff). Likewise, EXCS will collaborate with the group of Maarja Kruusmaa (Centre for Biorobotics/TUT).

- **Communication of the EU Structural Funds support:** The support from the Structural Funds will be duly advertised wherever appropriate by displaying the correct logotype.

5 Significance

The high level of integration of the research efforts in different CS topics in Estonia that EXCS will bring about will significantly enhance the **research capacity** of the groups involved contributing towards their long-term sustainability. The associated coordinated support actions will boost the internationalization and industrialization of the Estonian CS research. These are prominent contributions to building the **European Research Area** in information and communication technologies (ICT) as one of the **priority areas** of the EU FP7, to the development and transfer of cutting-edge technology in ICT as one of the **key technological areas** for raising Estonia's **economic competitiveness**, and to raising the **innovation capacity** of the Estonian economy.

Concerning cohesion with FP7, it must be remarked that the domain of research of EXCS has major overlaps with at least three of the five **technology pillars** of FP7: technologies of knowledge (including cognition, simulation); advanced and open communication networks; and secure and dependable software. EXCS is particularly strong in several **key bottleneck areas** identified by the European Commission's policy framework *i2010*: interoperability, security and reliability, identity and rights management, ease of use. *i2010* also identifies **inclusive Information Society** as one of main objectives; this is addressed by EXCS's research into language technology, as well as by its technology transfer and policy-making support actions. It is also relevant that the EU policy-makers find it alarming that in its current investments in ICT R&D EU is sorely lacking behind US and Japan.

The popularization actions will increase general interest in CS and attract young people to the discipline.

6 Additional Information

Budget rationale We prioritize personnel, which is by far the most critical resource for us. Overwhelmingly, the salaries are for *new* personnel, often from abroad: on the average, EXCS will at any moment finance 1 senior staff, 2 postdocs, 1 technical person (e.g., code developer, technical writer) and 1 support staff in each of the 4 TFTs.

The bulk of larger investments into equipment concern storage for large data sets and corpora, computing power (clusters), connectivity and specific phonetics and speech technology. Until 2010 (which the end date of the EKKTT national programme on language technology), we prioritize equipment for other research of EXCS.

Bigger investments into equipment will be made during the 1st half of the project period, to exploit it to the maximum benefit during the duration of the project.

Comment to the table "Indicators" ISI WoS does not measure CS research meaningfully. In the table we give a projection about our ISI WoS publications, counting, in addition to the strengthening of the centre, also on changes in Thomson Scientific's policies, foremostly resumed indexing of Springer's *Lect. Notes in Comput. Sci.*. To give an example, our ISI WoS count in 2007 was 13, whereas in 2006, when ISI WoS still indexed LNCS, it was 23!. Our prediction about high-level (= ISI WoS, ISI Proceedings, SCOPUS, INSPEC indexed) publications in journals and conference records for 2008-2015 is: 2008 - 70, 2010 - 95, 2015 - 135, 2018 - 155.

We interpret the ambiguous "funding per researcher" as "funding per full-time equivalents of senior staff" following the overall spirit of the rules.

As appendices to this document we enclose:

- An overview of the activities and results of TFT-U, TFT-B, TFT-V during their implementation (2006 to date) (as requested by HTM/Archimedes) + a similar overview about the activities of TFT-K in the same period (for comparability).
- Bibliometry data from ISI Web of Science for selected senior staff of the centre (as requested by HTM/Archimedes).
- Bibliometry data from Google Scholar for the same persons (for comparison).
- A complete list of publications by members of EXCS.