# Research Evaluation 2011-2015

## Department of Computer Science

**Computer Science, Aalborg University**
**Research Evaluation 2011-2015**

**Edited by:**

Manfred Jaeger
Jesper Kjeldskov
Hua Lu
Brian Nielsen

All pictures except front cover by Jiri Srba
Front cover photograph by Lars Horn, Baghuset

Printed by Novagraf

# Contents

# Evaluation

# 1 The Evaluation

## 1.1 The Evaluation Committee

### Peter Apers

Since 1985 Peter M.G. Apers is a full professor in Computer Science at the University of Twente, the Netherlands. In this role he is involved in a large number of externally funded projects.

After being a member of the Executive Board of the University of Twente (responsible for Research and Innovation) and Scientific Director of the Centre for Telematics and Information Technology (CTIT), he is currently Dean of the Faculty of Electrical Engineering, Mathematics, and Computer Science of the University of Twente.

At the national level, he has been involved in formulating the ICT-policy: as chairman of the Computer Science section of the Dutch Research Council (NWO), member of the ICT Taskforce, member of the ICT Forum, and the board of Exact Sciences of NWO, as member of the Advisory Board of IC-TRegie and as scientific leader of the ICT Roadmap. Together with others he has taken the initiative to found NIRICT (Netherlands Institute for Research on ICT), the ICT institute of the three Dutch Universities of Technologies of which he is now member of the board. Until 2014 he was chairman of the Dutch Technology Foundation STW. At the European level he is active as member of the Executive Steering Board of EIT Digital, a Knowledge and Innovation Community of the European Institute of Innovation and Technology.

---

Previous page, from left to right: Chris Hankin, Peter Apers, Heikki Mannila, Jan Gulliksen, Torben Bach Pedersen

## Jan Gulliksen

Jan Gulliksen is a professor in Human Computer interaction at KTH Royal Institute of Technology, Stockholm, Sweden, since 2009. He is also the Dean of the School of Computer Science and Communication. Before joining KTH he was a professor of Human Computer Interaction at Uppsala University. Jan is the chairman of the Digital Commission of Sweden's government, Department of the Enterprise and he has also been appointed Digital champion of Sweden by the European commission.

Jan is the chairman of IFIP TC 13 on HCI, chairman of the NordiCHI collaboration and founder of the NordiCHI conference series. Jan has conducted a number of larger Action Research projects with the purpose of increasing the usability of digital work support systems by introducing new user-centred design methodologies particularly at several public authorities. He is doing research on accessibility and the effect of digitalization on people with disabilities. Finally he is working on projects relating to the digitalization of industry, digital work environments and digitalization and leadership.

Jan is a fellow of the Academy of Engineering Sciences (IVA) in Sweden.

Jan founded the company Gulliksen Consulting AB in 1999 and is still chairman of the board.

## Chris Hankin

Chris Hankin is Director of the Institute for Security Science and Technology and a Professor of Computing Science. He was Deputy Principal of the Faculty of Engineering 2006-08, Pro Rector (Research) from 2004-06, and Dean of City and Guilds College 2000-03. He joined Imperial College in 1984 and was promoted to full professor in 1995.

His research is in cyber security, data analytics and semantics-based program analysis, in which he has co-authored three well-known textbooks and a large number of peer reviewed papers. He is currently Director of the Research Institute on Trustworthy Industrial Control Systems, a large partnership of 5 university teams.

He is a fellow of the British Computer Society (BCS), Institution of Engineering and Technology (IET), Royal Society of Arts, Commerce and Manufac-

tures (RSA), and the City and Guilds Institute. He has been Editor-in-Chief of ACM Computing Surveys and serves on the Editorial Board of Software Tools for Technology Transfer. the Journal of Logic and Computation, and Springer's Undergraduate Texts in Computer Science.

He was Chairman of the Scientific Board of INRIA from 2011-2015. He was a member of the IST Advisory Group for the European Commission's Directorate General for Information Society and Media (ISTAG) and then Vice Chair of the Connect Advisory Forum (CAF) for DG Connect. He was a member of the Expert Advisory Group for the European Commission's Future and Emerging Technology Flagships. He is a member of the council of ACM Europe and serves on the ACM Assessment and Search Committee. He serves on a number of UK Government committees on cyber security.

## Heikki Mannila

Heikki Mannila received his Ph.D. in computer science in 1985 from the University of Helsinki. After some time at the University of Tampere and various researcher positions, in 1989 he was appointed a professor of computer science at the University of Helsinki. He was a visiting professor in the Technical University of Vienna in 1993 and a visiting researcher at Max Planck Institute for Computer Science in Saarbruecken in 1995-96. He moved to Microsoft Research in Redmond in 1998, came back to Finland to Nokia Research in 1999, where he stayed until the end of 2001. Since 1999 he is a professor of computer science at Helsinki University of Technology (currently Aalto University). In 2004-2008 he was an academy professor. In 2009-2012 he was vice president for academic affairs and deputy president at Aalto University, a new university formed by the merger of Helsinki University of Technology, Helsinki School of Economics, and University of Art and Design Helsinki.

In 2012 Heikki Mannila was appointed the president of the Academy of Finland. The Academy of Finland provides funding for scientific research in all disciplines, handling over 4000 applications per year, of which about 15% are funded, and with a yearly budget of over 300 million euros.

Heikki Mannila is the author of two books and numerous refereed articles in computer science and related areas. His specific area of interest is in algorithms for data analysis, and applications in science and in industry. The

book "Principles of Data Mining", with David Hand and Padhraic Smyth, has been translated also to Chinese and Polish. He received the ACM SIGKDD Innovation award in 2003 and the IEEE ICDM research contributions award in 2009.

### Torben Bach Pedersen (chairman)

Torben Bach Pedersen is a Professor of Computer Science at Aalborg University (AAU), since 2009. Before, he was an Associate Professor at AAU (2000-08), and a business intelligence specialist/ Industrial Ph.D. Fellow in industry (1994-2000). He received his Ph.D. in Computer Science from Aalborg University in 2000.

His research covers big data analytics, multidimensional databases, data warehousing, business intelligence (BI), on-line analytical processing (OLAP), data integration, Linked Data, and data mining, with a focus on "Big Multidimensional Data" - the integration and analysis of large amounts of complex and highly dynamic multidimensional data. His research has been funded by a large number of grants from EU and national sources, including an Erasmus Mundus Joint Doctorate.

He has served as Area Editor for Information Systems and Springer EDBS, Editorial Board member for IJDWM, JCSE, and TLDKS, PC Chair for DAS-FAA, DaWaK, DOLAP, and SSDBM, and on more than 100 conference PCs.

He is an ACM Distinguished Scientist, and a member of the Danish Academy of Technical Sciences (ATV) and the SSTD Endowment. He heads the ATV Theme Group on Digitalization and Big Data and the national Danish (Infinit) Interest Group on Big Data and Business Intelligence. He received the Best Poster Award at the World Smart Grid Forum 2013 for his work on management of energy flexibility.

## 1.2 Evaluation Guidelines

Below are the guidelines for the evaluation committee members for the research evaluation of The Department of Computer Science at Aalborg University in January 2016:

- The members should:

  - Apply the best of their abilities, professional skills, knowledge and ethics, in the evaluation.
  - Work together with the department representative on the evaluation panel.
  - Take part in the evaluation meeting on January 19-21 in Aalborg.
  - Contribute to the written report as specified below, with the report being ready at the end of the meeting or very soon thereafter.
  - Contribute to the presentations of the evaluation results, as specified below.

- Five documents were sent to the committee members on December 18, 2015: the main report (this document) and four appendices for the four research units.

- Each evaluator is primary responsible for one research unit and secondary responsible for another one. The responsibilities are as follows (primary reponsible listed first): Database and Programming Technologies: Peter Apers and Chris Hankin; Distributed and Embedded Systems: Chris Hankin and Peter Apers; Information Systems: Jan Gulliksen and Heikki Mannila; Machine Intelligence: Heikki Mannila and Jan Gulliksen

- All members should read the main report. The primary and secondary responsible should read the relevant unit chapter in the main report with special attention.

- The primary and secondary responsible should read the appendix for the given unit.

- Both the primary and the secondary responsible should provide comments on all aspects of the unit chapter. The other members are encouraged to comment on selected aspects. The primary responsible should integrate the comments into the evaluation report.

- All committee members should provide comments on the department chapter.

- Comments should be on the actual research topics, results, and performance, but also on strategic issues like funding, internal organization and synergies, possible new directions, collaboration with industry, internationalization, positioning IT as a key enabler in society, etc.

- Comments should be given at the department, unit, and, if relevant, subgroup level.

- The main findings should be presented in dedicated feedback sessions, at the unit and department level, as specified by the evaluation program. The primary responsible will present the main findings at the unit level. The external evaluators will jointly present the main findings at the department level.

## 1.3 Evaluation Seminar

The evaluation seminar took place on January 19-21, 2016, at Cassiopeia. On the first day, overview talks on the department and each of the four research units were given. After reflection in the committee, initial feedback was given at the department level. The second day started with two parallel sessions where each unit presented further details and had question and answer sessions with their two evaluators. Following committee reflection, the afternoon started with dedicated feedback sessions for each unit, followed by a meeting with Eskild Holm Nielsen, the Dean of the Faculty of Engineering and Science to discuss strategic issues. The day ended with a dinner with the committee, Dean, Head of Department and the full professors. The last day started with committee discussions, followed by a plenary feedback session with assessment and recommendations for each research unit, the department, as well as the faculty/university level.

Following the seminar, the committee prepared written final versions of their assessments and recommendations. They are included in this report at the end of the chapters for the different organizational and research units.

# 2 The Department of Computer Science

## 2.1 Introduction

In 2015 Aalborg University celebrated its 41st anniversary. In its relative short lifetime the university has evolved from a regional education institution to an internationally oriented university. The establishment of the Faculty of Medicine in 2010 completed the picture, so that all major areas are now covered: Medicine, Humanities, Social Sciences, and Engineering and Science. The faculty of Engineering and Science is the largest of the four, and this is where The Department of Computer Science resides. The faculty is organized in 14 departments and centers and is spread over three campuses in Aalborg, Esbjerg and Copenhagen. The Department of Computer Science has all activities gathered in Aalborg. The department is roughly an average size department with 120 employees and 940 students. The activities are organized in four research units:

*Database and Programming Technologies* with particular emphasis on data management and data mining and on programming languages, environments, and tools.

*Distributed and Embedded Systems* with particular emphasis on real-time, embedded and distributed systems, networks, formalisms for the description and analysis of computer systems, verification and validation methods and tools.

*Information Systems* with particular emphasis on development and use of computerized systems in organizations, human-computer interaction, usability, interaction design, software development, and innovation.

---

Previous page: managerial and administrative staff of the department

*Machine Intelligence* with particular emphasis on machine learning, probabilistic graphical models, and intelligent web and information systems.

The department offers five study programs: Computer Science, Software Engineering, Informatics, Information Technology and Interaction Design. Common for these educations is a core of traditional computer science, and the three latter are interdisciplinary by nature - each flavored by inclusion of elements from the humanities, business or design.

The activities of the department are mainly funded by two sources: internal funding allocated by the faculty, and external funding granted by national and international research councils and agencies. The internal contribution is mainly derived from delivered teaching (app. 75%), but there are also contributions based on publications and turnover of external grants. Over the years a larger fraction of national research contributions have been directed to research agencies and these resources are obtained by application. At the beginning of the evaluation period 50% research time was allocated for each hour of teaching and at the end of the period this contribution made up 37%. Hence, larger emphasis is placed on external research grants, and the interaction between teaching and research is slightly loosened. This has been partially dealt with by increasing teaching by non-tenured staff.

We are probably entering a new economic reality, where increased activity is no longer automatically resulting in increased funding. There is a demand from society that we increase efficiency and are expected to sustain the existing activities despite a relatively declining funding.

## 2.2 Facts and Figures

The education activities play an important role for the internal funds and consequently also indirectly for research and dissemination activities. In the following we take a closer look on the development of students, funding, staff and research output.

### 2.2.1 Students and teaching

The educations offered by the department have attracted an increasing number of students over the past decade. In 2009 we needed three digits to count

**Intake of students**
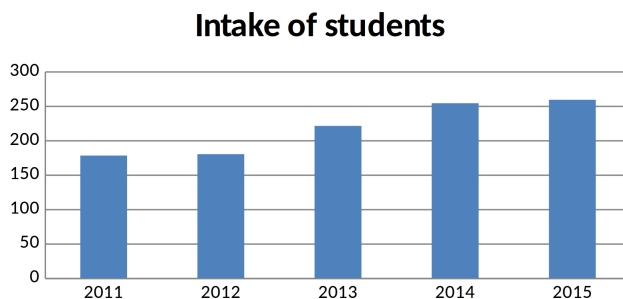


Figure 2.1: The development in student intake.

**Student population**
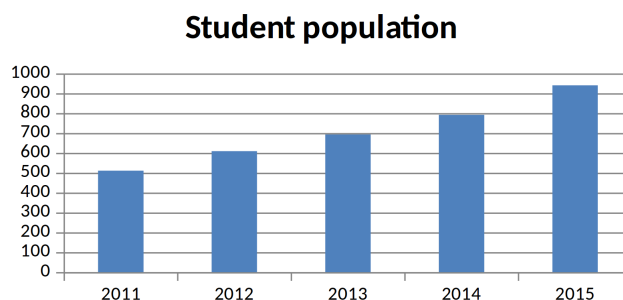


Figure 2.2: The development in student population.

the intake for the first time, and in the evaluation period the increase has continued, as depicted in Figure 2.1. Numbers for computer science and informatics have been quite stable, whereas there has been an increase for both software engineering and the bachelor program in information technology. In 2014 the most recent program in interaction design was introduced. In the period the yearly intake has increased 46% from 178 in 2011 to 259 in 2015.

With the increasing intake the total student population has grown as shown in Figure 2.2. Numbers have gone from 511 in 2011 to 941 in 2015, a growth of 84%. The growth is expected to continue over the next three to five years, before we reach a new state of equilibrium where the annual intake equals the number of students leaving, preferably as graduates.

The allocation of teaching is administered in an accounting system based on hours, and over the evaluation period the teaching delivered by the depart-

**Internal funding (1000 kr)**
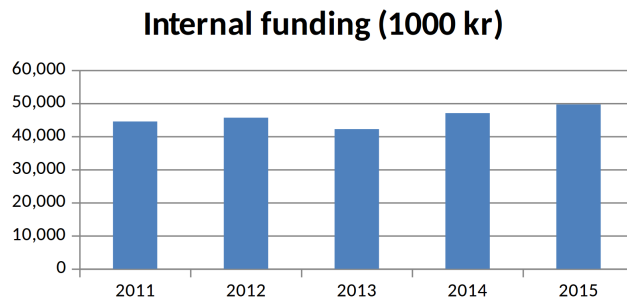


Figure 2.3: The development of internal funds.

ment has increased 23%. Thus we have been able to cope with an increase in student population of 84% with an increase in delivered teaching of 23%. This is a substantial gain in efficiency.

### 2.2.2 Funding

As mentioned earlier the funding of the department's activities is composed of internal funds from the faculty and external funds from research councils etc. The development of internal funding can be seen in Figure 2.3. As can be seen the funding decreased in 2013 despite the increasing student population. This was a result of a round of cost-cutting where all internal grants were trimmed.

Fortunately, we had a reserve of external funding that could be used to mitigate the immediate impacts of the budget cutbacks, and that gave us some time to adjust to the new situation. Thanks to this we avoided dismissals of staff, which became necessary in other departments. The effect of this was that the turnover of external grants peaked in 2013 as illustrated in Figure 2.4.

The economic reductions resulted in growing teaching duties, and the teaching norm was increased by approximately 10% for tenured staff, leaving less time for research. This is a substantial reduction of resources that will inevitably influence research production although it is not yet manifested in the figures for 2014-2015. As it is seen in Figure 2.5, the total turnover dropped slightly in 2014 while 2015 shows signs of a small recovery in absolute numbers.

**External funding (1000 kr)**
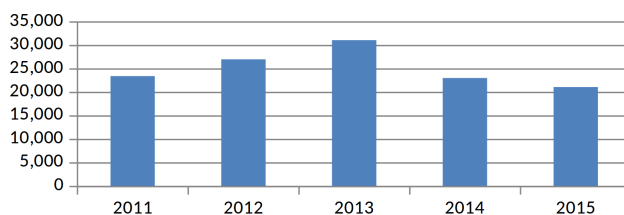


Figure 2.4: The development in external funding (spent).

**Total turnover (1000 kr)**
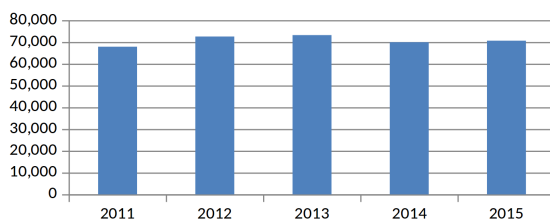


Figure 2.5: The development in total turnover (internal and external funds).

### 2.2.3 Staff

The reduced economy had effects on e.g. teaching load and was also manifested in other figures. The announcement of new positions was suspended and the immediate effect is clear in the enrollment of PhD's as shown in Figure 2.6. As it is also seen we are getting back in the neighborhood of our goal of twelve new positions per year.

A similar pattern is found for most staff categories as shown in Figure 2.7. The general picture is a slight decline in the middle of the period followed by a return to the earlier level. The exception is the support staff. This is partly due to a tightening in the sense that positions have not been refilled when people have left, and partly because technicians prviously directly employed by the department were re-assigned to a newly formed central IT organization of the university.

22

**Ph.D. enrolment**
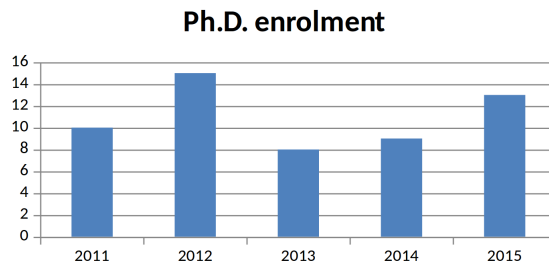


Figure 2.6: Number of new Ph.D. positions.
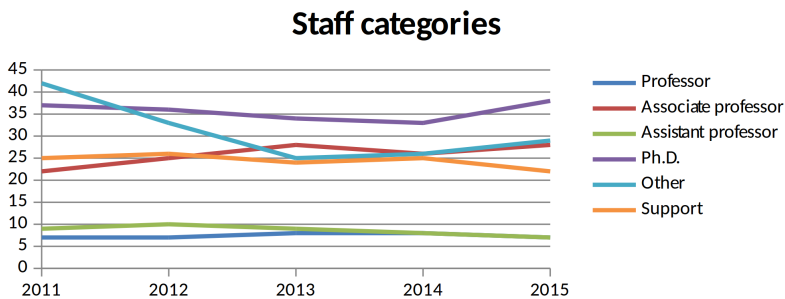
**Staff categories**



Figure 2.7: The number of employees in various staff categories.

### 2.2.4 Publications

The internal funding is partly based on scientific publications as measured by the Danish bibliometric research indicator (BFI). The indicator assigns points to publications in recognized outlets, where scientific publication channels are ranked according to reputation. Conference articles are given one to three points and journal articles between three and five points, and for each point the department is granted a fixed amount (currently 16.000 DKr.) per point. The points obtained by publications from the department in the last five years are shown in Figure 2.8. The BFI system has been controversial in the sense that it formed a partial basis for distribution of research funds, without taking differences in research areas and publication traditions into account. For our area in particular, the acknowledgement of full papers at conferences has been crucial in order to obtain a fair and accurate frame of reference. This issue was raised on several occasions by the bibliometric work
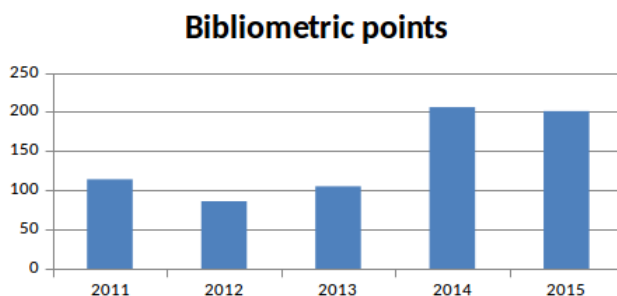
Figure 2.8: Points obtained in the Danish bibliometric system.

group in computer science, and under the chairmanship of professor Torben B. Pedersen the group succeeded with the full inclusion of conferences in the BFI system. This change took effect from 2013 and although the indicator thereby became more accurate within computer science, it is still questionable, whether it gives a valid reflection of publication traditions in different areas.

### 2.2.5 H-indices

The h-index has gained popularity over the past decade as an indicator for scientific impact. Jens Palsberg at UCLA maintains a list with the most influential computer scientists (that is, those with an h-index greater than or equal to 40) at his website. In the update from October 10th 2015 we recognize Kim G. Larsen and Christian S. Jensen as the two highest ranking professors affiliated to Danish universities.

Bibliometric analyses at the institutional level are based on Scopus and Web of Science. These indices may yield a true picture of many areas, and they are widely accepted within engineering, for example. However, for computer science a more accurate description is obtained by using Google Scholar. This fact is acknowledged by our faculty, but has not yet been implemented. To push forward the issue, we have decided that all tenured staff maintain a Google Scholar profile, and it is currently being considered if a national benchmarking of the Danish computer science departments could be based on these profiles. The h-index profile for the department in terms of counts for tenured staff is shown in Figure 2.9.
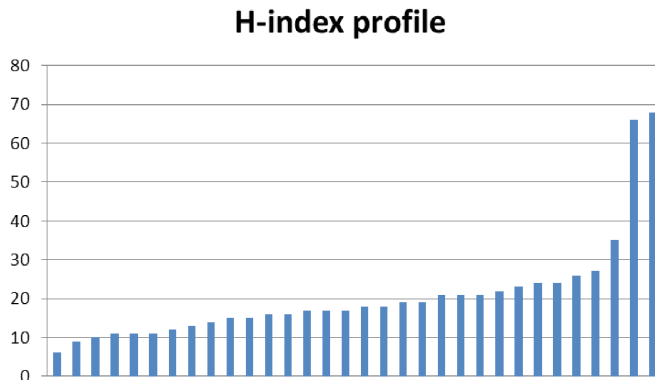
Figure 2.9: H-indices for tenured staff in the autumn 2015.

If we compare the profile to the one displayed five years ago we see a similar pattern but at a higher level. This is, of course, partly because the index is accumulative, but as some renewal of personnel occurs it is worth noticing that there is a general high impact of the department's researchers. Five years ago we had four persons with an h-index above 20 and although one of them has retired we now have twelve individuals with an index greater than 20. Furthermore, the peak is now well above 50 for the first time and in the other end the tendency is that applicants have substantial merit in terms of publication impact.

## 2.3 Achievements

At a number of occasions members of the staff have been asked to fill special assignments, have been granted prestigious awards or have attracted peer recognition in other respects.

**Advanced grant from the European Research Council**

Over the last decades the department has grown to become an important player in Europe and worldwide. As a culmination of this continuing evolution Professor Kim G. Larsen received one of the prestigious advanced ERC grants in 2015 for the project LASSO: Learning, analysis, synthesis and opti-

mization of cyber physical systems. Besides the recognition among peers for his outstanding contributions, the honor was accompanied by a substantial donation for further development of methods and tools within the fascinating field of cyber physical systems.

## Awards and honors

In 2011 professor Christian S. Jensen was selected for the Villum Kann Rasmussen Annual Award for Technical and Scientific Research – the largest Danish individual award for technical and scientific research.

In 2013 professor Kim G. Larsen received the CAV Award (Computer-Aided Verification) at the conference in St. Petersburg "for developing UPPAAL which is the foremost tool suite for the automated analysis and verification of real-time systems."

Professor Christian S. Jensen was appointed ACM fellow professor and Torben B. Pedersen was recognized a Distinguished Scientist by ACM.

Professor Anders P. Ravn joined the Danish Royal Order of Dannebrog for his lifelong contribution to science and society.

Several awards for best papers have been achieved in the period and, quite remarkable, the price for most influential paper at the conference MobileHCI, that is given to the paper that after ten years have received most citations, has been won in two successive years by professor Jesper Kjeldskov – the second time with associate professor Mikael Skov as co-author.

## Serving society and the scientific community

Professors Christian S. Jensen and Kim G. Larsen serve as members of Academia Europaea, The Academy of Europe and they were also appointed "Digital Sages" by the Danish Academy of the Technical Sciences.

Professor Torben B. Pedersen is also a member of the Danish Academy of the Technical Sciences and he has been elected as chairman for the new theme Digitalization and Big Data.

Professor Kim Guldstrand Larsen was reelected as Danish National Expert for the ICT Program Committee of EU.

Professor Christian S. Jensen has been appointed as president of the steering committee of the "Big Data" program of the Swiss National Science Foundation.

**Academic highlights**

Professor Jesper Kjeldskov and associate professor Radu Mardare successfully defended their doctoral thesis and obtained the Danish Doctor Degree, the highest academic recognition of a substantial contribution within a specific area.

A number of substantial external grants have been obtained in the period. We were fortunate to receive an Obel professorate that ensured Christian S. Jensen's return to the department and a Velux Visiting Professorship ensured a continued affiliation of Axel Legay. Other significant donations include a continuation of the Danish-Chinese basic research center and a research council grant from the YDUN program targeted at talented female researchers. With an acceptance rate of only 6% it is remarkable that one such grant was obtained by our, at that time, assistant professor Katja Hose.

## 2.4  Opportunities and Challenges

### 2.4.1 Collaboration and funding

The collaboration between units has been an issue of particular concern in earlier research evaluations, as an unexploited potential was seen for achieving new goals and insights by combining the existing strengths. The tendency that funding opportunities have become increasingly application oriented rather than focused on fundamental academic issues has enhanced such enterprises. The range of research topics in the department form a natural value chain: embedded systems communicating through the internet collect massive amounts of data - spread in time and space - that have to be stored and analyzed before the extracted results are presented. Interaction with the environment and with users become increasingly sophisticated with new devices and improved technologies. This somewhat simplified value chain has formed the basis for a number of successful research projects that have been

realized in collaboration between units and with other departments. The interplay between basic research activities and applications is a hallmark of the department's research. Basic research goes hand in hand with application of results in order to create value for the society and for companies. This effect of added value is illustrated by a number of contracts with private and public customers that subscribe to extractions and analysis of traffic data. Prominent examples of such collaborative projects are Totalflex aiming at optimizing energy consumption and distribution, and the Center Data-Intensive Cyber-Physical Systems, DiCyPS, which seek to establish generic solutions for the smart society, in particular by contributing user-friendly systems within the energy and the transport sectors. Different levels of collaborations may be identified in several of the projects to be describes in the following chapters. The world is in motion and we will have to anticipate future possibilities for growth and development. With an expectation of a public economy under increased pressure we will have to exploit the opportunities for external funding to the limit. The recent tendency to favor application oriented research challenges our ability to put our research activities into perspective and establish the connection to societal challenges. This direction has been successful in recent years and the value chain sketched above seems to generalize to other areas, simultaneously contributing to both basic knowledge and addition of value to society. The challenges will be to involve more (younger) people in the application process and to form strategic alliances through creation of strong consortia covering both computer science and application area expertise. With the recent success in attracting external funding we are well prepared for the near future. We currently have external funds roughly corresponding to four year's turnover, but with continued cuts in national funding we will have to be aware of all possibilities and opportunities for future funding.

## 2.4.2 Ph.D. education and recruitment

The enrolment of Ph.D. students is closely related to the external funding as Ph.D. students in Denmark are employees rather than students. As can be seen in Figure 2.6 the number of enrolled Ph.D.'s dropped in the middle of the evaluation period mainly due to the limitations in economic latitude. An important contribution to maintain the activity at a sustainable level is ob-

tained through professor Torben B. Pedersen's partnership in the consortium Erasmus Mundus Joint Doctorate In Information Technologies for Business Intelligence – Doctoral College (IT4BI-DC). This prestigious Ph.D. program was established in collaboration with leading universities in the area: Université Libre de Bruxelles (Belgium), Technische Universität Dresden (Germany), Universitat Politècnica de Catalunya (Spain), and Poznan University of Technology (Poland). With this program and the recent success in obtaining external funding the number of enrolled Ph.D.'s is expected to resume a level of at least 12 enrolled students per year. An important recruitment basis for Ph.D.'s in the past were elite programs in embedded systems and in data-intensive systems. In these programs talented students were under the supervision of full professors, they were challenged in their project work by actual research problems, and they could participate in Ph.D.-courses. Unfortunately, the university's elite student programs were terminated by the board in 2013 and a replacement has not yet been devised. We consider such programs important, offering an opportunity for talented students to become acquainted with an academic career. Another possibility for a more gradual transition between Master and Ph.D. studies is the so called 4 + 4 arrangements where students are enrolled as Ph.D. students earlier than usual. Initiatives to gain experience with such endeavors have been established, but we have not yet had the opportunity to exploit them.

### 2.4.3 Continuity and change

We have always considered ourselves a young department, and the first member of our staff retired only a few years ago. This raises the question of continuation and possible organizational changes. In Figure 2.10 the age profile of tenured staff is shown. Despite minor variation there seems to be a reasonable smooth distribution. In a perfect world there is a gradual transition where the younger takes over from those who leave, but in an academic setting this is not always the case.

A combination of retirement, resignations and shortage of new positions has left the Machine Intelligence unit in a vulnerable state, where the current staff is heavily loaded with teaching and administrative duties. This structural imbalance is a challenge that will have to be resolved and we have just started looking into possible solutions.

**Age profile**
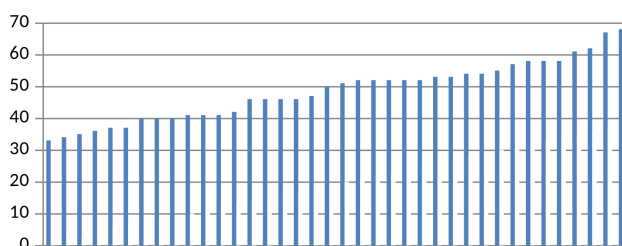


Figure 2.10:   Ages of tenured staff in the autumn 2015.

Another potential imbalance is introduced by an uneven teaching load between units. With new educations there is a need for particular teaching resources. We have a challenge in ensuring a development where the balance between education and research is maintained.

### 2.4.4 Collaboration and knowledge transfer

Computer science is an experimental and constructive science and the development of programs and tools is an important part of our merit. Unfortunately, our need for assistance in terms of technicians and programmers are not recognized as e.g. laboratory technicians are in a biochemical setting. Specialized software tools such as UPPAAL and HUGIN that are in widespread use throughout the world are rooted in research from our department. Such system's way into the future practice of institutions and companies could benefit substantially from practical oriented support for researchers by designated technicians and programmers. This could be helped by ranking computer science alongside more traditional engineering. Knowledge transfer and collaboration with industry are important areas, and we have launched a number of initiatives over the years. With the emergence of new organizations and networks the department's activities have evolved from a common centralized initiative to distributed activities. The national innovation network InfinIT is hosted at the department and is collaborating with the regional triple helix partnership BrainsBusiness. Most general activities are organized through these organizations, while specific collaborations on research and development are integrated in the activities of the research

units. Activities cover the full range from minor specific development activities for local companies to full range international collaborations in large research projects and basic research centers. A so-called Matchmaker helps companies to get in touch with the relevant researchers.

## 2.5 Committee evaluation: The University, Faculty, and Department Levels

**Observations**

Digitalization is one of the most important and pervasive changes in society today. It has implications for all aspects of the Department's activities. There will be a growing need for graduates with digital skills – a number of reports have indicated that just in the US there is a need for an extra 1-2 million data scientists by 2018 to meet that country's needs, while Denmark already today needs 3000 extra computer scientists and software engineers. In turn, this has implications for the educational processes but also brings new research challenges to the fore: for example Big Data/Data Science; Cyber Physical interaction and Systems; Sensor Technologies; Autonomy; Smart Societies; etc... Many of these areas require contributions from multiple disciplines but have Computer Science at their core.

Aalborg University is well-placed to benefit from this trend. It has a worldwide reputation for its innovative pedagogy based on problem-based learning (PBL). Coupled with this, the university has some world-class activities in the Computer Science Department and elsewhere which are ready to make substantial contributions. To realize the benefits of digitalization will require true multi-disciplinary efforts. The new Strategy document provides a framework to support these developments.

The Computer Science Department has two world-class groups and excellent staff in all groups. The Danish IT benchmarking exercise of 2014 shows that the Department is the best department in Denmark for number of refereed publications and BFI points per full-time faculty member. The Department is also top in a number of other metrics. During the review it was also reported that Aalborg Computer Science graduates are highly prized by industry. The Department thus deserves to be ranked even higher in the QS World Uni-

versity Rankings by Subject or the Academic Ranking of World University (ARWU "Shanghai") Subject ranking. The current rankings are to a large degree caused by the poor coverage of computer science publications in the commercial bibliometric indices used in these rankings (WoS, Scopus). Here, Google Scholar provides a much better coverage. However, the Department clearly has the potential to rise considerably in these rankings but will require support from the Faculty and University to achieve this.

We have reported in detail on each of the research groups and made recommendations for their continued evolution. We have also made recommendations that should be addressed at the university, faculty, and departmental level. We believe that the Department has very significant, but as yet unrealized potential and it is important that the Faculty and University actively support the continued development of this excellent activity.

**University and Faculty Level Recommendations**

At the University and Faculty levels, the following actions are recommended:

**Expansion:** The department has a potential to rise in international reputation and rankings, but it is difficult to imagine that this can be done without a significant growth in the size of the Department to better reflect the popularity of its education programmes and to allow it to develop its research profile. There are economies of scale and it is no accident that many of the world leading Computer Science departments are at least twice the size of the current Department. The Department has the building blocks to create a leading department and the trends identified earlier should support the necessary argumentation to justify expansion. It is thus recommended that the University and Faculty supports this expansion.

**Data Science and Data-Driven Science:** All sciences and scientific work as such is changing dramatically. This happens both in the "wet sciences" (technical, natural, and medical) but also in the "dry sciences" where Digital Humanities and Computational Social Science alter the fundamentals. Underlying this is the key notion of data-driven scientific discovery, also known as the "4th paradigm" of science, which means

that all areas require new skills (data science). It is paramount that any ambitious university is at the forefront of this change. Thus, it is recommended to immediately start one or more initiatives in data science and/or data-driven science. Options include both a new interdisciplinary education programme in data science and new educational elements bringing data science skills into existing education programmes.

**Digitalization of PBL:** The new university strategy identifies actions to bring IT to PBL. However, the digitalization aspects could be even more emphasized. In fact, AAU has a unique opportunity to transform PBL so that IT is not merely one tool among many, but instead the nexus around which PBL evolves. Thus, it is recommended to start a significant multi-disciplinary initiative on digitalization of PBL with Computer Science at the core.

### Department Level Recommendations

At the Department level, there are a number of actions that can be taken to ensure that it is able to play a full role in these developments:

**Strategy:** Strategy is important and needs to come both from top down and bottom up. For example, if the Department is to be more than the sum of its parts, there should be a top-down discussion about whether the Department should approach the Faculty to create a new center of excellence to complement the existing strengths.

**Organization:** The organizational structure needs to support both the research and education; the research groups should recognize that they gain from being partners in a larger organizational unit. Some of the structural issues which have arisen from historical alignments could be addressed by separating management and scientific organization, e.g., by a matrix structure.

**Leadership:** This follows from the previous point. Some competition between groups is inevitable and even healthy but at the Department level cooperation and shared vision is important and beneficial for all.

**Funding:** To address some of the worries about the funding opportunities for the future, there could be more coordination and support at Departmental level of the funding application process.

**Recruitment:** Whilst it is important to only appoint excellent candidates, a Departmental strategy should support a more strategic approach to recruitment and expansion of the department. It might be necessary to actively headhunt for positions in specific areas.

**Expansion:** All groups said that they would like to expand. The Department should also consider approaching the Faculty/University to create a new center of excellence.

**Relationship education-research:** The Department should clarify its strategic view on the relationship between education and research. The best education programmes often draw on the research strengths of a department. Education is important both in serving societal and industrial needs but also in preparing the next generation of researchers. If staff see education as a burden, the Department should take active measures to draw on best practice from elsewhere to relieve the burden as far as possible.

**Gender issues:** The gender (im)balance in Computer Science is a worldwide problem but there are active measures that could be taken to try to improve the situation. The Department seems to be ignoring this issue maybe assessing it as "too difficult". There are good practices (for example the UK's Athena SWAN programme) which may help. The Department is recommended to explore (partial) solutions which have been adopted elsewhere.

**Digitalization:** Higher education is not immune to the effects of the digitalization trend that has been described above. The adoption of new technologies in support of the educations offered in the Department may be part of the solution to relieving the "burden" discussed above.

**Sabbaticals:** Sabbaticals allow individual staff to refresh their skills and gain from new experiences but can also benefit the Department in a number of ways:

- The Department profile is raised by having staff on secondment in other institutions.

- The Department has the opportunity to develop non-sabbatical staff through assignment of duties vacated by staff on sabbatical.

- The Department may have more opportunity to host in-coming visitors as a result of staff being away from Aalborg.

# Programming Technologies
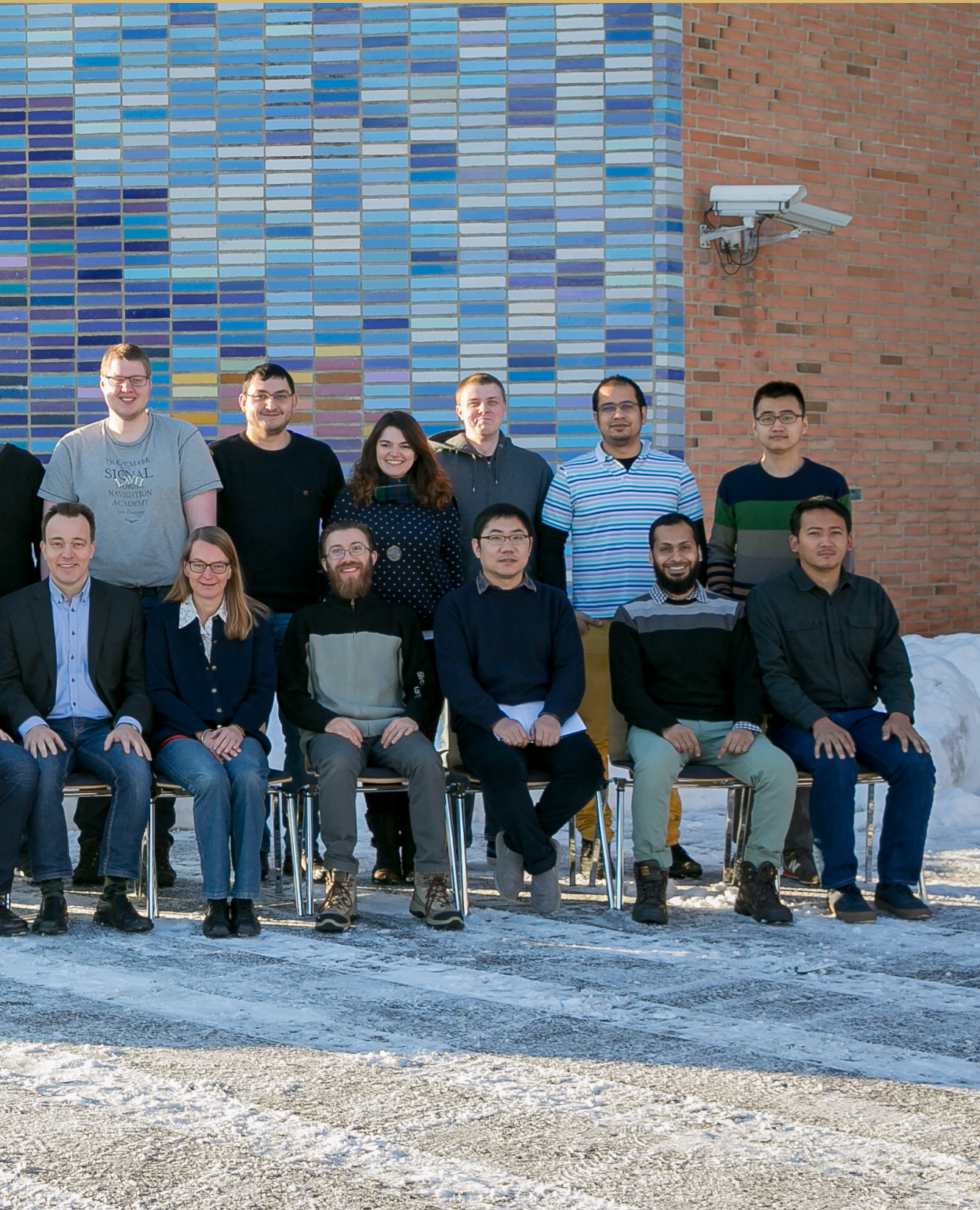
# 3 Database and Programming Technologies

## 3.1 Executive Summary

The Database and Programming Technologies (DPT) unit conducts research in data management, data mining, and programming languages, environments, and tools. The unit hosts Center for Data-intensive Systems, Daisy, that embraces all activities concerning data-intensive systems. Daisy aims to be an internationally leading research center within its areas of specialization.

During the reporting period, the unit has performed very well according to key performance indicators, including publications, external funding, Ph.D. education, scientific service, and industrial collaboration and dissemination.

The unit has published 213 refereed journal and conference papers in the period of which 56 are in the very top outlets within data management and a further 57 in top specialized outlets relevant to Daisy.

The unit has participated in 32 externally funded projects, with some DKK 62 million in funding from a wide range of sources, in the period. Most projects have substantial industrial participation, and many involve other groups in the department or the university. One spinout company has been established. Several software packages and tools have been released as Open Source.

The unit has actively educated new researchers, with 13 Ph.D. degrees having been awarded in the period, and some two dozen Ph.D. projects in progress.

DPT researchers have received substantial peer recognition and have performed extensive service to the scientific community. The unit has been successful in attracting top scholars for its positions in data management and data mining, but has been less successful in programming technology.

In conclusion, the unit believes it has been very successful research-wise in the

reporting period. In particular, the unit believes that its publication performance in top outlets and level of industrial collaboration compare favorably with those of similar groups across Europe.

## 3.2 Profile

The research in the DPT unit covers a wide range of areas within data-intensive systems and programming technology.

The unit's research approach is primarily constructive: theoretically well-founded, purposeful artifacts such as frameworks, data structures, indexes, algorithms, languages, tools, and systems are prototyped and subjected to empirical study. Thus, the research often involves labor-intensive software development and experimental studies.

Further, the research is mostly driven by (perceived) real-world applications, the objective being to provide solutions that enable new or better applications. The primary application areas include embedded systems, intelligent transport systems, (mobile) web querying, logistics, energy grids, and healthcare.

In data-intensive systems, substantial research concerns aspects of what is now termed big data. Central research topics belong to three areas:

- **temporal, spatial, spatio-temporal, spatio-textual and mobile data management** Increased activity in this area is fueled by the increasing ubiquity of geo-positioning and mobile devices, notably smartphones, and by the proliferation of geo-referenced content, including GPS data and a variety of web content such as business directory entries and micro-blog posts. Focus is predominantly on advanced indexing and query processing. The unit has been active in this area for several decades. Almost half of the publications during the reporting period fall within this area.

- **business intelligence, data analytics, data warehousing, data integration, OLAP, multidimensional databases, data mining, cloud data management** Research in this area is in large measure motivated by the explosive growth in available data and the increasing focus on extracting value from data. Research in this area generally adopts multidimensional models of data and often aims to offer efficient

support for novel functionality. Activities in this are growing rapidly and account for an increasing part of the unit's activities.

- **semantic and web data management, Linked Open Data** This line of research is motivated by new web infrastructure for data management and by models of data that aim to meet the needs for web-based data and data management. This area of research is relatively new in the unit and has become active due to hiring and new project activities.

Within the above topics, the research covers modeling and database design, data models, query processing, indexing, and applications.

In programming, the research concerns general-purpose programming languages as well as special-purpose languages, e.g., for embedded, real-time, mobile, parallel, and distributed systems. Studies also cover environments and tool support for program development and analysis. Central research areas cover integration of functional and object oriented programming, Programmatic program construction, Java for hard real-time embedded systems and applications in the area of indoor positioning. Research on teaching programming now also forms part of the research interest within this area.

## 3.3  Staff

### 3.3.1 Current Staff

**Professors:** Christian S. Jensen, Torben Bach Pedersen.

**Associate Professors:** Katja Hose, Hua Lu, Kurt Nørmark, Simonas Saltenis, Bent Thomsen, Christian Thomsen, Lone Leth Thomsen, Kristian Torp, Bin Yang.

**Assistant Professors:** Thomas Bøgholm, Xike Xie.

**Post-docs:** Chenjuan Guo, Benjamin Bjerre Krogh, Laurynas Siksnys.

**Ph.D. students:** Tanvir Ahmed, Ove Andersen (e), Asif Iqbal Baba, Gaston Bakkalian (e), Johannes Lindhart Borresen, Davide Frazetto, Bhuvan Gummidi, Nurefsan Gur, Kai Hermann (e), Jilin Hu, Dilshod Ibragimov (e), Kim Ahlstrøm Jakobsen, Søren Kejser Jensen, Dalia Kaulakiene, Ilkcan Keles, Kasun Perera (e), Rudra Pratap Deb Nath, Bijay Neupane, Faisal Moen

Orakzai (e), Muhammad Aamir Saleem, Saulius Samulevicius, Lawan Subba, Emmanouil Valsomatzis, Jovan Varga (e).

The Ph.D. students labeled with an "(e)" are enrolled at AAU, but *externally* employed by partner universities or, in one case, a private company.

### 3.3.2 Staff Development

Over the period, the staff evolved substantially around a stable core.

In 2013, Christian S. Jensen returned from Aarhus University to take up an Obel Professorship in 2013, and Christian Thomsen was promoted to Associate Professor. In 2015, Katja Hose and Bin Yang, who had joined earlier in the period, were promoted to Associate Professor.

A number of assistant professors and postdocs joined the unit and left it during the period. Ralf Rantzau joined as Assistant Professor in 2010 and left in 2011 for a position at Cisco Systems. Rene Hansen joined the unit as a postdoc in 2011 (after being a Ph.D. student and a research assistant) and left for Systematic in 2014. Yoann Pitarch joined the unit as a postdoc in 2011 and left in 2012 for University of Lyon. Mohamed Khalefa joined the unit as a postdoc in 2011 and left in 2012 for Alexandria University. Shuo Shang joined the unit as Research Assistant Professor in 2012 and left in 2013 for China University of Petroleum-Beijing.

A large number of Ph.D. students and research assistants pursuing Ph.D. degrees were employed during the period, but are no longer with the unit: Sari Haj Hussein, Jinling Jiang, Simon Kongshøj, Xiufeng Liu, Kasper Søe Luckow, Darius Sidlauskas, Kostas Tzoumas, Carmen Ruiz Vicente, Dingming Wu. In addition, Edwin Lewis-Kelham, Claus Pedersen, Dovydas Sabonis, and Chaithanya Vadhri were employed as research assistants.

We also note that we hosted a number of visitors during the period, including Ph.D. students that were supervised by members of the unit during their visits.

## 3.4 Goals 2011–15

The detailed plan for the evaluation period may be found in the report on the previous evaluation period. The unit targeted primarily a use-inspired approach to research, where research decisions were to be guided by concerns for use in practice or anticipated practice. In line with this principle, we aimed to interact with practitioners on a regular basis, e.g., via participation in advisory boards or boards of directors for technology companies, via collaboration with practitioners in funded research projects, via participation in industry associations, and via lecturing to industrial audiences.

In order to achieve focus and synergy among activities, the unit chose to focus on a few advanced key application areas, namely mobile services, intelligent transport systems, energy, and logistics. Specifically, an umbrella research area with three focus areas was defined.

### 3.4.1 Main Area: Data-Intensive Services - Mobile, Ubiquitous, Cloud, and Beyond

The main area was inspired by the increasing proliferation of IT services. These services are increasingly data-intensive, requiring the processing of ever larger and more complex data sets in order to provide the most optimized service. The services span an increasingly complex sphere, with some used on mobile devices and others used in ubiquitous settings, with some used by persons and others used by computers, with an increasing interplay among many diverse services, and increasingly based on cloud computing. As part of the overall theme, the unit planned to develop a number of building blocks for data-intensive systems in a cloud setting. This was to be underpinned by research on algorithms, data structures, and query processing techniques capable of exploiting the inherent parallelism found in new hardware architectures such as multi-core CPUs, as well as in cloud computing environments.

### 3.4.2 Focus Area: Cloud Intelligence

In business intelligence (BI) and data warehousing, the main focus was to be cloud intelligence, which emerges from the migration of BI and analytics technologies to a cloud environment. Many different types of data ex-

ist in the cloud, including relational, multidimensional, text, semantic web XML/RDF/OWL, geo-tagged, and sensor data, all of which should be analyzed in an integrated way. Our vision was to develop technologies that enable a cloud warehouse based on a novel kind of data model that supports both multidimensional concepts and the flexibility seen in semi-structured models. Concrete planned activities included spatio-temporal data warehousing and data mining (prediction and forecasting) for traffic services and mobile phone networks, investigating the predictive power of sentinels and cloud deployment; effective compressed bitmaps indices; techniques for the warehousing of semantic web data; real-time data warehousing; an extract-transform-load framework for cloud computing; and architectures and query processing techniques for massively distributed large-scale data warehouses with tight integration of historical, streaming, and forecast data.

### 3.4.3 Focus Area: Spatio-Temporal Data Management and Mobile Services

The unit planned to build on its strengths in spatio-temporal data management, in particular in spatial and spatio-temporal indexing and query processing, exploring important trade-offs, such as how query result accuracy or freshness depends on update costs. In addition to experimental studies, deeper theoretical understanding and cost modeling of workload-dependent performance were targeted.

The plans also included studies that aimed to support more realistic, application-specific assumptions, such as studies of models involving location privacy and models for indoor and hybrid indoor/outdoor spaces. A data management foundation that integrates models of indoor and outdoor spaces was envisioned that was able to support Euclidean and network distances and that was able to accommodate both outdoor and indoor points of interest.

### 3.4.4 Focus Area: Umbrella Project in Programming Technology

Programming technology remains at the core of software development, and we found it important to have activities in this core area. We thus planned a new project intended as an umbrella activity to coordinate activities that would otherwise be carried out on an individual basis.

We planned to continue to take part in collaborative efforts, especially in mobile and embedded systems, but also anticipated a new direction in data-intensive applications such as business intelligence and eScience.

In the short term, we anticipated engaging with (local) industry to help them understand the shift in software construction from the object-oriented approach to the more declarative and dynamic approach being offered by new language features. This was to occur through participation in projects such as DaIn, CaIn, and InfinIT. We found it important to raise the profile of the (sub)unit. To give focus to this endeavor, we planned to pursue a new direction called "programmatic programming," and we intended to research the transitions between programming with dynamic versus static types.

We planned to utilize research funds from the department and involve M.Sc. student projects and seek external funding for at least one Ph.D. student.

## 3.5  Activities and Results

In the following, we give an overview of the research activities and the results achieved in the evaluation period. We cover each of the three planned focus areas and then cover substantial research topics not anticipated in the plans.

### 3.5.1 Focus Area: Cloud Intelligence

**Extract-transform-load** (ETL) continued to be an active research topic, resulting in around half a dozen papers in the reporting period. The framework pygrametl for Python-based programming of ETL processes, created in the previous reporting period, was extended with parallelism (both SMP and cloud-based) and more functionality and matured due to user feedback. First, pygrametl was ported to the Disco MapReduce environment to enable support for processing of fact tables and different kinds of dimension tables in star and snowflake schemas without the user has to understand technical details about the parallelism. Dimension processing techniques for different scenarios were also proposed.

Later, a Java-based ETL framework was created for Hive on Hadoop, both of which are Java based. This work, among other challenges, investigated how to bring support for slowly changing dimensions to Hive (that did not support

44

UPDATEs) and how to do efficient lookups in big dimension tables during fact processing. This made it possible to define ETL processes by means of high-level constructs without worrying about technical MapReduce details. The tool showed to be efficient and effective: For example, it was nearly 4 times faster than Hive in processing a slowly changing dimension and needed far fewer statements in the code.

**Cloud data management** was investigated. Here, we studied how to achieve the fastest or cheapest executions of analytical data processing jobs on virtual machines acquired from Amazon's spot market of leftover instances. Due to the many possibilities and varying prices, it is hard for users to make the best choices. The proposed SpotAdapt framework estimates the runtime from small sample runs and uses this to make near-optimal suggestions for either the cheapest or fastest deployment. Work has also been done on ETL processing on MapReduce as described earlier. The International Workshop on Cloud Intelligence was organized twice resulting in two panel papers and two workshop reports that point to important research directions in the field.

**Predictive analytics** was another prominent research topic, resulting in around a dozen papers. One topic was how to mine so-called *sentinels*, which are causal relationships where changes in one or more source measures are followed by changes to a target measure within a given time period, We optimized sentinel mining using bitmap encodings, specific CPU instructions, and multi-cores to be two orders of magnitude faster. Integration into the TARGIT BI Suite was also considered.

Another topic was context-based spatio-temporal prediction with applications in reducing the energy consumption in mobile broadband networks (MBNs). A number of increasingly powerful prediction techniques were proposed for context-based spatio-temporal time series, including hierarchical prediction contexts (temporal, spatial, ...), dynamic optimal context selection, concept drift, rule-based methods, and using knowledge of planned events. Overall, prediction results improved the state of the art substantially. Motivated by applications in traffic management, we considered mining streams of moving object positions in road networks to find current evolving trajectories, incrementally mine closed frequent routes, and predict near-future locations and densities based on the evolving trajectories and frequent routes.

Finally, motivated by large time series of energy supply and demand, we

developed several techniques and systems based on two main design principles. First, as data has significant (self-)similarities over different periods and time series producers, higher-level models can efficiently approximate the data. the time series, both storage and query performance-wise. Second, data is often forecastedand many applications query historical and forecasted data together, so the same kind of models can be used for forecasted and historical data. storage. This enabled 1–2 orders of magnitude better performance.

### 3.5.2 Focus Area: Spatio-Temporal Data Management and Mobile Services

This focus area has been a key research topic throughout the evaluation period and is documented in a large number of papers. The research conducted within this area is motivated by a multitude of available real data in both outdoor and indoor contexts, spanning the spectrum from theoretical studies of novel ideas to applied research documented in prototype systems.

**Transportation related data management** forms a strong line of research, with more than 60 papers being published in the reporting period, that is generally motivated by the rapidly growing volumes of available data, including GPS data, Bluetooth positioning data, and fuel consumption data from vehicles. Many studies have been conducted that use GPS data, with specific topics such as these: (i) tracking of moving objects, (ii) trajectory data indexing, (iii) travel-time computation with complex weights, (iv) query processing for outdoor LBS, (v) eco-routing and eco-driving, and (vi) skyline queries.

GPS data has been used as the data foundation for tracking moving objects and for studies on how to ensure that large sets of moving objects can be queried and updated simultaneously and efficiently. High-frequency GPS data can be converted to trajectory data that has served as the foundation for a wide range of road-traffic analysis tasks, such as studies aiming to enable green waves and studies of turn times at intersections.

A core aspect of road transportation is travel time. Therefore, substantial efforts have focused on this aspect. During the evaluation period, increasingly advanced travel-time weights assigned to road-network graph edges have been invented with the goal of increasing the accuracy of travel-time estimates.

The combination of GPS data and fuel-consumption data has been used for

studies that aim to enable eco-routing and eco-driving. Because fuel costs are the second highest expense for fleet owners, the ability to compare both routes and drivers with respect to fuel consumption is important in practice and thus has been the subject of several studies. Finally, travel time and fuel consumption are independent aspects of traversals of routes in a road network. Thus, studies of stochastic skyline routes have been conducted that integrate these aspects.

**Advanced indexing** has been an ongoing research topic throughout the evaluation period, resulting in eight papers. In particular, building on the foundations established in the past, main-memory indexing of geographical positions of moving objects was explored. Main-memory, multi-threaded spatial indexes were developed that leverage the parallelism of modern processors to support high update rates inherent in applications where censors, such as GPS devices, track continually changing processes, such as the movements of vehicles. First, tree-based and grid-based main-memory indexes were compared experimentally. Next, a series of techniques were developed and compared to isolate concurrent updates and queries on spatial grid-based main-memory indexes. One of the developed approaches isolates the conflicting operations by allowing them to operate on different snapshots of data. Another approach exploits the specifics of the targeted application domain. Other indexing work focused on indexing for similarity search in generic metric spaces. Specifically practical, B-tree-based solutions were proposed.

**Indoor spatial awareness and indoor data management** was an increasingly prominent research topic in the unit during the evaluation period, fueled by several funded research projects. In the Indoor Spatial Awareness (ISA) project, we studied how to enable hybrid indoor positioning by exploiting both Wi-Fi and RFID such that improved positioning accuracy is achieved. We also investigated how to identify typical frequent movement patterns among indoor objects from their symbolic indoor tracking data. The results in the ISA project were published as two research papers in MDM conference.

In the projects BagTrack and NILTEK involving industrial partners specialized in airport baggage handling, we worked on innovating data management technologies for handling massive indoor RFID tracking data. In relation to RFID baggage data cleaning, we focused on removing false positives (cross readings) and on recovering false negatives (missing readings). We designed

graph-based approaches as well as a probabilistic model-based approach. In relation to warehousing and analysis of big RFID baggage data, we designed a warehouse solution and proposed techniques to mine hotspots and risk factors from RFID baggage tracking data. For continuous queries on RFID baggage data in a mix of indoor/outdoor spaces, we designed models and algorithms for reasoning about RFID-tracked moving objects, predicting their future locations and performing continuous queries on both observed and predicted positions. We also studied distance-aware spatial queries (range, $k$NN, and join) on indoor moving objects. In a more general context of indoor location-based services, we studied how to generate indoor space models that can support indoor navigation from digital building information (e.g., IFC files).

The research on indoor settings resulted in some two dozen publications in prestigious scientific outlets (e.g. ICDE and TKDE). The research delivered several prototype systems, some of which were demonstrated at international conferences (e.g., ICDE and SSTD). Furthermore, a number of techniques (e.g., data warehousing and data cleansing) were presented to industrial partners with the goal of achieving integration into their operational systems.

**Skyline queries** is another line of research that was carried out in the contexts of spatial and multidimensional data. Our research, published in seven papers in outlets including TKDE, ICDE, EDBT, and IS, contrasts previous work in several respects. First, we combine the dominance relationship that underlies skyline queries with spatial distances and locations, which enables novel spatial decision making and spatial object/location ranking that are not studied in the literature. Second, we study how to upgrade disadvantaged multidimensional data points (e.g., representing products with multiple attributes) economically such that those points enter the skyline. Third, we propose a flexible framework that is able to efficiently resolve arbitrary user-specified size constraints on skyline queries, whereas most previous research focuses on selecting representative points from the skyline. Fourth, we develop skyline algorithms in non-centralized computing settings, including distributed data sites, streaming environments, and MapReduce.

### 3.5.3 Focus Area: Umbrella Project in Programming Technology

Despite an exceptionally high teaching and administration load, the programming technology staff has produced 26 papers, four Ph.D.'s, and has been

involved in a number of projects, in industrial technology transfer, and in an attempt to create a spin-off company. In line with recommendations by the previous evaluation committee, we engaged in collaboration with the data management staff and with staff in the DES unit.

The long term research has focused on integration of object-oriented and declarative programming as well as on interactive and programmatic program construction. Research on educational issues has also been conducted.

In collaboration with data management and DES staff, the programming technology staff engaged in mobile applications, smart houses, and indoor positioning, through the SmartCampusAAU project, partly funded by the EU regional development fund and the Region of North Jutland, the Smart-Campus2.0 project, partly funded by CaIn, and the SmartCampus3.0 project, partly funded by InfinIT. The SmartCampus projects produced a number of demonstrators in collaboration with four industry partners: MapsPeople, AskCody, ConLan, and RTX. MapsPeople and AskCody developed products based on the results. We investigated the possibilities of setting up a spin-out company with AAU Innovation, but abandoned the plans when large players such as Cisco started delivering similar products. The SmartCampus software has been released as open source via GitHub and is now used in many universities. The software has also been made available to the Danish App Lab, set up at AAU CPH in collaboration with Microsoft and Nokia.

In the latter part of the reporting period, work was started on technologies and applications for big data analysis. A joint Database and Programming Technology master's course was offered in the autumn of 2014. In February 2015, the Popular Parallel Programming (P3) project, funded by the Danish Research Council, started. Its goal is to help spreadsheet users harness the power of multicore computers and enable big data analysis using higher order functional programming within spreadsheet applications.

A collaboration with staff from the DES unit and Via University College, on enabling real-time programming in Java, that started during the previous reporting period remained active. Three Ph.D. projects in this area were completed. A time predictable version of the Java Virtual Machine and two open source tools, TetaSARTS and SymRT, for analysis of timing properties of Java programs, written in the Safety Critical Java (emerging) standard, were developed. Both tools translate SCJ programs to networks of timed

automata suited for analysis by the UPPAAL model checker. This work was funded partly by the EU COST Action IC1001: Transactional Memories: Foundations, Algorithms, Tools and Applications (Euro-TM) and the EU COST Action IC0701 Formal Verification of Object-Oriented Software. In this area, we engaged extensively in technology transfer through the Center for Embedded Systems (CISS), the InfinIT working group on high level programming languages, and the InfinIT working group on Embedded Systems Engineering, working with companies such as Polycom (Kirk Telecom A/S), Wirtek A/S, Mechatronic Brick ApS, Aalborg Industries A/S, Prevas A/S, Blip Systems A/S, RTX Telecom A/S, Dore Development, Intelligent Systems A/S, Grundfos, Skov, Aalborg Industries, Attira, ConfiCore, Dansk Computing Centre A/S, and MVC-Data.

In addition to the successful project applications we have been involved in some unsuccessful ones. We coordinated the BIMPEL project application submitted for the EU FET Unconventional Computation (UCOMP) programme and the Rapid development of reliable software for multi-core embedded systems for the Danish Research Council for Technology and Production (FTP) programme. We collaborated on the DIACUT EU project proposal for Horizon2020, two COST actions and one further FTP application.

### 3.5.4 Unanticipated Research Directions

**Energy/flexibility data management** This area emerged as a major research topic during the period, resulting in some 15–20 papers. The work was motivated by the need to accommodate large amounts of renewable energy sources in the energy supply, which in turn necessitates that the energy demand can follow the supply, so-called demand response. Thus, flexibility becomes a key concept. Traditionally, flexibility has been modeled only implicitly (time-varying energy tariffs or price signals). We have instead modeled the flexibility explicitly, proposing the key concept of *flex-offers* which specify the inherent flexibilities in time, amount, and/or price across different time intervals explicitly. A range of topics were investigated based on the flex-offer concept. A number of different models and metrics were proposed for to measure and compare flexibilities, especially when combining, e.g., time and amount flexibility. We proposed several techniques for aggregating small individual flex-offers into larger ones as required by energy

scheduling and energy trading, including heuristics and techniques for keeping as much flexibility as possible. Other aspects considered were how to respect given constraints and priorities, e.g., in size or shape and how to balance demand and supply to respect grid constraints. Our techniques were shown to be efficient, scalable, and with good result quality. Flex-offer aggregation and scheduling techniques were integrated into a service-oriented architecture based on IoT technologies to enable a virtual market of energy. Further, techniques were developed for scheduling and visualizing flexibilities. The flex-offer technologies were employed in a number of case studies based on real data in residential homes, commercial buildings, electromobility, and industrial production. The flex-offer work led to the Best Poster Award at the World Smart Grid Forum 2013.

Another line of work considered fine-grained prediction and extraction of both energy demand and energy flexibilities at the device/appliance level. Our analyses showed that there was significant amounts of (predictable) flexibility for household devices. Based on these findings, we developed techniques for predicting device activation at time granularities varying from hours to a complete day. We also evaluated the (high) financial value of flexibility and how it varies for different scenarios.

**Semantic Web data management** emerged as a major research topic during the period, resulting in some two dozen papers. The primary underlying technologies are the Resource Description Framework (RDF), Linked Open Data, and the SPARQL query language. One subline of work followed our proposed vision of *Fusion Cubes* as a foundation for self-service business intelligence exploiting both internal and external (open) data. Here, we have worked on an extract-transform-load (ETL) framework for turning Open Data into RDF. Results were published in the Linked Open Data cloud and at a SPARQL endpoint. We proposed a significant extension of the QB4OLAP vocabulary and a tool for semi-automating the annotation of statistical Linked Data (QB) datasets with with multidimensional semantics. We also developed the CoDA approach for processing aggregate queries in a federation of SPARQL endpoints based on a cost-based optimizer. Moreover, we worked on optimizing RDF data cubes, mostly at the logical level.

Another subline of work aimed to query the multitude of knowledge bases available on the Web in an integrated and efficient way, while obtaining reli-

able results, adressing problems like knowledge base heterogeneity and data errors. In this context, we developed a scalable approach (AMIE) applying association rule mining in ontological knowledge bases, and used the obtained rules to deduce and add missing knowledge to a knowledge base. We developed FrameBase, a linked open knowledge base meant to tackle semantic heterogeneity among knowledge bases. It captures n-ary relationships and integrates knowledge from different large-scale knowledge bases. To enable scalable query processing, we have built a distributed engine (Partout) that partitions big RDF graphs into fragments, assigns the fragments to machines in a cluster, and efficiently executes SPARQL queries by exploiting parallel processing. We also proposed an extension of the dataflow language Pig Latin to handle Linked Data in RDF with the goal of bridging the gap between Pig and SPARQL for analytics.

**Social networks** accumulate huge amounts of user-generated data. The unit's research concerns mainly the topical, spatial, and temporal aspects of such data, the overall goal being to extract and deliver additional value to the users of such data. In the evaluation period, three papers have been published. The first paper focuses on location-related privacy issues in geo-social networks (GeoSNs) and formulates four privacy aspects that can lead to privacy leakage for users in GeoSNs. Possible solutions for these aspects are discussed in the paper. The second paper works on effective detection of stable and temporal topics in social network data. It proposes a unified user-temporal model and enhance the model with regularization and smoothing techniques, such that stable and temporal topics can be detected simultaneously. The third work makes use of social network data to search for local experts that can offer expertise relevant to user needs which are captured in keywords. To make the search effective and efficient, the work proposes a user ranking method and a scalable framework with search algorithms for massive user-generated social data.

**Spatio-textual data management** concerns primarily the efficient and effective querying of content, often termed spatial web objects, with associated geographical locations and textual descriptions, possibly in addition to other attributes like time. Examples of such content include geo-located microblog posts (e.g., tweets), web pages, and business directory entries.

This research is motivated by the increasing mobile use of the web, by increas-

ingly capable geo-positioning infrastructures, and by the increasing availability of spatio-textual content. Well beyond a billion keyword-based queries are issued on a daily basis that target such content.

A range of query functionality has been studied. Key functionality extends spatial $k$ nearest neighbor and text retrieval queries to enable the retrieval of spatial web objects that are both relevant to query keywords and are near a query location. Recent focus has been on supporting queries that retrieve groups of objects that satisfy both geographical conditions related to a query location and each others' locations and that are relevant to given query keywords. For example, one type of query returns groups of objects with combined textual descriptions that satisfy given query keywords and that are near each other and near query locations. Another type of query aims to support users who wish to visit several alternative objects and thus returns groups of objects such that each object in a group is relevant to the query keywords and such that the objects in a group are near each other and near a query location.

Queries that return sets of objects are often NP hard, and three types of algorithms are often proposed: Exact algorithms, approximation algorithms with provable approximation bounds, and algorithms that are shown empirically to provide results with good or very good accuracy efficiently in realistic settings.

Query result authentication and so-called why-not querying have also been studied in the context of spatial keyword queries. In addition, topics such as clustering, spatio-textual term querying, and topic modeling have been studied. The results are reported in a dozen and a half papers.

**Education-related research** The programming technology staff maintains a strong focus on teaching, including experimenting with new teaching methods, curricula design, and involvement in study board and school activities. These activities have led to five research publications and the participation in the project Future Education and Training in Computing: How to support learning at any time anywhere (FETCH). Research has been done on the use of video in the teaching of programming, on automating the handling of assignments, on problem based learning in software engineering curricula, and on the benefits and cost of the accreditation of university educations.

## 3.6 Own Evaluation

### 3.6.1 Research Areas

The general theme and goals described in the main area, data-intensive services – mobile, ubiquitous, cloud, and beyond, were addressed within the contexts of the three focus areas. Thus, the evaluation considers these three areas in turn. We end by assessing the research on unanticipated topics.

**Cloud Intelligence**   Looking back at the plans, the term "cloud intelligence" has served well as an umbrella for a wide range of activities. Following the general trend in the data management area during the period, the data model, storage, and query processing efforts moved away from "one-size-fits-all" solutions and instead focused on specialized solutions optimized for specific types of data and applications. The envisioned integration of past and future (forecasted) data was first demonstrated in the TimeTravel system and is, along with the envisioned work on massively distributed large-scale data warehouses of historical, streaming, and forecast data ongoing within the larger scope of efficient management of big sensor data.

The flexible platform for modeling, acquiring, storing, and querying complex multidimensional data was initiated according to the proposed *Fusion Cubes* vision and based on RDF and Linked Open Data. The envisioned activities within spatio-temporal data warehousing/mining, sentinel mining, compressed bitmap indices, real-time data warehousing, and cloud-based ETL framework(s) were all carried out more or less as planned, with good results and resulting in a number of publications in prestigious outlets. As planned, a spin-out was created for commercializing the Position List Word Aligned Hybrid (PLWAH) compressed bitmap index technology. Thus, we find that the research in the cloud intelligence area has been successful.

**Spatio-Temporal Data Management and Mobile Services**   The unit pursued most of the areas mentioned in the plan. The trade-offs between accuracy, freshness, and performance in the context of spatial indexing and query processing were explored in the setting of main-memory multi-core systems, reflecting the plans mentioned in the main (umbrella) area of the plans. A number of main-memory indexing methods were developed pushing the boundaries of high-update-throughput spatial indexing. On the other hand,

the challenging topic of workload-dependent cost-modeling of performance trade-offs was not explored.

Collaboration with FlexDanmark was started during the evaluation period as part of the DaIn project. The main result of this collaboration has been a data warehousing system prototype that is used to compute travel time in all of Denmark. The system is now used by FlexDanmark for estimating the travel times that serve as the foundation for calculating payments to their subcontractors (amounting to 125+ million Euro/year). The data warehouse and its functionality have been the subjects of a number of research papers. The close collaboration with FlexDanmark was not planned, but has emerged during the evaluation period and is still ongoing, with FlexDanmark being a key industry partner in the DiCyPS project.

Research on data management in indoor settings were conducted mainly in the settings of the ISA, BagTrack, and NILTEK projects. Consistent with the plan, the ISA project designed a hybrid indoor positioning approach that exploits both Wi-Fi and RFID technologies. As planned, ISA also studied how to implement hybrid indoor positioning on mobile hand-held devices. Although the BagTrack and NILTEK projects were not acquired when the plan was made, the two projects covered other planned items including integrated indoor/outdoor space modeling and indexing and querying of indoor moving objects. The two projects also worked on data cleaning and warehousing problems, which was not planned but was needed in the projects. The record of resulting publications in prestigious outlets and technology transfer to the industrial partners, we believe suggests that the unit's research on indoor data management has been successful.

**Umbrella Project in Programming Technology**  Research in this area have generally followed the plans. We have pursued collaborative efforts in mobile systems, especially indoor location based systems, and in embedded systems. We managed to obtain several externally funded projects, four Ph.D. degrees were awarded, and we published several papers.

Especially in the beginning of the reporting period, we made good progress on our new area of programmatic programming with three published papers. However, mainly due to heavy teaching and administration load in the latter part of the reporting period, this area has stagnated. We started activities in the eScience area, especially Ph.D. courses on Supercomputing programming

and master thesis work on programming technology for big data analysis. We also received funding for the P3 project that started in January 2015.

Early in the reporting period we collaborated with data-intensive systems staff, but this collaboration stopped when Christian S. Jensen left and was not resumed upon his return. In the latter part of the reporting period, apart from running a joint specialization course on database and programming technology, there were no joint activities between the data-intensive systems and programming technology staff.

We held a number of seminars to engaging with (local) industry to help them understand the shift in software construction from the object oriented approach to the more declarative and dynamic approach being offered by the new programming languages and new language feature in existing languages. However, we did not succeed in engaging with industry through projects in this area as we had hoped.


**Unanticipated Topics** Research on *energy/flexibility data management* became a prominent activity in the period. Our proposed concept of flex-offers is more general and powerful than competing approaches, and it has been successfully deployed in a number of real-world use cases. Papers were published in leading data management outlets, e.g., TKDE and leading energy-related outlets, e.g., ACM e-Energy, several keynotes were givenm and significant funding obtained. Thus, we find this effort to be successful.

Although not planned, research on *Semantic Web* data management came naturally into the unit with the addition of new faculty members, a number of joint Ph.D. students working on Semantic Web issues, and the increased availability of (Linked) Open Data. We believe to have made good contributions in our focus areas of efficient querying over knowledge bases (QWeb) and Fusion Cubes self-service BI on semantic web data (IT4BI-DC).

We have made a good start on *social data management* in this period, resulting in a number of publications in prestigious outlets, including two conference papers at ICDE and one journal paper at IEEE Internet Computing.

Substantial efforts were devoted to enabling different functionality in relation to *spatio-textual data*, most notably keyword-based retrieval of such data. Members of the unit have been first movers in this research area and have been able to publish a series of papers in top outlets that have received

substantial attention, as judged by citations and invitations to give lectures and for collaboration. The some 18 papers on this topic include papers in ACM TODS, VLDBJ, TKDE, PVLDB, ICDE, and ACM SIGSPATIAL GIS. Half a dozen keynotes were given on the topic in the evaluation period. In addition, prototype systems have been demonstrated.

Research on the *teaching of programming* came naturally with the increased teaching load and needs for adopting new and more efficient teaching methods. As part of the FECTH project, we have researched the effects of using short videos in a first year programming course. In addition we have experimented with variations of "flipped classroom" approaches in the course. Both have been backed by questionnaire surveys among participating students.

In summary, the research on the unanticipated topics were motivated by important developments during the period and achieved good results. Thus, we plan to continue our research on these topics in the next period.

### 3.6.2 Scientific Output

As in the previous evaluation period, the publication strategy of the unit focuses on promoting research quality and visibility of the results. Thus the unit strives to publish its best research results in the best general journals and conferences within the general scientific communities covered by the unit (mainly data management and programming languages). Other results are published in the best specialized journals and conferences within specialized subareas, e.g., spatio-temporal data management, mobile services, data warehousing, semantic web, e-energy systems, and embedded or functional programming. The lowest priority is given to publication in general outlets with little quality control and consequent relatively low quality. This strategy aims to increase the impact and visibility as well as the peer-reputation of the unit and its staff. The unit's ranking of publication outlets is as follows. All those in bold are at the top level in the Danish national bibliometric ranking, BFI.

**General Journals**
> **Tier A TODS**, **TOPLAS**, **VLDBJ**, **TKDE**, and **PVLDB**.
> **Tier B DKE**, **The Computer Journal**, CLS&S, **SP&E**, **IEEE SW**, and **IEEE Computer**, Eurasip Journal on Embedded Systems, ISSE.

**Specialized Journals**

**Tier A** GeoInformatica, **TGIS**, **DSS**, **IS**, and **IEEE ToITS**.
**Tier B** ITS, JOT, **HOSC**, IJDWM, and JMC and TLDKS.

**General Conferences**
**Tier A** SIGMOD, **ICDE**, **EDBT**, **CIKM**, **SIGKDD**, **SIGIR**, **ICDM**, **WWW**, **OOPSLA**, HOPL, **POPL**, **PLDI**, and **ECOOP**.
**Tier B** **DASFAA**, **DEXA**, **ADBIS**, and **IDEAS**.

**Specialized Conferences**
**Tier A** **SSTD**, **ACM SIGSPATIAL GIS**, **MDM**, **SSDBM**, TIME, **SOCC**, ICDCS, **ER**, **SDM**, **ECML/PKDD**, **PAKDD**, **DaWaK**, ISWC, ESWC, ACM e-Energy, ISMIR, MMM, ICME, **OOPSLA/SPLASH**, **ICFP**, **ECIR**, SAS, and **UbiComp**.
**Tier B** ASE, BIaW, Cloud-I, COLD, DaMoN, DLS, e-learning, EnDM, MobiQuitous, MobiDE, W2GIS, STDBM, MobiLight, PerTrans, ITS, DOLAP, BIRTE, DARE, DeRiVE, GIR, GeoRich, ICMB, ISA, JIST, JTRES, ISORC, NWPT, Scheme&FP, LISP, **LCTES**, BYTECODE, RTAS, WCET, and Trafikdage AAU.

The top-tier outlets are the most selective. In particular, publication in tier A general conferences is highly competitive. Papers are typically 10 to 12 pages long (2-column, 9 or 10 point type), and most accepted papers report on experimental studies that require substantial software development. The acceptance rates for the first tiers of general conferences and for some of the specialized conferences are below those of the best journals. In particular, the acceptance rates at tier A general conferences are typically below 20%.

There is a very good match between the unit's top-ranked outlets and the top level in BFI ranking. All the tier A (general and specialized) journals are at the top level, and all but one of the tier A and B general conferences and most of the tier A specialized conferences are at the top level.

In the evaluation period, 213 refereed publications were co-authored by DPT members. A significant portion of them, 52, are in journals, of which 27 are in tier A general journals. Of the 161 conference and workshop papers, 29 are in tier A general conferences and 51 in tier A specialized conferences. In the following table, publications under "Other" spread over a range of journals and conferences in different areas not central for the unit, typically with only 1 paper in each. The table shows that 63.4% of all journal publications were published in tier A general or specialized journals. Similarly, 49.7% of all conference publications appeared in tier A general or specialized conferences.

| | General | | | Specialized | | Other | Total |
|---|---|---|---|---|---|---|---|
| | A | B | C | A | B | | |
| Jour. | 27 (51.9%) | 3 (5.8%) | 3 (5.8%) | 6 (11.5%) | 3 (5.8%) | 10 (19.2%) | 52 |
| Conf. | 29 (18.0%) | 9 (5.6%) | 6 (3.7%) | 51 (31.7%) | 50 (31.1%) | 16 (9.9%) | 161 |

The number of tier A general conference papers was lower in the first half of the period, due to a number of faculty members leaving the unit at the end of the previous period. However, it climbed again in the second half of the period as the unit grew in size. and it meets the unit's own expectations and the high level achieved in the previous period.

Further, we give the breakdown of journal and conference publications according to BFI. Note that some journals (conferences) have not been included in the ranking system and the percentages in the table are relative to the total number of journal (conferences) publications. We are very satisfied to see that by far most of our publications are in BFI Level 2 (top level) outlets.

| | |
|---|---|
| Level 2 journals (the top level) | 37 (71.2%) |
| Level 1 journals | 10 (19.2%) |
| Level 2 conferences (the top level) | 86 (53.4%) |
| Level 1 conferences | 48 (29.8%) |

### 3.6.3 Funding

The unit has had substantial success at attracting funding for its research.

Table 3.1 lists funded projects grouped by starting year, reporting for each group the count of projects, the total amount of funding, and the funding sources. The reported figures are the amount from external grants that go specifically to DPT. As can be seen, the unit has received funding from 32 projects during the period, with a total funding of DKK 62.4 million (EUR 8.4 million), an increase of more than 65% over the previous period. Of these projects, 25, totaling DKK 50.6 million (EUR 6.8 million), started during the period. As can be seen, the amounts of projects and funding vary considerably from year to year, but was generally stable over the period.

The funding sources are varied and exhibit a balance between Danish and European sources and private sources, as well as sources that fund independent

| Year | New Projects | Amount in DKK | Funding Sources |
|---|---|---|---|
| 2008-10 | 7 | 11,747,285 | European Regional Development Fund, FP7, Danish Research Council for Independent Research, COST, InfinIT |
| 2011 | 2 | 2,000,000 | VKR Foundation (private), COST |
| 2012 | 7 | 13,069,445 | ForskEL, Danish Advanced Technology Foundation, FP7, European Regional Development Fund, COST |
| 2013 | 8 | 17,636,361 | Erasmus Mundus, Obel Foundation (private), FP7, Danish Energy Agency, Danish Agency for Science, Technology and Innovation, InfinIT |
| 2014 | 3 | 1,011,840 | Danish e-Infrastructure Cooperation |
| 2015 | 5 | 16,917,864 | Innovation Fund Denmark, Danish Research Council for Independent Research, Weco A/S (private) |
| Total | 32 | 62,382,795 | Sources contributing in excess of DKK 2 million: Danish Research Council for Independent Research, Innovation Fund Denmark, FP7, Erasmus Mundus, European Regional Development Fund, Obel Foundation, ForskEL, Danish Advanced Technology Foundation, VKR Foundation |
| Total new | 25 | 50,635,510 | |

Table 3.1: DPT Funding

versus application-oriented research. The Danish research councils, including the Danish Research Council for Independent Research, Innovation Fund Denmark, Danish Advanced Technology Foundation, Danish e-Infrastructure Cooperation, and the Danish Agency for Science, Technology and Innovation, account for 39% of the funding. European sources, including FP7, Erasmus Mundus, and the European Regional Development Fund, account for another 39%. Private Danish foundations and companies account for 14%, while (semi)public councils and agencies in the Danish energy sector

account for the last 8%. There is a reasonable balance between independent and application-oriented research, with the latter category being the largest.

Four projects exceed DKK 6 million in funding: Center for Data-intensive Cyper-Physical Systems (DiCyPS), Erasmus Mundus Joint Doctorate in Information Technologies for Business Intelligence - Doctoral College (IT4BI-DC), The Obel Professorship, and Querying the Web of Data Easily and Efficiently (QWeb). Smaller, but important, projects include P3 and Reduction. The project families Bagtrack/NILTEK and MIRABEL/Totalflex/Arrowhead are examples of synergy between projects and successful follow-up projects.

In summary, the unit is very satisfied with the level of funding and the fact that it has acquired more large projects from attractive sources than previously.

### 3.6.4 Scientific Service and Recognition

Members of the unit have served the scientific community in a number of leadership roles that indicate substantial peer recognition, including as Editor-in-Chief of ACM Transactions on Database Systems and The VLDB Journal, the two top data management journals, and as Associate and Area Editor of several journals and an encyclopedia. The service also includes a term as Vice Chair for ACM SIGMOD. They have served as program committee co-chair and vice chair and as general chair for a range of conferences, including IEEE ICDE, IEEE MDM, IEEE IC2E, DASFAA, SSDBM, APWeb, and ACM SIGSPATIAL GIS, and have co-organized several workshops. This is in addition to other leadership service and a very substantial number of program committee memberships of the top conferences.

In the reporting period, one unit member who is already an IEEE Fellow became an ACM Fellow, a member of Academia Europaea, and received the Villum Kann Rasmussen Annual Award for Technical and Scientific Research; in addition, he was appointed President of the Swiss National Research Programme dedicated to Big Data, was appointed Distinguished Visiting Professor in the Sa Shixuan International Research Center for Big Data Management and Analytics at Renmin University of China, and was appointed Honorary Professor at Renmin University of China. Another unit member became an ACM Distinguished Scientist and a member of the Danish Academy of Technical Sciences, who picked him to represent Denmark in the 2013 US/EU

Frontiers of Engineering Symposium.

In the reporting period, members of the unit received several best paper, demo, and poster awards, including the ACM SIGSPATIAL 2014 Best Paper Award, the ACM SIGSPATIAL 2014 Best Poster Award, the WWW 2013 Best Student Paper Award, the IEEE MDM 2013 Best Demo Award, the World Smart Grid Forum 2013 Best Poster Award, and the ACM IWGS Workshop 2012 Best Paper Award.

### 3.6.5 Education of Researchers

Educating new researchers has also been an important activity throughout the period. In total, 13 Ph.D. students graduated during the period. Most graduated in the beginning of the period, with 6 in 2011 and 4 in 2012, after which there was a decline, due to a gap in started students, delayed students, and Christian S. Jensen's departure. In the last part of the period, the number of Ph.D. student increased significantly, so it is currently 25. The most important reason for this was the start in 2013 of the Erasmus Mundus Joint Doctorate program "IT Technologies for Business Intelligence - Doctoral College" (IT4BI-DC). The program is (as far we know) the first in Denmark to offer true *joint* Ph.D. degrees, where candidates are enrolled at, and obtain a single joint degree from, two universities. The partners are Universite Libre de Bruxelles (coordinator), Aalborg University, UPC Barcelona, TU Dresden, and Poznan University of Technology. Currently, 16 Daisy Ph.D. students participate in the program, with a similar number expected to join later.

Ph.D. students who have left during the reporting period now hold positions at EPFL, Technical University of Denmark, Shenzhen University, NASA Ames, Via University College, University College North Jutland, Data Artisans, and Google (London, Denmark).

Nine postdocs have worked in the unit in the period. Five are still at Daisy, either as postdocs or assistant professors; three hold academic positions at University of Toulouse, University of Alexandria, and China University of Petroleum; and the last works at the company Systematic.

In summary, the unit finds that it has been successful in educating new researchers and that this area will have increasing importance in the future.

### 3.6.6 Industrial Uptake and Dissemination

In a number of cases, the research results from Daisy have been exploited commercially. By dissemination to practitioners, Daisy contributes to increasing the competent use of technologies and the use of new technologies. In addition, Daisy contributes by training of staff that are subsequently hired by companies and universities. Prominent examples are provided next.

The *sentinels* data cube mining technology co-developed with TARGIT in the DaIn project is now a key feature in TARGIT Decision Suite. *Algorhyme* is the AAU spinout company established to commercialize the Position List Word Aligned Hybrid (PLWAH) compressed bitmap index technology in the form of a so-called Data Cartridge for the ORACLE DBMS. The Algorhyme Query product, now marketed by Algorhyme, and can answer complex queries on combined text and structured data up to 100 times faster than can ORACLE by itself. In the *BagTrack*/*NILTEK* projects, data cleaning and warehousing techniques developed by Daisy are being taken up by the partner Lyngsoe Systems that has started to integrate these into their operational systems for handling airport bags. The *pygrametl* ETL programming framework is publicly available and had in September 2015 more than 650 downloads/installs from the Python Package Index (and an unknown number of downloads from GitHub) and 640 visits to its homepage. The users come from different domains, including finance, advertising, health, and public administration.

The data warehouse solution built for GPS and fuel consumption data has since March 2011 been used by the company *FlexDenmark* to compute the travel times between any two points in Denmark. These travel times are used as the foundation for calculating the payments for taxi trips scheduled and coordinated by FlexDenmark (some 5 million trips in 2014, totaling about DKK 1 billion). Daisy has also made the data available to the public in a number of web-based prototypes. These web pages are heavily used by the Danish municipalities because they are the only web pages that show the congestion on all roads in Denmark. The general public also uses the web sites; during days with media coverage, there were in excess of 70,000 unique visitors. Further two licenses to using the data have been sold.

Indoor positioning software developed in the unit was open sourced via Github and is being used in several places. Software on real-time and embedded systems (*HVM*, *TetaSarts*, *SymRT*) has also been released as open source,

as has software on Programmatic Programming.

In terms of dissemination, Daisy founded the national Danish interest group on Big Data and Business Intelligence, under the InfinIT umbrella. This interest group is headed by a Daisy staff member, who also heads the Digitalization and Big Data theme under the Danish Academy of Technical Sciences. The InfinIT interest group on Embedded Systems Engineering is also headed by a Daisy group member.

Several companies have benefitted substantially from collaborations on indoor positioning. MapsPeople is now 100% focused on delivering indoor mapping solutions and AskCody has a number of customers for this. A former research assistant and a former postdoc are working on indoor positioning at AskCody and Systematic, respectively.

Former Ph.D. students contribute to transfer knowledge from Daisy to their new employers, see details above. Two of the former Ph.D. students have founded companies working on Big Data (Data Artisans and Social Quant).

In summary, we find these results very satisfactory, and generally according to or beyond what was planned.

### 3.6.7 Collaboration

The unit maintains collaboration with a large number of international colleagues. The coauthorships of papers published during the reporting period offer an accurate description of the collaborations.

In particular, the unit has substantial academic collaborations with colleagues at several Asian universities (e.g., NTU, CAS, PKU, RUC, USTC, ZJU, CUP, HKUST, HKPU, HKBU) and Australian Universities (UM, UQ, RMIT), in addition to US universities and labs (e.g., Brown, Auburn, HP Labs) and European Universities (e.g., UV, AU, DUT, ULB, UPC, UJI). The unit also attracts visitors from many universities, e.g., HKUST, ZJU, UCAS, CUP, USTC, BUAA, QUB, AU, and UM.

The unit has substantial collaboration with industry, as can be seen in the appendix. Here, we especially value relationships where companies participate in a series of projects over a longer time period, for mutual benefit. Examples of such long term collaboration includes BI vendor TARGIT which collaborated new BI technologies, FlexDanmark which collaborated on driv-

ing times in road networks, the Danish Energy Agency which collaborated on analyzing EV data, and MapsPeople and AskCody which collaborated on indoor positioning and navigation systems.

Our collaboration increasingly takes place in larger projects with other research units from the department, from the university, or from national or international partners. Often, the project teams are interdisciplinary and working on "grand challenge" problems, e.g., in energy, transport, or learning. Such projects include Popular Parallel Programming (P3), DiCyPS, Totalflex, and a number of EU projects. This development is in line with national and international funding opportunities and is expected to increase in the future, so we well positioned in this respect.

Internal collaboration with other units in the department has continued and even increased over the period, on topics like real-time programming for embedded systems in Java and other high level languages for embedded systems, embedded systems engineering, and combining data management, embedded systems, and human computer interaction for flexible energy systems and data-intensive cyber-physical systems. Again, such broad collaborations are expected to increase in the coming years.

### 3.6.8 Hiring

In general, it has been a substantial challenge to hire high quality Ph.D. students. The unit has not been good enough at hiring internal graduates in data-intensive systems. This is in part due to the shutdown of all the elite masters programs in 2013, including the elite program in data-intensive systems. Hiring external Ph.D. candidates is challenging, although we offer very good salaries and conditions. Establishing elaborate ranking and interviewing procedures has helped somewhat, but the challenge remains. One reason is that Ph.D. recruitment often occurs internationally after the B.Sc. degree rather than the M.Sc. degree. It is hoped that establishment of the 4+4 Ph.D. model at AAU can have some effect.

At the postdoc and assistant professor levels, the hiring opportunities have been significantly better, especially in data-intensive systems where four highly qualified assistant professors were hired, three of which are external and one is internal. Three of these have been promoted to tenured associate professors.

In addition, five very well qualified postdocs, three external and two internal, were hired in data-intensive systems. The external hires have generally been Ph.D. students or postdocs at top European or Asian-Pacific universities.

The permanent staff in programming technology have had an abnormal high teaching and administration load, with some members having loads of nearly 100% during the last couple of years. During the reporting period we did not manage to hire assistant or associate professors in ordinary positions. In the latter part of the reporting period one assistant professor was hired jointly funded by the P3 project and the department.

## 3.7  Plans for 2016–20

Plans are described for data-intensive systems and programming technology.

### 3.7.1 Data-Intensive Systems

We cover first general considerations and then specific research focus areas for the coming evaluation period. In particular, with a substantial number of permanent and junior faculty and a large number of Ph.D. students, currently some two dozen, the staff that focuses on research in data-intensive systems is relatively large. This state of affairs offers new challenges and opportunities.

**Excellence and Impact**   The unit views fundamental and applied research as equally valuable and believes that the excellence and impact of both can be improved substantially through effective collaboration. We aim to focus on excellence in both kinds of research.

In terms of fundamental, or long term, research, performance indicators include publication in top scientific outlets (excellence) and citations (scientific impact), as well as awards, invited talks, and other forms of peer recognition.

In terms of application-oriented research, a more diverse and less established range of indicators exist: demonstrations, industry and user funding, industrial uses of results and software, spin-outs, software downloads, and system and service users.

The unit intends to keep monitoring its performance with respect to the above indicators.

**Synergies—Cross-Cutting Themes**   We aim at realizing latent potentials for synergies in data-intensive systems by identifying and focusing some of our activities around selected, cross-cutting themes. Cases exist where unit members study the same types of techniques in different settings or application domains. By bringing together such unit members, we expect to leverage each others' insights. We expect that this will improve research quality and focus, and we believe that the creation of new connections has the potential to improve the innovative nature of our research.

As an example, we study the theme of ingestion of data into databases in different application domains, including data warehousing, mobility and transportation analytics, and smart-grid data management. Data ingestion covers topics such as extract-transform-load, main-memory index updates, multi-threaded processing, and streaming. Specifically, some members of staff have studied techniques for extract-transform-load in the setting of general multi-dimensional data. Bringing this staff together with staff working in the setting of mobility and transportation analytics holds the potential for advancing the techniques as well as for benefiting the studies in mobility and transportation analytics. Envisioned means of doing so include meetings, talks, and workshops and shared Ph.D. students and postdocs.

Other types of techniques to be considered encompass: scale-in (scale-up), including main-memory and multi-core processing; scale-out, including Hadoop and parallel and distributed processing; dimensional query processing; predictive techniques; best practices for experimental studies, including documentation and automation; open data and services practices; open source infrastructure.

**Synergies—Pipelines**   We also aim at realizing latent potentials for synergies by means of collaboration along fundamental-to-applied research pipelines. No single unit member can focus on and excel in fundamental research and applied research at the same time. While working in the same general topic area, different members of staff position themselves differently along the fundamental-to-applied dimension. A fundamental-to-applied research pipeline is established by means of collaboration among these unit members. The potential benefits are many: Those focusing on more fundamental aspects get inspiration for new and relevant research problems and may see their results being implemented in systems that create value for collaborators

and society in an accelerated fashion. Those focusing on challenges closer to applications may benefit from the availability of fundamental insights and advanced techniques.

The most notable fundamental-to-applied pipeline in the unit is found in the area termed spatio-temporal data management, or mobility and transportation analytics. In this area, the unit has a demonstrated ability to publish in the most prestigious scientific outlets, to make prototype systems available, and to create value for its external collaborators. We aim to maintain close collaboration and to achieve additional synergies.

**Laboratory Support**   In the area of data-intensive systems, the unit is developing, maintaining, supporting, and using increasing amounts of software and an increasing number of systems. This is important for long term research and industrial collaboration and uptake, and it contributes to the visibility of the unit. These activities are increasingly labor-intensive and are difficult to sustain via external funding, in part because externally funded projects are relatively short-lived and because they often address new challenges. Thus, the research activities would benefit substantially from longer-term laboratory support. It is an objective to achieve such support.

**Ability to Seize Opportunities**   We aim to leverage our size to maintain and build our capacity for seizing opportunities for achieving further research activities and results in our area. For example, this includes our ability to compete for funding in areas, where we expect future funding to be available, such as big data, smart societies and cities, energy, and open data. It also includes our ability to collaborate with industrial partners.

Means of achieving this aim include a continued focus on relevant networking and collaboration. In addition, we will consider the feasibility some unit members proactively specializing in areas that show particular promise.

Next, we cover three clusters of envisioned research topics for the coming period. Each cluster is described briefly in the following.

**Spatial, Temporal, Spatio-Temporal, and Spatio-Textual Data**   This cluster represents an evolution of activities that have been ongoing in the unit for more than two decades.

In the coming period, we anticipate that focus will be on enabling real-time *mobility and transportation analytics*, which calls for real-time data ingestion and online querying capabilities. This in turns suggest that main-memory technologies are to play an important role. While focus has been on GPS data from vehicles, we also expect to consider additional data sources and the integrated use of multiple data sources in analytics. Example outcomes include (i) a state-of-the-art platform for network-based mobility and transportation analytics that supports real-time, on-line analytics, (ii) a new paradigm for trajectory-based routing, (iii) enablement of a range of new applications developed in collaboration with external collaborators.

In the previous period, we have had considerable success in the area of keyword based querying of *spatio-textual data*. We aim to continue activities in this area. One important direction is studies of how to evaluate query result quality. We will explore the use of crowdsourcing for this purpose. Depending on how successful we are, this can open to a large number of research directions. Notably, the ability to evaluate result quality allows us to study advanced ranking functions that will challenge existing indexing and query processing techniques.

Finally, we were among those researchers who early on realized and studied the challenges of enabling *indoor location-based services*. We will continue research in this area that provides efficient support for functionality needed by indoor services that utilize cloud computing and are delivered to smartphone users. We will also study analytics on indoor mobility data, to the extent we can gain access to such data.

**Advanced Analytics and Data Warehousing**   This cluster builds on the activities within data warehousing, business intelligence, and analytics that have been ongoing in the unit for almost two decades.

One topic will be *data warehousing and analytics for sensor data* such as large-scale fine-grained time series, e.g., from wind turbine sensors or cyber-physical systems. A key characteristic will be storing and querying data in the form of higher-level models, e.g., based on mathematical functions rather than as individual low-level data items. Data ingestion and extract-transform-load will be investigated, both for historical and streaming data, as will indexing and query optimization. The developed techniques will be integrated in a scalable cloud-based platform.

Another topic will concern tighter *integration of analytics with decision making* and optimization. One aspect of this is scaling multi-criteria decision support techniques such as skyline queries to big data, including novel algorithms for cloud-based computing paradigms and novel recommendation approaches applying dominance analytics to big data. Another aspect is the tight integration of (predictive) analytics with optimization problem solving in order to prescribe the best course of action, so-called prescriptive analytics. Here, we will investigate how to integrate solvers for different problem classes tightly with data management functionality using a single data model and query and optimization language, and how to optimize such analytical workflows and scale them to big data.

Finally, we will continue the successful work on *energy data management* initiated in the last period. The flex-offer concept will be extended to cover even more advanced cases, e.g., to include (running) state information. New aggregation and analytics techniques for flex-offers will be developed, along with new techniques for predicting available flexibility based on user behavior and suggesting flex-offers for it. Flex-offer concepts and techniques will also be integrated and aligned with existing smart grid technologies and (existing and emerging) energy markets.

**Web and Social Media Data Management**  This cluster connects more recent activities in the unit that have emerged in the last period.

Linked Open Data forms a Web of Data that is a promising data source for analytics. In the previous period, a number of activities emerged around adding multidimensional (OLAP) support to RDF, Linked Open Data, and SPARQL, covering the modeling and management of metadata, distributed and federated query processing, semantic extract-transform-load, and physical level optimizations. We plan to integrate these activities with the overall goal of enabling easy and efficient querying of the Web of Data, covering both "self-service" business intelligence and more general user queries. A number of novel aspects will be considered. Provenance, quality, and context become even more essential in the (Linked) Open Data setting. Techniques for discovering new data sources and determining their relevance and interestingness will be developed. Techniques for guidance and recommendation as well as explaining query results will be investigated. Techniques and tools for annotating and integrating data sources with additional multidimensional

semantics will be developed. Further, we will build on existing collaborations and establish new ones for the novel aspects.

Social network platforms have accumulated huge volumes of user-generated data by digitalizing many aspects of people's daily lives, e.g., social relationships, behaviors, and locations. Such data can be leveraged to improve peoples' lives. In the previous period, we started to study social media data with a focus on search for local information for users with specified needs. We plan to continue this line of research, putting emphasis on the value and relevance of the search results. We also aim to study how to detect interesting human-involved patterns from social data, thus achieving insight into how to solve particular types of transmissible societal problems, e.g., food poisoning. For this new line of research, we will seek concrete collaborations with researchers from the humanities and social sciences and other relevant domains at Aalborg University. We also aim to develop a scalable platform that is able to support a wide range of geo-social analytics queries efficiently.

### 3.7.2 Programming Technology

Programming technology is at the core of any software development and the demand for software keeps increasing. Thus, research into making programming easier and faster and at the same time safer and resource optimal is paramount to the IT industry in particular and society in general. We foresee that the trend towards a more declarative programming style in mainstream computing will continue, especially combined with more declarative approaches to concurrency, parallelism, and distributed computing. Although lagging behind this trend will reach apply to embedded systems. We also foresee a more interactive and programmatic approach to programming. Finally, we foresee a need for predictability of time, energy and cost of computations.

Embedded system development is still dominated by C (and ADA), but C# and Java are becoming more widespread, especially Safety Critical Java is now starting to become a viable alternative to C in hard-real-time embedded systems. Although we foresee that the programming technology will continue to be fragmented, we expect a trend towards a more integrated and coherent approach integrating object and function oriented programming with high level concurrency constructs. We also foresee a more interactive and programmatic approach to programming. Finally, we foresee a need for predictability

of time, energy and cost of computations.

The subgroup expects to pursue research in the following areas:

- Interactive and programmatic programming in live programming environments.
- Teaching of programming and technologies and techniques for supporting teaching.
- Declarative and interactive, parallel and distributed technologies for Big Data Analysis.
- High level programming of low level components for embedded systems and the Internet of Things.

Underpinning the latter two subjects are methods for predicting time, energy, and cost of computations. We expect to continue and extend our collaborative efforts, especially in embedded systems and in data intensive applications.

Our goal is to maintain or even increase our current activity level by seeking funding for new projects and Ph.D. students. We hope to decrease the teaching and administration load on permanent staff by attracting competitive candidates for departmental assistant and associate professorships and further boost the staff level by jointly funded positions. We will also decrease teaching load by ''outsourcing'' some duties to other parts of the department.

We expect to continue our participation in the FETCH and P3 projects. We aim to establish two new collaborative projects on a platform for interactive real-time big data analysis and on high level programming of embedded systems, respectively. Funding for these activities should be sought for from National and EU funding agencies. We may also attempt to start activities in unconventional computing if the right collaboration partners and funding opportunities arise. Research on teaching programming will continue.

## 3.8 Committee evaluation

**Observations**

The Database and Programming Technologies (DPT) unit consists of two Full Professors, nine Associate Professors, and two Assistant Professors. Given this size and looking at the output of the unit, the conclusion is that the unit is of excellent quality. The unit performs at world-class level.

During the evaluation period 213 refereed publications were co-authored by members of the DPT-unit. They appeared both in journals and high-level conference proceedings. A substantial part of papers appeared in Tier A+B general journals and proceedings, among which are highly ranked general journals such as TODS, TKDE, and PVLDB, and highly ranked general conferences such as SIGMOD, ICDE, EDBT, CIKM, ICDM, and WWW. Furthermore, a significant number of papers appeared in Tier A+B specialized journals and conferences. In total, 113 publications appeared in Tier A.

Quite a few papers are cited extremely well, resulting in a h-index for Christian S. Jensen of 70 which is outstanding and places him among the top echelon of researchers in this area. Furthermore, the other full/associate professors have high h-indexes as well. Excellent results have been achieved in the planned topic areas of Cloud Intelligence and Spatio-Temporal Data Management and Mobile Services, and in the unplanned data management topic areas such as Energy, Semantic Web, Social Networks, Spatio-Textual data, and the programming technology topic area Education.

In the evaluation period 13 PhD defenses took place. Given the size of the group this could have been a bit higher. After the return of one of the full professors the external funding has increased substantially. In the evaluation period, 25 new projects started with total grants to the unit of 50 M DKK. The funding came from both national and European sources (among which the Erasmus Mundus Joint Doctorate program). This resulted in a substantial increase of the number of PhD students. So, it is expected that the number of PhD defenses will improve significantly.

The Center for Data-Intensive Systems (Daisy) has been an excellent platform to combine the various research activities in the unit and make them visible to the outside world. Daisy makes it possible to address real world problems together with industry and to combine research and innovation. One spinout has been established, and several software packages are available as Open Source.

The members of the unit are very active in PCs of many high quality conferences and editorial board of prominent journals. Worth mentioning are: Editor-in-Chief of ACM TODS and the VLDB Journal, and Vice Chair of ACM SIGMOD, showing the high peer recognition.

On the whole, the unit is of excellent quality, both in research and in inno-

vation. The unit has a good mix of fundamental and use-inspired research, making it robust for changes in funding. More funding in future will go via application areas. The unit has a good vision on Big Data and a clear focus on specific application areas, such as energy (e.g., balancing available supply and the current demand) and transport (e.g., travel prediction based on available GPS data).

## Recommendations

The number of PhD defenses given the size of the unit has been rather low. More external funding is required to appoint more PhD students. In the last phase of the evaluation period, quite a few new PhD students have been appointed. So, this weakness is being addressed properly.

The report and the presentation mention "perceived" real world applications. Given the maturity of the computer science field, it is important to look at real applications to be aware of the full extent of the challenges of the applications.

There is a number of opportunities. Big Data is getting more and more attention, also outside the academic world. In many companies and organizations, there is a need to handle large volumes of data and gain knowledge from it to improve products and services. The number of possible applications is enormous. It is important to focus on specific applications while still producing generic solutions or results.

Furthermore, society needs people that know how to handle data, so it is important to educate students in Data Science. The recommendation is to start a Data Science program as soon as possible.

Given the change in funding, that goes more along the lines of applications, there is the threat that Computer Science becomes invisible. Because of this, it is important to claim a strong position of CS in these projects to cover also the CS challenges. For DPT, this means that it should have a clear research agenda for the database technology challenges it wants to address.

Given the changes in the requirements coming from society and the advances in CS, it is important to have a flexible organization to adequately support cooperation across units. It is suggested to have an open mind about the positioning of Programming Technology in the departmental structure, e.g.,

to consider whether a matrix structure can help in this case as well as others.

# 4 Distributed and Embedded Systems

## 4.1 Executive Summary

The research of the distributed and embedded systems unit concerns modelling, analysis, and realization of computer programs, with an emphasis on distributed and embedded systems. The contributions of the unit include work on semantic foundations, algorithms and tools for verification and validation, and methodology and application. The research activities are both theoretical/foundational in nature as well as practical/experimental.

In the period, the unit has strengthened its established stronghold and leading position within verification and validation of embedded systems and foundational semantic theories. The unit has successfully expanded its research area towards the emerging area of cyber-physical systems which is witnessed by several new projects in the area, including the Sino-Danish basic research center IDEA4CPS,the innovation center DiCyPS, and the prestigious ERC advanced grant LASSO. The unit continuously develops the leading and internationally recognized supporting design and verification tool suite Uppaal. For this work, Kim G. Larsen was co-awarded the 2013 CAV award. Two notable extensions have been developed and publicly released in the period: SMC (statistical model-checking) supporting scalable performance evaluation, and Stratego enabling controller synthesis and optimization.

The unit has been involved in extensive research collaborations. Internationally, the unit has been involved in several (15) EU projects, and has coordinated the STREP projects Encourage and Sensation. It has attracted 37 new research projects with a total grant size of 103.8 M. DKK. Nationally, the unit renewed and continued coordination of the ICT Innovation Network Infinit, that serve as an important vehicle for national dissemination. The unit

also emphasizes industrial impact by involving numerous industrial partners in strategic research projects, collaborating with the Danish Industrial Association, and by running the competence center CISS; here a permanent staff serves as match-maker and conducts several smaller collaboration projects.

A high international research impact is also witnessed by the high number of high quality publications in international peer-reviewed conferences and workshops (175), journals (63), books (2) and book chapters (6), awards and invited talks, numerous services to the research community, and a very high number of external visitors.

## 4.2  Profile

The research of the unit concerns modelling, analysis and realization of computer programs, with the emphasis on distributed and embedded systems. During the last decade the unit has established a recognized profile and substantial expertise in foundations, methods and tools for embedded software systems design. In 2001 the Centre for Embedded Software Systems (CISS) was established in collaboration with the Institute for Electronic Systems at Aalborg University, with the ambition to create an industry oriented research center of excellence. During the period 2006-2010 the unit amplified its research on model-driven and component-based development, enabled by the Advanced Technology Platform Project DaNES, and semantic foundations for integrated modelling of quantitative system properties like discrete, stochastic and continuous behaviour, enabled by the VKR Center of Excellence.

Leveraging these strongholds, the unit has in the current evaluation period (2011-2015) broadened its research scope towards cyber-physical systems (CPS) in which networked compute nodes interact with the physical environment, collaborate, and operate in context of other systems and human users. Applications include automotive systems, energy optimization, home automation and SmartGrids. The research thus spans the entire range from foundational research, strategic research to industrial application.

The research involves the following key elements: 1) Semantic theories for modelling the behavior of computer programs and systems. 2) Design, implementation and models for analysis of distributed and embedded systems. 3) Algorithms, methods and tools for verification and validation of programs

and systems.

Each of the three research areas mentioned above constitutes a subject in its own right. Moreover, the areas are interrelated in a number of ways: semantic models offer important guidelines for development of languages and paradigms for distributed systems; semantic models are necessary prerequisites for development of verification algorithms and tools; the development of validation tools provides new insight into the underlying semantic models on one hand, and it is applied in environments for the construction and analysis of distributed systems; the evolving nature of distributed systems provides insight into the strengths and weakness of existing semantic models, and serves as an inspiration for the development of new ones; finally, distributed systems truly expose the limits of developed verification algorithms.

## 4.3  Staff

### 4.3.1 Current Staff

**Professors:** Kim Guldstrand Larsen, Anders Peter Ravn (20%), Axel Legay (20% Velux guest professor).

**Associated Professors:** René Rydhof Hansen, Hans Hüttel, Josva Kleist (10%), Radu Mardare, Brian Nielsen, Ulrik Nyman, Arne Skou, Jiri Srba.

**Assistant Professors:** Giovanni Bacci, Mads Christian Olesen.

**Permanent Researcher:** Marius Mikučionis.

**Post Docs:** Giorgio Bacci, Marco Mũniz, Petur Olsen, Bingtian Xue.

**Project Related Researchers:** Mads Kronborg Agesen, John Bjerregaard.

**Ph.d.-students:** Andreas Engelbredt Dalsgaard, Søren Enevoldsen, Louise Foshammer, Peter Gjøl Jensen, Simon Laursen, Thibaut Le Guilly, Mikkel Hansen, Anders Mariegaard, Thomas Pedersen (leave), Mathias Ruggaard Pedersen, Danny Bøgsted Poulsen, Jakob Haahr Taankvist, Zhengkui Zhang.

### 4.3.2 Staff Development

The full time staff of the unit currently consists of 1 full professor, 7 associate professors, 2 assistant professors, 4 post-docs and 13 Ph.d students. These

are complemented by 3 part time staff members. Compared with the previous evaluation period, the scientific staff has increased by one associate professor and one assistant professor. However this is partly offset by one professor working part time and now completely retiring. The unit has maintained approximately the same level of Ph.d. students and post-docs.

In the period an assistant professor in the area of computer networks was hired, but desired to move back to his home country. To strengthen the area of distributed systems and network, a new associate professor will join the unit ultimo 2015. A project related researcher has been promoted as a permanent senior academic staff to consolidate tool application and development, and the maintenance of the Uppaal tool-suite.

In conclusion, the staffing of the unit has been maintained, but is still challenged by an increased teaching load, and the demand for funding applications and resulting project administration.

## 4.4  Goals (2011-2015)

The following outlines the goals and plans setup for the current period.

### 4.4.1 General plans

We want to strengthen our collaborations and current leading role within Embedded Systems both nationally and at the European level, e.g. involving creation and offering of a national set of PhD courses on Embedded Systems.

We want to maintain strong collaboration with research groups from the Department of Electronic Systems within Control Theory and MPSoC, overcoming the distribution in location. We want to maintain high internationally visibility within the areas of model checking and concurrency theory.

We want to expand our leading position with respect to verification and validation of distributed systems by a clear strategy on our effort on tool development. This includes application of our existing mature tool engines to specialized tools (e.g. RT Java Workbench), establishment of a lightweight model checking tool box for easy prototyping, and a full exploitation of its potential by combining existing expertises (model checking, testing and static

analysis) as well as bringing in new scalable techniques (constraint solving, SMT solving, statistical model checking and learning).

The effort of the unit on GRID computing, currently active on the Nordic DataGrid Facilities, should focus more on the unit both research- and teaching-wise. This will allow the unit to engage in the emerging area of Cloud Computing and link to the existing Embedded Systems Platform.

We will extend the domain of distributed systems considered in our formalisms, tools and methodologies from Embedded Systems into the direction of Multi-Core and Cypher-Physical Systems.


### 4.4.2 Concrete plans

We also stated a series of concrete goals that are summarized in the following. Within *Concurrency Theory* the main goals were to extend our formalisms with quantitative notions of energy, cost, and probabilities, and to develop a theory of metrics generalizing established equivalences and preorders. Within *verification and validation* main goals were to combine model-checking with static analysis techniques and implement a statistical model-checker for Uppaal. We also envision a technique for model-based testing incorporating stochastic distributions and further tool integration. Within *Networks and Operating Systems* goals were to increase activities on cluster and high-performance computing, parallel algorithms, and handling computation on large amounts of distributed data. We also envisioned a resource aware embedded systems platform using high-level programming languages and enabling service oritented applications. Within *Embedded Systems Methodology* the goals were to formal model web services, and to develop methods for component based systems. We also planed to write an industrial handbook. Finally, with respect to *Staff and Funding* we would work for better career opportunities and (especially tenure) positions. We also recognized the increasing need for external funding to maintain the level of Ph.D. students and post-docs.

## 4.5  Results

We report on the research of the unit during the period 2011-2015 within the following strands:

*Foundational Theories:* Semantic theories and meta-theories for concurrent processes and their logical properties. Models, logics and analysis for hybrid, real-time, and probabilistic processes. Semantic theories for processes whose communication topology changes dynamically, including security protocols.

*Verification and Validation:* Development and implementation of data structures, algorithms and tools for model-checking, static analysis, synthesis, and testing of real-time systems and CPS. Applications to communication protocols, control programs, and planning and scheduling.

*Methodology and Application:* Methodologies for specification, analysis and testing of CPS, including networking, home automation and SmartGrids, web-services, and biological systems. Application to industrial case studies.

### 4.5.1 Foundational Theories

**Compositional Specification Theories:**  Compositional specification theories provide the means for developing correct implementations in a step-wise refinement manner.  In particular, given a composite system, such a theory allows properties of the overall system to be inferred from (and reduced to) properties of its components.

Together with Legay, Traonouz (INRIA, Rennes) and Wasowski (ITU) we have introduced the first complete specification theory for real-time systems, with component specification being given as timed I/O automata and refinement as alternating simulation. The resulting theory has been implemented in the Uppaal branch ECDAR with two journal publication in STTT.

**Modal Transition Systems:**  Modal transition systems (MTS) is a specification formalism introduced more than 25 years ago, which extends the classical notion of labelled transition systems (LTS) by decorating transitions with modalities (*must* and *may*) with an accompanying co-inductive notion of when an LTS implements a MTS. Since the introduction, the problem of the exact complexity of the notions of thorough refinement and consistency

between MTSs have remained open. In a sequence of papers in collaboration with researchers at Brno (CZ) we have settled the complexity of these open problems.

Despite its many advantages, MTS are incapable of expressing some practically needed aspects in the refinement process. We have introduced a new model called parametric modal transition systems (PMTS) together with a general modal refinement notion that overcomes many of the limitations and investigated the computational complexity of modal and thorough refinement. Also we introduced and studied dual-priced MTS with two kinds of quantitative aspects reflecting (static) investment cost and (dynamic) running cost. These two topics constitute the core of the PhD thesis of Michael H. Møller.

**Metrics:**   For the comparison of the behaviour of quantitative systems, metrics are preferred to equivalences since the latter are not robust with respect to small variations of the numerical values used in their specifications (e.g., time delays, energy consumption levels, and transition probabilities). Pseudometrics and hemimetrics respectively generalize the concepts of equivalence and simulation between systems by measuring to what extent the behaviours are different.

In collaboration with Legay (INRIA, Rennes) we have studied both the theoretical and computational aspects of several metrics for weighted transition systems. In a journal publication in TCS, we proposed the first complete axiomatization for two simulation metrics over regular weighted transition systems: the point-wise and accumulation hemimetrics. The computational problems for the two metrics are shown to be in P and NP $\cap$ coNP, respectively, by means a reduction to the problem of computing payoff for discounted games. The relation between metrics and games has been further investigated in two following papers, where we developed a general framework for reasoning about distances between transition systems with quantitative information. The framework generalizes and unifies a large variety of previously considered metrics, and it has been proven useful for the study of their topological properties.

A large number of results have been published on stochastic systems, including discrete and continuous-time Markov chains (MCs, CTMCs), Markov decision processes (MDP), semi-Markov chains (SMCs), and Markov processes

with continuous state space (MPs).

From a theoretical perspective, the algebraic properties of metric-based semantics for MPs have been studied via a duality between Stone MPs and metrized Aumann algebras[1]. This work have been conducted together with Panangaden (McGill, Montreal) and Kozen (Cornell University, Ithaca). In other two works (published at MFCS'12 and QEST'14) we focused our attention on the study of continuity properties for several metrics on stochastic systems from their logical, algebraic, and topological accounts. More recently, in two papers published at FoSSaCS'15 and ICTAC'15, we related the linear-time metrics over (semi-)Markov chains to the probabilistic model checking problem over LTL and MTL formulas. We also showed that linear-metrics can be approximated via converging sequences of increasingly-weaker branching-time metrics. Interestingly, even though the problem of approximating the linear-metric is NP-hard (the exact decidability is still an open problem), we showed that each of our proposed branching-time approximants are polynomial-time computable. With Gebler (CWI, Amsterdam) and Tini (Insubria University) we developed specification rule formats for compositional operators on non-deteministic labelled probabilistic systems.

From a computational perspective, we developed on-the-fly algorithms for computing the bisimilarity metrics on MCs, CTMCs and MDPs. These algorithms have been published at TACAS'13, MFCS'13, and QEST'13 and implemented in a tool-library that accounts also for improved compositional techniques. The TACAS'13 paper has been nominated for the best paper award at ETAPS'13.

**Logics:** Complex systems that combine artificial (software-based) components and natural components are the new challenges today in Engineering and Technology. They can be found in areas as diverse as aerospace, automotive engineering, chemical processes, civil infrastructures, energy, healthcare, manufacturing, transportation, and consumer appliances. From a modelling perspective this systems combine entities with discrete nondeterministic behaviours with entities having a continuous behaviour modelled either as stochastic/probabilistic systems or some sort of systems caring continuous information about time or resources consumption.

---

[1]This duality extends the one described in the section on Logics.

Regarding the stochastic/probabilistic systems, in collaboration with Cardelli (Microsoft Research Cambridge, UK), Panangaden (McGill Univ., Canada) and Miculan (Univ. of Udine, Italy) we have investigated various issues regarding the logical and mathematical foundations of such systems that allowed us to deeper understand their nature.

To specify and study properties of such systems, a plethora of logics have been proposed in the literature. For the probabilistic and stochastic systems in particular, we have done a comprehensive model theoretical analysis in collaboration with Luca Cardelli (Microsoft Research Cambridge, UK), Dexter Kozen (Cornell Univ. USA) and Prakash Panangaden (McGill Univ., Canada). In this research we analyse for the first time proof systems for these logics aiming at achieving sound-complete axiomatizations against various stochastic semantics, also including compositional issues. These are non-compact modal logics and hence, the weak and strong completeness do not coincide. We clarified these aspects and propose strongly complete axiomatizations.

The study of the complete axiomatizations for stochastic/probabilistic systems, in the context of the mathematical foundations that we provided for these models, allowed us to prove Stone-type dualities for Markov processes against Aumann algebras (which are the Boolean algebras corresponding to the aforementioned logics). Our stone duality generalizes, for stochastic/probabilistic systems, previous work by Stone and Tarski. This work is the result of a collaboration with Dexter Kozen (Cornell Univ. USA) and Prakash Panangaden (McGill Univ., Canada).

Another class of systems that we have intensively studied in the last years are the weighted transition systems carrying informations about various types of resources. In recent works we took the challenge of solving long-standing open problems regarding weighted logics involving clocks (including the alternation-free weighted Mu-calculus), and their decidability and axiomatization. In particular, we have also developed, for the first time, a complete axiomatization for timed logics against timed transition systems.

**Theory of Timed Systems:** The theory of timed systems has been one of the most active research area since the origin of the DES unit. The two most studied models in the past five years were timed automata and their different quantitative extensions as well as the model of timed-arc Petri nets.

The theory developed for these timed systems has a strong connection to the development of verification tools and experimental evaluation stemming from many different application domains.

The main focus in the theory of *timed automata* was on implementability and robustness issues, priced timed automata and optimization questions with connection to timed games, automatic synthesis of parameters for parametric timed automata, multicore reachability/LTL model checking as well as the development of alternative techniques for verification of timed systems that do not depend on difference bound matrices (DBM). In particular, we studied the use of max-plus algebra as an alternative to continuous-time analysis of timed automata, as well as the use of discretization of time and development of compact data structures for storing and searching through large discrete state-spaces. Numerous case studies and tool development are mentioned separately in the sections to follow.

Model checking of *timed-arc Petri nets* was initially performed by translating the problems to the setting of timed automata but later on dedicated techniques for continuous time, discrete time and untimed reachability/liveness analysis were developed too. A particular focus was on the use of discrete time verification, development of suitable data structures and efficient approximation techniques. The theory was applied to the area of timed workflow analysis, including efficient tool implementation and a few case studies concerned with, for example, medical workflows.

**Process Algebra and Behavioural Types:** A main focus of the work on behavioural type systems has been that of providing a general account of the many different type systems that exist for nominal process calculi. This has led to generic type systems for the $\psi$-calculus framework: a generic type system that captures a variety of extensions of the simply typed $\pi$-calculus and another generic type system that captures resource-conscious type systems that use a notion of environment splitting.

Another line of work on behavioural type systems has been that of devising algorithms for type inference. Finally, another activity in the area of process calculi has been that of understanding how to capture multicast-related notions of communication, and in particular the relationship between broadcast and aggregation.

**Energy Games:**  Priced and game extension of timed automata was introduced in 2001 and 2005 and offer ideal formalisms for modelling a number of dynamic resource allocation and optimization problems. During the period we have been invited by Communication of the ACM to publish a review article describing these models and the associated complexity and decidability results. In 2008 we introduced the notion of Energy Automata, being and extension of priced timed automata, where the cost-rates may be both positive and negative. Closing an open problem we have shown that the existence of an infinite lower-bound-constrained run is undecidable for weighted timed automata with four or more clocks. The above work has largely been carried out with collaborators in LSV, Chachan (F).

In addition, effort has been put on understanding lower- and/or upper-bounded run-problems in the untimed setting. In particular, we have extended (untimed) energy games to a multiweighted and parameterized setting, allowing us to model systems with multiple quantitative aspects.

Most recently we have introduced the notion of average-energy games, where the goal is to optimize the long-run average of the accumulated energy. We have shown that this objective arises naturally in several applications, and that it yields interesting connections with previous concepts in the literature. We also considered the case where the system has to minimize the average-energy while maintaining the accumulated energy within predefined lower and upper bounds bounds at all times, corresponding to operating with a finite-capacity storage for energy as in energy-games. This work has been carried with our collaborators in LSV, Chachan (F) and in Saarbrucken (D).

Other quantitative games considered include Nash-Equilibria for concurrent priced reachability games, as well as Action Investment Games, where we have studied the trade-off between investments limited by given budgets and resource constrained (energy) behaviour of the underlying system.

**Synchronizing Words:**  The problem of synchronizing automata is a classical problem in automata theory, and is concerned with the existence of a word that sends all states of the automaton to one and the same state. This problem has classically been studied for complete deterministic finite automata, with the existence problem being NLOGSPACE-complete.

We have considered the synchronizing-word problems for weighted and timed

automata in several variants and combinations of these including deterministic and non-deterministic timed and weighted automata, synchronization to unique location, as well as synchronization with a safety condition.

Furthermore we have considered the existence of synchronizing *strategies* in the presence of partial observability. We have made a comprehensive complexity analysis of several such problems in the setting of finite automata, and obtained a surprising polynomial time decidability result in the setting of deterministic weighted automata, where the adaptive strategy is only allowed to observe whether the current weight-level is negative or nonnegative.

**Security:**   Work on *cryptographic protocols* includes the design and analysis of concrete protocols using the applied $\pi$-calculus and ProVerif as well as the development of type inference strategies for proving authenticity properties of injective agreement. This inference strategy is now part of the TyPiCal tool developed at the University of Tokyo. Finally, there have been results on generic type systems that can be instantiated to existing type systems for authenticity and secrecy, respectively.

Most of the work on *security in socio-technical* systems has been performed in the context of the TREsPASS project. Here the focus is on applying formal methods to modelling and analysis of complex organisational structures, taking human behaviour and (informal) social organisation into account, with the goal of finding novel attack vectors, including attack vectors that exploit human/socioal factors through insider knowledge and social engineering. This information can then be used to perform a more precise risk assessment and countermeasure prioritisation than what is currently possible. Research in this area also includes formalisation and analysis of various attack modelling formalisms, e.g., attack trees.

*Automated vulnerability analysis* of code and systems to detect potential security problems is a well-established discipline in security research. Our work resulted in two new analyses: one for *Dalvik bytecode*, the low-level language underlying Android systems; and one for PHP, one of the most widely used languages for web-site development, and also a very dynamic language with very few static controls making analysis challenging. Furthermore, the use of the Coccinelle tool for secure coding standards compliance has been explored, and some exploratory work on pointer analysis and value-set analysis of low

level code has also been completed in this period.

## 4.5.2 Verification and Validation

**Statistical Model Checking:** Substantial effort has been made towards developing and implementing *statistical model checking* (SMC) techniques within the Uppaal tool. In essence, given a probabilistic or stochastic model $M$ and a linear temporal property $\phi$ , SMC estimates or settle hypothesis about the (unknown) probability measure of the set of run of $M$ that satisfies $\phi$ by simulation with the price of only providing answers with a certain (required) confidence.

The SMC engine of Uppaal is based on an extended stochastic semantics of networks of timed automata. The stochastic semantics of a network is given by repeated output-races between components, which are supposed to be independent. In particular, the stochastic semantics empowered by the SMC engine, allows for a large range of performance properties to be addressed, These performance properties complement and refine the hard real-time safety properties addressed by the classical symbolic model checking engine of Uppaal.

The work on SMC has been in collaboration with INRIA, Rennes and addressed initial estimation, sequential hypothesis testing of cost- and time-bounded reachability probabilities for stochastic priced timed automata. Later, a substantial extension of the theory and SMC engine were made supporting performance analysis of stochastic hybrid systems. Also, a distribution implementation of SMC was implemented, being almost "embarrisingly parallisable" except for the sequential testing method where care had to be taken in synchronizing the findings of runs from different cores to avoid bias.

In collaboration with ISCAS, Beijing, China, and as part of the Sino-Danish Basic Research Center IDEA4CPS, SMC methods for dynamical networks and for bounded MITL properties have been developed These aspects constitute the core of the submitted PhD thesis of Danny Poulsen.

The SMC engine of Uppaal has by now been applied to numerous applications including analysis of schedulability for mixed critical systems, subway control systems, energy-aware sensor networks, comparison and optimization of strategies for room heating, biological systems, battery-aware schedulabil-

ity analysis.

**Synthesis:**   We have continued work on controller synthesis for timed games in different directions, based the branch Uppaal-TIGA and its on-the-fly synthesis algorithms from 2005. For games with partial observability, we have devised an efficient method for determining a subset of observers that is sufficient for control yet has minimal cost. Moreover, extending the long line of research on robust analysis of timed automata, we have offered an efficient method and implementation for synthesizing *robust* strategies for timed games, i.e. strategies that are winning even under small pertubations of constraints on clocks in the game automata. Finally, we have extended controller synthesis from safety and reachability objectives to objectives expressed in a fragment of MITL.

Most recently we have made significant advances in combining our methods for symbolic controller synthesis using in Uppaal-TIGA with the simulation based performance analysis of Uppaal-SMC. A first important contribution, allows a synthesized strategy to be viewed itself as a timed automaton (though with an internal representation), restricting the behaviour of the original timed game. In particular, additional safety and performance properties of the timed game under the synthesized strategy may now be verified. Even more ambitions, in the new branch Uppaal-STRATEGO, strategies are first-class objects that may be obtained by either symbolic synthesis (with respect to safety objectives) or by highly scalable reinforcement learning using simulation (with the objective of optimizing the expectation of a specified performance-measure). Moreover, additional properties – or even further synthesis – of the original timed game under the restriction of such a strategy may be conducted. In this manner, strategies that are both guaranteed safe and almost optimal may be synthesized. Uppaal-STRATEGO has already been applied to different applications including synthesis of safe and optimal cruise controllers, and to synthesis of deadline-safe scheduling for battery-powered embedded systems that optimizes the expected life-time of the battery.

**Distributed and Parallel Verification:**   Practical verification problems are often bigger than what can be tackled by a single processor, or even by a single computer. One area of research is in utilizing the large multi-core computers

and compute clusters available nowadays.

Often algorithms cannot directly be ported from a sequential setting to a parallel/distributed setting, as the architecture has different attributes: memory access times are no longer uniform, and shared objects need to be protected by locks or by very careful coding, i.e. lock-free data-structures. With the Formal Methods and Tools group at the University of Twente, we have combined our tools with the LTSmin multi-core model checker. We have created specific algorithms to take into account the partially-ordered nature of the state spaces we consider, and in particular we have derived new algorithms for checking Büchi acceptance conditions of timed automata in a multi-core setting. The applications have been shown to scale linearly up to around 40 cores.

Recently we have developed algorithms for time optimal reachability analysis using principles of swarm verification, and implemented these on AAUs model-checking cluster with promising speedup.


**Testing and Learning:**   In *Model-based Testing (MBT)*, the goal is to automatically generate and execute sound and effective test cases from models specifying the required behavior to check that system under test (SUT) correctly implements that behavior. In context of the MBAT project we have developed methods for combining model-based testing with static analysis and model-checking in industrial workflows. We have also extended our techniques, theories, and tools for MBT of real-time systems in several directions, including monitoring hybrid systems, data transformations, and support for remote testers.

A practical limitation of MBT is the prerequisite of a system model. In *learning*, the goal is to automatically construct an abstract behavioral model of a an actual SUT through observed runs (passive learning) possibly supported by stimulation (active learning). Properties of the SUT as represented by the model may then examined though analysis of the learned model. We have improved learning of deterministic finite state-machines by using abstractions and refining these using counter-example guided abstraction refinement techniques. More original is our work with colleagues from the Machine Intelligence unit where the aim is learning probabilistic system models (labelled Markov decision processes, probabilistic automata) from observed

input/output runs with the purpose of model-checking the system. Our results includes a new learning algorithm for this setup (inspired by the Allergia algorithm), proofs of convergence between generating and learned model, experiments demonstrating the accuracy of preservation of PLTL properties.

**Tool Development**   We emphasize the importance of validating the theoretical results via experiments and our unit has been involved in the development of numerous tools, among others the Uppaal-family (including Uppaal, SMC, SMC, TiGa, Stratego, TRON, Yggdrasil, OPAAL, TAPAAL), Caal and Coccinelle.

### 4.5.3 Methodology and Applications

**Real-Time and Safety Critical Java:**   Much of the work on real-time and safety critical Java was done as part of the CJ4ES project and, in particular, by the PhD projects of Stephan E. Korsholm and Andreas E. Dalsgaard. The main focus has been on developing novel insights, analyses, and tools to aid programmers for developing real-time and/or safety critical applications, using specialised variants of Java. The advantages of using a (variant of a) high-level language, such as Java, includes a higher abstraction level making programs more portable and maintainable; easier recruitment of (new) developers to a highly specialised field; and the availability of strong tools and tool chains both for development and for analysis/certification. One of the major topics in this area, have been investigations into worst-case execution time (WCET) analysis, both from a foundational perspective as well as from a purely tool-based perspective.

As a consequence of AAU's strong commitment to real-time/safety critical Java, professor Anders P. Ravn is currently on the standardisation committee for Safety Critical Java.

**Schedulability Analysis:**   Ensuring the timely operation of embedded systems is as important as ever. In this context model checking techniques have been applied to check the schedulability of mission critical embedded systems, in the form of model-based schedulability analysis. This included a detailed case study of the schedulability of the Herschel-Planck satellite. *Model-based*

*Schedulability Analysis* can including many more details in the analysis compared to classical schedulability analysis, but it can suffer from state-space explosion. The Herschel-Planck case study was also the inspiration for a line of research that was conducted in the CRAFTERS research project exploring the possibilities of model-based schedulability analysis. Most of the research deals with schedulability in the context of *Hierarchical Scheduling Systems* (HSS) where the scheduling system is divided into components that can be analyzed more or less independently. Apart from time, different aspects such as energy consumption, shared resources, quality of service and voltage/frequency scaling have been dealt with in varying versions of the model-based scheduling framework. The contribution of model-based schedulability analysis is the flexibility compared to classical schedulability analysis frameworks. This line of research has also lead to an ongoing cooperation with Insup Lee from University of Pennsylvania.

Related to this but taking a very different approach PACoR was developed. PACoR is a process algebra based formalism for specifying timed systems.

In cooperation with the programming technology research unit an approach which analyses the Worst Case Execution Time (WCET) of safety critical java programs has also been developed. A model of the tasks detailed behavior is generated directly from the java object code, which entails a very detailed analysis. As part of the SENSATION project there has also been a focus on modeling mixed criticality systems running on battery power. All of the above mentioned research utilizes some version of Uppaal for the verification of the schedulability models. Effort has also been put into trying to tackle the fundamental question of what a timing anomaly is. Research has also been conducted on a schedulability model with a discrete division of time. Finally a result shoving the decidability of a certain class of adaptable task automata has also been obtained.

**Hybrid Systems:** Embedded software for cyber-physical systems interacts with control systems. Thus their design requires an understanding of hybrid systems, and the interdisciplinary area between control theory and computer science. Research in this area continues with investigation of clear links between the formalisms of the two disciplines through the use of linking predicates that generalises the conventional simulation relations. This research is

a collaboration with Professor Olderog, Oldenburg, Germany and Professor Wisniewski at Automation, Aalborg University. On a tool side, there is work that has focused on invariant checking through the use of (compositional) barrier certificates with timed automata for verification.

**Safety:**  In a systems setting, safety is assessed through the likelihood of some malfunction that causes serious damage to the environment where the system is deployed. The assessment is often supported by analyses of the expected reliability of the system under given component fault probabilities. The SMT tool has here been demonstrated in a systematic procedure to assess reliability of systems, where fault models are derived systematically from Fault-Tree specifications.

**Biological Systems:**  The analysis of biological systems comes as a natural application of statistical model checking (SMC) techniques as the models are no longer limited to capabilities of symbolic techniques and the syntax is enriched with stochastic primitives. In particular we developed two distinct approaches for modeling bio-chemical reactions: Discrete counting of molecules based on reaction rates modeled as exponential rate of transition firing, and continuous dynamics of chemical concentrations modelled as ordinary differential equations. Both techniques have been demonstrated on a genetic oscillator responsible for the circadian rhythm in the brain. Translation utilities have been developed for biological models in ANIMO, XGMML and the SBML (Systems Biology Markup Language) formats and applied on hundreds of models available on SBML.org page.

**Web Services:**  The work on web services modelling and verificaiton was mainly carried out in the connection with the PhD thesis of Saleem Vighio. The main focus was on automatic analysis of protocols from WS-Business Activity. WS-Business Activity specification defines two coordination protocols in order to ensure a consistent agreement on the outcome of long-running distributed applications. We used the model checker Uppaal to analyse the Business Agreement with Coordination Completion protocol type and the analyses showed that the protocol, as described in the standard specification, violates correct operation by reaching invalid states for all underlying com-

munication media except for the perfect FIFO. We further generalized the model, and developed an approach for practical and automated verification.

**Home Automation and Smart Grid:** Smart Grids consist of traditional power distribution networks extended with distributed energy resources (DER) – including buildings and homes. In order to control Smart Grids in an efficient manner, all constituent components must be interconnected in a way that allows for monitoring and control at all levels of granularity. Standards, protocols and architectures for such infrastructures are currently under development, and there is a need to model, analyze and do experiments with new proposals. To this end, the following contributions for building systems have been made in the period – based on previous experiences with modelling methodology and applications of timed automata.

A flexible middleware architecture and infrastructure (Homeport) has been developed. It supports existing wireless standards and allows for easy integration with new proposals. On top of this, a full tool chain based on timed automata and Uppaal has been proposed and implemented at a prototype level. Applications of the tool chain in the home automation domain include a framework for analysis of feature interactions and a user interface for the definition and analysis of home device scenarios. Furthermore, an architecture for the exchange of flexibility market bids between DER (including buildings) and the energy market has been developed.

**Network Monitoring:** The aim of this work is to use network monitoring techniques to create statistical models of internet-work traffic and use these models to understand and enhance network performance, security and reliability aspects, thus supporting network management and operation. In several cases, we have applied flow-based techniques, and as part of this work written a comprehensive tutorial.

By combining flow-level measurements, minimal efforts on packet captures, and an analytical traffic model, we have developed a new efficient and accurate method for estimating required network link capacity. Internet traffic is often assumed to fit a Gaussian distribution well. We have re-validated this assumption despite the changing applications of the internet like social networks, clouds and video streaming websites. However, we also found that

traffic bandwidth is a more reliable indicator of Gaussianity than number of users for high-speed links. Other applications are network based monitoring of the availability of cloud services, and learning of whitelists in SCADA networks.

In another line of research, we have shown that malicious traffic is not evenly distributed across the internet, but is concentrated in areas. We offer a precise definition of such "bad internet neighborhoods", and investigate how these are spatially (geographically, sub-networks, and ISPs) and temporally distributed with the aim of improving blacklisting and attack prediction.

### 4.5.4 Other

**Higher Education Research:** There has been work on education research. Most of this has stemmed from experiences with teaching activities conducted within the department. Some of the work considers flipped teaching approaches that encourage active learning strategies as applied in courses on *Syntax and semantics* and *Computability and complexity*. Other work considers how the changes in the curriculum since 2010 have influenced problem-based learning, which is the core pedagogical approach used at Aalborg University.

### 4.5.5 Scientific Output and Impact

**Scientific Output** In the period, the unit has produced 63 journal papers (46%, 46%, 8% resp. in rank A, B, C), and 175 peer-reviewed papers in international conferences and workshops (22%, 53% 25% resp. in rank A, B, C), thus most of these appearing in recognized and high quality outlets. The journal rankings A, B, C aim to be consistent with the official BFI rankings. Further, 30 other publications have been written, including two (Danish) industrial dissemination books, 6 book chapters, and 11 Ph.D theses.

The number of journal publications have more than doubled from 29 in the period (2006-2010) to 63 for the current period. The number of conference and workshop papers have increased from 140 (2006-2010) to 175. We conclude that the amount of publications have increased, and that there also in an increasing focus on producing journal publications.

**Towards Industry**   DES has since 2002 been conducting industrial dissemination of research results through the regional ICT competence center CISS. This serves the purpose of contributing to industrial innovation, to competitiveness, and to stimulate further industrial collaboration. In the period 2011-2015 this activity has been funded by a number of national and international dissemination projects, where the central projects are CISS Connect (regional dissemination), InfinIT and ITOS (national dissemination) and a large number of ARTEMIS projects (international dissemination). These projects often inclued demonstrator work packages and generates a number of so-called *demonstrators* which have been developed and presented at industrial exhibitions together with regular industrial collaboration results. The demonstrators include among others home automation scenarios and production plants mock-ups. Also, dissemination has been pursued through handbook material and through the development of industrial courses that have been taught both as stand-alone courses and as parts of accredited life-long learning activities. Researchers from other units of the department have participated in several of the national projects on dissemination and applied research. Below, we outline the most important of the projects and activities:

- INFINIT: As grant holder, DES/CISS coordinates the project activities and is also in charge of one of the two projects offices. Industrial dissemination is carried out through a series of national events (INFINIT Summits) including keynote talks, panel sessions, focused parallel tracks and technology demonstrations. Also, CISS is leading several of the special interest groups of INIFINIT and has also been in charge of a number INFINIT mini projects - each involving at least two companies.

- ITOS: As a main result, the project has developed a Field Book describing different approaches to the development of embedded software. The project is carried out in close collaboration with 30 industrial companies, and the Field Book has been presented at the conference 'Embedded Everywere', October 2015, with participation of more than 400 industrial persons.

- Software Corridor, CISS Connect, CISS Innovation Factory, MoDES SME programme: These projects (coordinated by CISS with a total funding of more than 13 MDKK) have been focused on demonstrators,

embedded platform prototypes and staff competence upgrade together with more than 20 companies. The funding has mostly been from the European Regional/Social funds, but also by invitation from the Danish Strategic Research Council as a follow up on a successful strategic, national research project running in the previous evaluation period.

- ECSEL/Artemis JU projects: As a follow-up on the Danish participation in the ECSEL/Artmis programme, CISS has been active in setting up international consortia which have resulted in six large industrial dissemination projects with a total (CISS) funding of more than 15 MDKK and resulting in prototype systems as well methodology demonstrators.

- Other projects: In addition to the above industrial dissemination, CISS has coordinated a large number of minor, focused dissemination projects involving typically one or two SME's and typically developing prototypes or doing feasability studies.

**Towards public authorities**  As part of the interaction and dissemination towards public authorities, DES has participated actively at the 'Peoples Meeting' series at Bornholm twice, with contributions to the Ministerial booth for Innovation Networks (2013) and to the Northern Jutland booth on Smarter Regions (2015). DES has also interacted with the regional growth forum on future innovation policies using 'Smart Specialization' - through the EU CLINES project. Furthermore, a DES member has been appointed 'Digital Wisemen' of ATV, participating in the public debate on the future digital society.

**Dissemination towards Students**  We consider the dissemination of our research results and its integration into students' curricula as an important aspect of our research. In the past we have written books for our courses aiming at undergraduate and graduate students, and in this period the lecture book "Reactive Systems: Modelling, Specification and Verification" (published by Cambridge University Press) has been complemented with CAAL, a new web-based tool for modelling and analysing concurrent processes expressed in the CCS language.

Members of the unit have been consistently elected best teachers of the year in the department, at faculty level and even at the university level, providing a strong witness of the units commitment to dissemination to students.

Since 2008 and until 2014 we have offered an Elite M.Sc. Education in Embedded Software, accepting yearly 5-10 students. This has allowed students to be involved closely in our several basic and industrial research projects and essentially serve as a pre-start of Ph.D. studies. In addition we are offering an international M.Sc. on Embedded Software.

ITOS is a collaboration with Federation of Danish Industries aiming at raising awareness and uptake of embedded system technology both in industries but also towards students. As such our unit have been active on the selection committee for the annual Embedded Thesis Award, as well as active in the embedded development student competition eFactor2014.

Members of the unit have been invited to serve as lecture at numerous international Ph.D. school including several appearances at ARTIST PhD schools and the NATO school Marktoberdorf. Also, in 2011 we organized the international PhD school *Quantitative Model Checking* with more than 100 participants and 9 international lectures. Most recently, together with Technical University of Denmark we organized the EMSIG Autumn School on Cyber-Physical Systems with 35 participants and 8 lectures.

**Funding**   Table 4.1 shows the number of projects and grant size of projects that are granted in the period, including in the first row projects granted in the previous period that extends into the current. The unit has been very active and successful in obtaining external funding via several small and large grants for basic and strategic research, as well as industrial dissemination. The unit has been active in 60 projects with a total grant size of 162 M. DKK.

Despite that three larger grants ran out in the first part of the period (DaNES, MTLAB, Infinit), the unit has sustained a high level of funding. In the period, the unit has obtained new grants with a value for DES of 103.8 M. DKK, thus exceeding the previous 5 year period (65M. DKK) by 60%.

In year 2013 the funding was a bit lower due to the completion of the large basic research project, MTLAB. Despite this, we have managed to maintain a high level of funding for basic research with the grants for Sensation and Cassting, the continuation of IDEA4CPS, and especially the Advanced ERC

| Year | Projects | Total Grant | Significant Projects |
|------|---------|-------------|----------------------|
| -2010 | 23 | 58.2 M. DKK | DaNES, MTLAB, Infinit |
| 2011 | 12 | 22.1 M. DKK | IDEA4CPS, MBAT, Encourage |
| 2012 | 9 | 21.3 M. DKK | Totalflex, TREsPASS, Crafters, Sensation |
| 2013 | 7 | 9.4 M. DKK | Arrowhead, Cassting |
| 2014 | 4 | 13.6 M. DKK | IDEA4CPS2, Infinit2 |
| 2015 | 5 | 37.4 M. DKK | ERC (LASSO), DiCyPS, ASAP |
| Sum | 37 | 103.8M. DKK | |
| Total | 60 | 162.0 M. DKK | |

Table 4.1: New Projects by year

| Project Type | New Projects | Total Grant | Av. Grant/Project |
|--------------|-------------|-------------|-------------------|
| Basic Research | 9 (4 intl.) | 48.2 M. DKK | 5.4 M. DKK |
| Strategic Research | 15 (9 intl.) | 41.7 M. DKK | 2.8 M. DKK |
| Dissemination | 13 (2 intl.) | 13.9M. DKK | 1.1 M. DKK |
| Total | 37 (15 intl.) | 103.8M. DKK | 2.8 M. DKK |

Table 4.2: New Projects by Type

grant LASSO awarded towards the end of the period.

Table 4.2 show the distribution and grant size of new projects obtained in the period categorized by their degree of basic research, strategic research, and industrial dissemination. It can be seen that all categories have a significant portion of funding, but we notice a decrease in funding for dissemination compared to a very high level in the past period. We also notice that part of the large increase in funding came at a cost of an increased number of relatively smaller grants for strategic research.

In conclusion, the current and near future funding situation is very good, although it will be challenging to surpass.

### 4.5.6 Scientific Services and Recognition

Members of the unit have proved substantial services to research community that indicate significant leadership and recognition. This include serving on the editorial boards of the journals FMSD, TCS and NJC as well as the newly established OJCS. Members of the unit has also served PC-chairs for

the conferences FORMATS 2011, JTRES2011, ICTSS2012, VMCAI 2015, and organized several workshops. In particular, the unit has been playing a leading role in co-organizing a number of international PhD schools including QMC and EMSIG. The high international recognition is also witness from the fact,that the unit is to organize 5 international conference and workshop in the next two years (WATA, ICT-Energy, RP, NWPT and MFCS). Moreover, members of the unit serve on the steering committee on a number of leading international conferences including ETAPS, TACAS and CONCUR, as well on EMSIG the Embedded Systems Special Interest Group. This is in addition to a very significant number of program committee memberships.

Within the reporting period 2011-2015, one member of the unit, who is holding Honary Doctorates from both Uppsala University and LSV Cachan, received as the first non-American the CAV Award, became member of the Academia Europaea, and most recently won the prestigious ERC Advanced Grant. In addition he is (has been) serving as member of Villum Kann Rasmussen Annual Award committee, the EATCS annual award committee as well as the CAV Annual Award committee. During the period he was relected Danish national expert for the ICT committee, served as advisory board member for the Computer Science Department at Copenhagen University (DIKU), as well as vice-chair for ATV group on Electro- and Information Systems.

Another member has received the prestigious Sapere Aude Postdoctoral position, and achieved most recently the highest Danish degree Doctor Scientiarum. Also recent, a Velux Visiting Professorship was granted.

A Ph.D. student of the unit obtained the ElitForsk travel-stipend (as only 1 out of 3 Pd.D. students from Aalborg University), and was later awarded the Huawei Visiting PhD fellowship (1 out of 11 Ph.D. students in Denmark). Also, M.Sc. students of the unit are performing at a high level witnessed by several publications during the study, and by winning the Best Embedded Systems Thesis (2913) offered by Federation of Danish Industries.

In the reporting period, members of the unit received several other awards including two consecutive Silver Medals at the Model Checking Competition, best paper award at FACS13, and best paper award nomination at TACAS13. Also, several conference publications from the unit have been invited for journal publication in special issues.

## 4.6 Own Evaluation

In our opinion the research activities carried out by the DES unit in the period 2011-2015 have been very satisfactory both in quantity and quality. The number of publications have increased significantly totalling 271, including 63 journal papers and 175 conference/workshop papers.

We consider these figures as an indication of the high quality and impact of our unit, which is further witnessed by the ERC Advanced Grant recently won by Professor Kim G. Larsen (2015), which also achieved the prestigious CAV Award for the Uppaal tool (2013). Also witnessing the excellence is the Sapere Aude Fellowship held by Radu Mardare (2010-2012), the ElitForsk travel grant awarded to Andreas Dalsgaard (2013), as well as the Velux Visiting Professorship offered to Axel Legay from INRIA, Rennes (2015).

**Overall Strategic Goals** In the previous period (2006-2010) the unit was building up high international visibility within the area of Embedded Systems, with strong involvement in national and European intiatives (e.g. the NoE ARTIST) and with the national ICT competence center CISS (Center for Embedded Software Systems). In the present period (2011-2015) we have maintained this stronghold witnessed by several national and European intiatives including ARTEMIS projects – e.g. ENCOURAGE (DES coordinator) and MBAT (DES WP leader) – as well as continued coordination of *the* danish ICT Innovation Network InfinIT.

In the present period (2011-2015) the scope of our research has been succesfully extended to the emerging area of Cyber-Physical Systems (CPS), both in terms of basic research and industrial collaboration. This is witnessed for instance by the ERC Advanced Grant recently won by Professor Kim G. Larsen, the Sino-Danish basic research center IDEA4CPS, the Willum-Kahn Rasmussen Center of Excellence, MT-LAB, the DFF—FNU project ASAP, and the new Innovation Center DiCyPS all being related to CPS. During the period 2011-2015 the DES unit has substantially increased the joint research collaboration with other units of the department, as well as with other groups at the university. The long-standing collaboration with Department of Electronic Systems, has been maintained at a somewhat reduced but stable level.

The DES unit has a strong focus on tool-development serving as vital link between the foundation research and the applied/methodological research of the unit. Undoubtedly, the continuous improvement and support of the Uppaal tool is a key factor behind our succesful industrial collaboration and our international leading position with respect to verification and validation, witnessed by the CAV 2013 awarded given to Professor Kim G. Larsen, and the tool TAPAAL being runner-up for two consecutive year in the Model Checking Contest.

Within the period 2011-2015, a major effort has been in developing and implementing a new, highly scalable statistical model checking engine for Uppaal, and its application in a variety of domains. The grant obtained from DeiC to our 576 core Model Checking Cluster (MCC) has already lead to scalable, parallellel implementations of optimal reachability for Uppaal.

Within foundational research, significant results have been obtained on quantitative specification theories as well as complete axiomatizations covering both timed, weighted and stochastic systems. For stochastic systems, substantial effort towards identification of suitable behavioural metrics, their logical characterization and algorithmic classification has lead to a series of high-impact publications. Our work on modal transition systems, priced time automata and most recently energy games is by now widely recognized[2].

With respect to our methodological research, the effort on real-time systems – real-time Java and schedulability analysis – have resulted in several publications, whose impact is witnessed by professor Anders P. Ravn currently serving on the standardisation committee for Safety Critical Java. Also *Homeport* a flexible middleware and infrastructure for home-automation has been developed through strong industrial collaboration and with links to Uppaal.

The external funding of 103.8 M. DKK exceeds that of the previous period with 60%. In particular, the newly started projects LASSO (ERC Adv Grant) and DiCyPS (InnovationFund) secures significant funding to basic as well as industrial research in the next period.

Finally, we are sad that the eight Elite M.Sc. Educations at Aalborg University accredited by the Ministry of Education have been abandoned by the university. Our own Elite M.Sc. Eduction on Embedded Software was extremely succesfull in involving Master Students in the several industrial development

---

[2]invited article in CACM

and basic research projects conducted by our unit. Moreover this allowed an early and smooth introduction to the several Ph.D. positions our unit is constantly offering.

## 4.7 Plans

In the following we detail the plans for future research in terms of research directions, possible collaborations and projects for the unit, as well as measure to be taking for future management and development.

**Research Directions**   We will strive to maintain our strong international collaboration and leading role within Embedded and Cyber-Physical Systems, and our high international visibility and recognition in the areas of model checking, logics and semantics for concurrent systems.

We will strengthen further our focus on tool-development. This will sustain the bridge between foundational research and applications. In order to scale to complexity of CPS, techniques from other areas such as SMT and machine learning will be exploited, as well as taking full advantage of the recently established MCC cluster.

The current trend in our foundational research in replacing Boolean verdicts with metric notions and quantitative results will be expanded. We will invest substantial new effort in research on cloud computing, (wireless) networks and IoT.

**Research Collaborations and Projects**   We want to maintain current balance between basic research and application-oriented projects. Even though we have a number of large, long-term projects, they are typically less attractive to SMEs. We therefore want to maintain a fair number of smaller, short-term projects despite the increased overhead in managing these.

Our stong international collaboration must be sustained. We have been extremely succesful in the ARTEMIS program, but significant effort is required to ensure similar performance in the successor ECSEL, though we are already involved in two new projects ECSEL MANTIS and ENABLE-S3. We also intend to be partner/coordinator of future FET projects: participation in the

FET projects SENSATION and CASSTING have been very rewarding in terms of collaboration with top-research groups in Europe. For similar reasons our active participation in (future) COST actions must be continued.

In terms of national collaboration, the directing role of the Innovation Network InfinIT provides us several possibilities. Also, the ongoing projects PPP and Trespass ensures collaboration with colleagues at DTU and ITU. In collaboration with other units of the department we want to persue new collaboration with SDU and ITU within the area of Smart Society.

Within the department, significant collaboration with other units is already planed in the context of various projects, e.g. DiCyPS, PPP, LASSO as well as the Innovation Network InfinIT. Also, the ERC Advanced Grant provides an opportunity for closer collaboration with researchers from the MI unit.

Within the university, we want to pursue new, strategic collaborations with Department for Electronic Systems within WSN, IoT and Security.

**Research Management and Staff**    To minimize the overhead on researchers to engage in industrial collaboration with SMEs, the CISS VIP staff must be increased, and should be involved more directly in research projects.

We wish to identify and establish better ways of managing our several and interconnected projects.

Furthermore, we wish to strengthen the alignment and synergy between MSc/PhD/post-doc projects with industrial research projects, and initiate exploitation of existing financial sources for industrial PhDs/post-docs. Finally, we encourage members of the research unit to plan sabbaticals in order that new ideas from the outside may be brought in.

## 4.8  Committee Evaluation

### Observations

The Distributed and Embedded Systems group is a world-class group. It is involved in a broad range of activities from semantic foundations through tool development for verification and validation to real-world applications. The group is making excellent contributions across the whole spectrum of

activity; this is internationally recognized by prestigious awards such as:

- The ERC Advanced Grant LASSO

- The 2013 CAV Award for Uppaal - the first time that this award has been granted to a non-US team

- The ranking of the "Uppaal in a Nutshell" as the 9th most influential paper in Software Engineering since 1972

- Best paper awards, medals and other awards to Associate Professors

The h-index of Kim Guldstrand Larsen is outstanding and places him among the top echelon of researchers in this area; his h-index is higher than some Turing Award winners in cognate areas. It is also pleasing to note that some of the Associate Professors also have high h-indices for their career point. The world-leading status of the group is further illustrated by the impressive collaborations with other international leaders such as Luca Cardelli, Dexter Kozen and Prakash Panangaden. In addition to these collaborations with established researchers, Professor Larsen also has important collaborations with younger ERC laureates (Patricia Bouyer and Jean-Francois Raskin).

The group has published well during the review period with 175 conference papers - 75% of which are in A and B venues - and 63 journals - 92% of which are in A and B venues. At the foundational level, there are important contributions to compositional specification of real-time systems, complexity theory of modal transition systems, timed automata and timed-arc Petri Nets, Energy Games and Security. Particularly noteworthy are the new areas of work on metrics for measuring the differences between transition systems and logics - these have both involved fruitful international collaborations. In the verification and validation work the major developments have involved the addition of statistical model checking to the Uppaal tool set and work on the synthesis of controllers and strategies which has now been implemented in UPPAAL-STRATEGO. The introduction of machine learning techniques into the model-based development work has shown some good early results and is a promising theme to develop further in the next period. Noteworthy contributions to methodology and applications include the work on Real-Time and Safety Critical Java and the application work on Home Energy systems

and Smart Grid. The group has secured 37 new grants to a total value of DKK103.8M. It is notable that a significant proportion of this funding, particularly from ARTEMIS and the Danish Strategic Council, is associated with the Centre for Embedded Software Systems and INFINIT - activities which are focused on industrial pull-through and exploitation of the research results. Compared to the previous review period, the group has significantly expanded its activities to encompass cyber-physical systems - an area that is likely to be of growing importance and which will offer many opportunities for multi-disciplinary projects and large-scale funding.

The major strength of the group is the people; not only the group leader but the strong group of more junior academic staff and an excellent group of support staff. The broad span from foundational work to applications is also unusual in such groups in other universities and is a considerable strength of DES. The profile and reach of the group is enhanced by its dissemination activities but also the engagement of senior staff in policy-related activities at national and European levels.

One area of concern is the relatively low number of PhD students during the period (14) and the number of non-completions (3) - although we recognize that, because of the employment status of PhD students, the group already spends half of its external funding budget on PhD students and has made a decision to give more emphasis to postdoctoral positions. However, it seems that the group is not making use of all possible sources of PhD funding such as the industrial PhD scheme. The group suffers a little from lack of visibility to prospective students; this is quite common for groups in this area but DES is better placed than most because of the engaging application areas that it has and maybe these should be used more for student-oriented publicity.

There are growing interactions with other groups in the Department and other departments in the University and there are likely to be many further opportunities in the future. The group has an excellent research base to respond to the new challenges that the increasing digitalization of society is likely to bring.

Given the emphasis on dissemination, the number of tools produced by the group will increase and the support for development and maintenance of such artefacts is a significant burden on the group. This aspect is an important part of their reputation and the Department must find a way to provide

appropriate technical support.

**Recommendations**

- Continue to develop all areas across the spectrum from foundations through to applications. The expansion to consider the role of machine learning in model checking and the continued development of work on Cyber Physical Systems are important developments.

- Consider the introduction of a mentoring scheme for PhD students - separate from the scientific supervision.

- Make better use of all funding schemes for PhD students and consider developing publicity material targeted at prospective students and emphasizing some of the applications work.

- Continue the reflection started with the SWOT analysis and decide which threats are significant and should be treated.

- Work with the Department and other groups to develop a strategy for tool support.

# Information Systems

# 5 Information Systems

## 5.1 Executive Summary

The research unit Information Systems focuses on design, development, and application of interactive information and communication technologies. The research is empirical and solution-oriented and pursued in relation to the international research communities of systems development and human-computer interaction.

The unit has successfully increased its quality and impact of publications and improved its internationalization and outreach. It has been less successful it continuing its previous strategy in attracting research funding and has had to change strategy. The unit has suffered significantly from the university's economic downturn during 2013-14 and has despite large investments in teaching including creating a new education only recently been able to begin to recover financially.

## 5.2 Profile

The research unit Information Systems focuses on design, development, and application of interactive information and communication technologies. The research unit has a long tradition for empirical and solution-oriented research with a focus on processes, products, and theory. The process-oriented interest contributes to methodological knowledge of agile development, design, and innovation of interactive systems. The product-oriented interest contributes to knowledge on the specifics of interactive systems and applications in work and leisure and in organisations, teams, as well as other social arrangements. The theoretical knowledge interest is on innovation, design, and

the application of interactive systems in practice. This focus is pursued in relation to the international research communities of information systems and human-computer interaction.

The unit has a strong element of empirical research involving people, whether it is the individual users and developers of software applications, groups of users and teams of developers, or the use and development in a larger social or organisational context. One part of the empirical research is experimental in the form of design, construction, and evaluation of particular interactive prototypes to explore and solve specific problems. Another part is directed at the professionalisation of systems development. This takes place through close collaboration with companies in improving their working practice and encompassing agile processes and the integration of agile development, usability engineering, and interaction design.

The IS unit has recently established the Research Center for Socio+Interactive Design (S+I). The purpose is to create and foster synergies between systems development research and HCI research by engaging in joint research and by connecting with software development and design companies.

The unit does research activities within three main areas: interaction design, usability engineering, and systems development.

### 5.2.1 Interaction Design

The interaction design research area is concerned with the design of everything that is both digital and interactive, and specifically with designing interactive products to support the way people communicate and interact in their everyday and working lives. The unit's research into interaction design is driven by the design and implementation of concrete functional prototype systems and technology probes, and empirical studies of these in use. Prototype systems are designed and built to fit specific use contexts, explore new technologies or interaction techniques, or to be purposely provocative in order to explore user reactions and behavioral change. Empirical studies are done through longitudinal real world deployment, or controlled experiments, depending on the research question and aim. Closely related to the technical and empirical work, the research into interaction design also focuses on research methodology, such as developing new methods and techniques for

designing and studying interactive systems.

During the period from 2011 to 2015, the unit's interaction design research activities can be divided into three streams: ubiquitous and mobile computing, interaction techniques and changing behaviour with technology.

**Ubiquitous and mobile computing:** The research stream on ubiquitous and mobile computing is about designing new interactive systems for people's everyday and working lives, enabling them to do things that they couldn't do before. The research is driven by the opportunities offered by the increasing prevalence and maturity of ubiquitous and mobile technologies in society, such as situated displays, smartphones, smartwatches, and Internet of Things, allowing for fundamentally new applications of interactive systems. This research has a long history in the unit, dating back to its work on mobile technologies in the early 2000s, and has been very successful in terms of quality, impact, and visibility.

**Interaction techniques:** The research on interaction techniques is about exploring new ways for people to interact with computerized systems. This research is driven by the continuing emergence and improvement of new input technologies for human-computer interaction, such as computer vision, multitouch and context sensors, which enable radically new interaction designs. This stream of research was initiated in the late 2000s, and has gained a strong momentum in the unit within the last 5 years.

**Changing behaviour with technology:** The research stream on changing behavior is about using interaction design for the specific purpose of making people act differently than they did before. This research is driven by the "grand challenges" of improving human life through, for example, improved environmental sustainability or improved health. This stream of research has been initiated in the unit within the last 5 years, facilitated by a number of collaborative research grants, and is showing great potentials for the future.

### 5.2.2 Systems Development

The research in systems development seeks to understand and improve the organizational practices pertaining to agility, innovation, and management. The research involves empirical investigation of organizations that either develop or use software-based systems and applications. Research efforts engaged with practice in organizations are based on qualitative and proactive research methods such as case studies and action research. The research results are published in outlets from the fields of information systems, and software engineering.

**Agility:** The focus on agility attends to the rapid changes that are circumstantial to most systems development projects and how this has motivated the use of agile development methods in software companies. The purpose is to empirically understand and improve the emerging practices from combining agile methods with distributed development, usability work, and documentation-driven organizations.

**Innovation:** The research stream on innovation attends to the pressing needs for creativity and high-value solutions at the level of the software developing team and up to the organizational adoption. The purpose is to develop methods, tools, and techniques as well as infrastructures and conceptual models to support teams and organizations in creating valuable solutions throughout the span of systems development.

**Management:** The research stream on management attends to the human and social enablers for the successful development of new software based systems that are valuable in organizational practice. The purpose is to contribute with empirical understanding of how to manage: project-level processes, knowledge, and risks; company-level process improvement, outsourcing, and benefits; and e-government within and across municipalities.

### 5.2.3 Usability Engineering

The research on usability engineering contributes to understanding and improving current usability practices within software development projects. This entails the development and assessment of specific usability evaluation methods, understanding and improving design activities after conducting usability evaluations, and finally how usability engineering can be further integrated into software development practices.

**Usability evaluation:** Our research in this area entails methods for conducting usability evaluations. In particular, the activity has focused on how to reduce efforts in evaluations through new ways of conducting evaluations and analysing usability data. This also includes methods that address the challenges of conducting usability evaluations over a distance, e.g. when evaluating web applications.

**Redesign:** This stream of research deals with the interplay between usability evaluation results and redesign activities. Redesign activities focus on improving the level of usability in systems that have been evaluated. Our work contributes to understanding the interplay between evaluation and redesign and to assess different approaches for creating redesign proposals in practice. Within this stream we have also studied the practical application of various materials (e.g. guidelines), which were developed to enhance a usability focus when developing public IT systems. Such studies include redesign proposals to improve the utility of supporting materials when applied by practitioners.

**Integration into software development:** This research stream deals with the limited application of usability engineering methods in software development practice. Although usability engineering activities are considered relevant in the software development process, there are obstacles that limit complete, effective and efficient integration into software development. One substream of research deals with identifying barriers for (further) integrating usability engineering into organizational settings, e.g., survey studies. The second substream represents intervention studies of possible solutions to reduce identified barriers.

## 5.3  Staff

### 5.3.1 Current Staff

**Professors:** Peter Axel Nielsen, Jesper Kjeldskov and Jan Stage
**Associate professors:**  Ivan Aaen, Mikael B. Skov, Jeni Paay and John Persson
**Assistant professors:** Anders Bruun and Dimitrios Raptis
**Ph.d students:** Jane Billestrup, Nis Bornø and Rikke Jensen
**Research Assistant:** Michael Kvist

### 5.3.2 Staff Development

Figure 5.1 is a graphical overview of scientific staff during the five-year period.

Peter Axel Nielsen and Jan Stage have been professors throughout the period. During the period we have had several promotions on all levels of the academic ladder. Jesper Kjeldskov was promoted from associate to full professor. Jeni Paay and John Persson went from assistant to associate professors while Anders Bruun, Dimitrios Raptis and Lise Heeager went from Ph.D. students to assistant professors. Jane Billestrup and Rikke Jensen were initially employed as research assistants, and are now Ph.D. students.

Members of staff have also left the research unit during the period. Jeremy Rose left for a position as professor at the University of Skövde, and Lise Heeager left for a position at Aarhus University. Jeremy Rose has, in a transition period during 2014, been employed 20% (dashed line in Figure 5.1). John Persson left the unit temporarily in 2014 for a position at another department, before returning in a position as associate professor in 2015. Henrik Sørensen, Fulvio Madriz, Frank Ulrich, and Karsten Jahn left the unit after successfully completing their Ph.D. studies.

## 5.4  Goals (2011-2015)

The following goals for the period 2011-2015 were set as part of the previous research evaluation (January 2011).
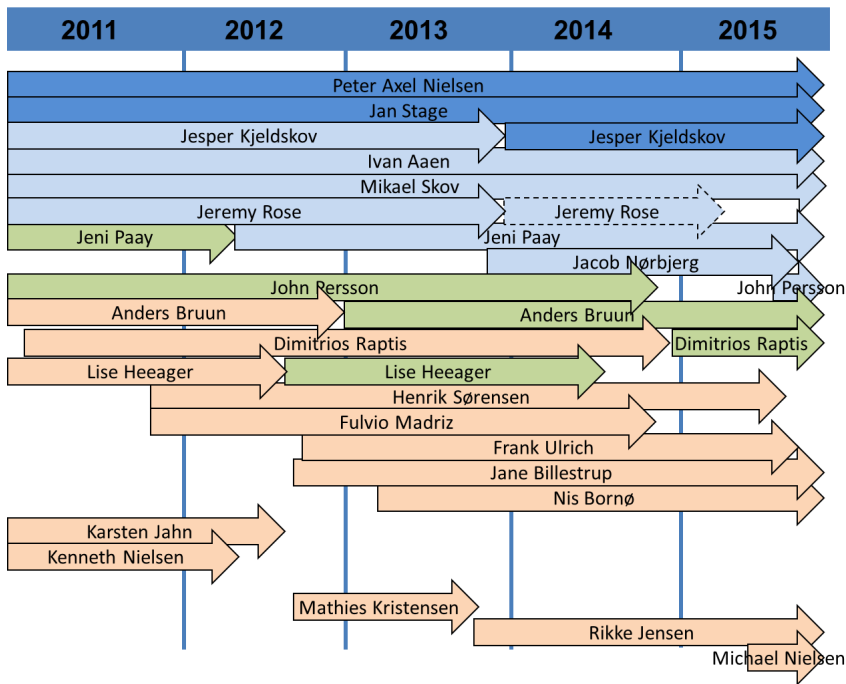
Figure 5.1: Scientific staff 2011-2015. Key: Tenured staff (dark blue = professor / light blue = associate professor); assistant professor/postdoc (green); Ph.D. student/research assistant (beige).

**Quality and impact of publications:**    The goal was to increase quality and impact of our research by aiming at publishing in outlets of higher standard, and with larger audiences (A-level conferences and journals). We intended to pursue this strategy even if it meant a reduction of the total number of publications (information about ranking of journals and conference can be found in the appendix for the IS unit).

**Strategic research project:**    The goal was to establish a strategic research project as a setting for collaboration across the unit, and to exploit our research competences. The project was to focus on the area of digital public service.

**Extend international research collaboration:**  The goal was to extend our collaboration with existing international and national partners, leading to closer collaboration and joint funding. In particular we wanted to capitalize on our previous investments into networking and forming partnerships.

**Research education:**  The goal was to maintain a high level of involvement in research education through PhD courses, supervisor workshops, national collaborations to improve PhD education, and by developing the learning environments we provide for our PhD students and postdocs.

**Recruit more staff:**  The goal was to recruit more scientific staff on post-doc/assistant professor level, as well as associate professor level. We also had a goal of recruiting an increased number of PhD students.

**Research funding:**  The goal was to maintain the level of national funding, and to gradually increase the amount of EU funding attracted by the unit.

## 5.5  Activities and Results

### 5.5.1 Interaction Design

The IS unit's research activities in interaction design are divided into three streams.

**Ubiquitous and Mobile Computing**

The unit's research on interaction design for ubiquitous and mobile computing has dealt with the design, implementation, and use of interactive systems in response to the new opportunities offered by ubiquitous and mobile technologies. The research activities on mobile computing have extended the unit's previous research in this area, while the activities on ubiquitous computing represent the unit's more recent development toward addressing the opportunities and challenges of an emerging landscape of very varied interactive technologies, digital ecosystems and the Internet of Things.

119

In the area of mobile computing, we have conducted experiments into different physical form factors of mobiles, such as their screen size and perceived "coolness", and on their user experience. This has been done as part of a PhD project involving collaboration with the University of Patras, Greece, and with local software industry funded through Cassiopeia Innovation. Our work on mobiles has also involved technical research activities where we have, for example, investigated the use of smartphones for mobile augmented reality. This was done through implementation and studies of use for a number of specific applications, such as on-site architectural planning, and urban social networks. Another activity has revisited our previous work on research methodology in the area of mobile HCI, specifically following up on two highly cited papers from 2003 and 2004 in relation to them being awarded "most influential research" in 2013 and 2014. Finally, the research activity on mobile computing has included two years of work on a Dr.Scient. (higher doctorate) dissertation on "Designing Mobile Interactions" (see https://en.wikipedia.org/wiki/Doctor_of_Science)

In the area of ubiquitous computing, we have continued our research into technology design and use in domestic settings with a particular focus on kitchens and the activities of cooking and eating together as facilitated by interactive technology. This has involved continued close collaboration with the University of Melbourne, and Microsoft Research Cambridge. We have also initiated new research activities into the opportunities and challenges of digital ecosystems, including empirical and conceptual work on multi- and cross-device interaction. In this work we have, for example, developed a conceptual framework for interaction with digital ecosystems (4C), and conducted a large digital ethnography of Apple's Continuity. Parts of this have been done in relation to a PhD project involving collaboration with Microsoft Research Cambridge, and postdoc research in relation to the TotalFlex project. Through deliberately studying an extreme case of ubiquitous computing, another activity has investigated the user experience of living with implanted networked technology through a longitudinal field study of cardioverter-defibrillator patients. Taking a research methodology perspective in relation to these activities, we have continued our research into techniques for studying technology use in places that are difficult to reach by traditional observational means. Specifically we have done this by exploring the approach of digital ethnography by, for example, using video recordings posted on YouTube as a data

source.

In the period 2011-15, this research activity has produced 40 publications of which 4 are A-level journal articles, and 12 are A-level conference papers (see the IS appendix for information about ranking of journals and conferences). Two PhD students have graduated from this activity, and it has produced one Dr.Scient. degree.

**Interaction Techniques**

The unit's research on interaction techniques has partly grown out of related activities in the mobile and ubiquitous computing activity and hence, we have had a primary focus on techniques for mobile technologies or in ubicomp environments. Here we are concerned with the design, programming, and evaluation of techniques for interaction on such mobile or ubiquitous technologies. One of the activities received funding namely the Tele-presence in 3D Blended Spaces project funded by the Obel Foundation. As several of the projects under interaction techniques are highly technical, this research activity has closely collaborated with students with advanced programming skills, e.g. software engineering or computer science students.

For interaction techniques for ubiquitous computing environments, we have extended our previous work on blended interaction spaces for collaborative work. Here we focus on video conference systems and a common problem with such video conference call technologies is that they do not facilitate eye contact due to the offset of the camera and the screen, e.g. on desktop or laptop computer. We designed and implemented EyeGaze that uses multiple Kinect cameras to generate a 3D model of the user, and then renders a virtual camera angle giving the user an experience of eye contact. We found that EyeGaze provided additional value in terms of e.g. involvement, turn-taking and co-presence.

More recently, we have started to investigate interaction techniques for cross-device interaction, that is interaction that takes place over or across several devices, e.g. mobile devices. Cross-device interaction is currently receiving a great deal of attention within HCI as people often have several devices at their disposal, and devices are being used in social situations and settings. Our work has primarily focused on cross-device interaction techniques for

mobile devices through development and deployment of prototype systems. For example, we developed JuxtaPinch, which was a prototype system that facilitates photo sharing across multiple mobile devices through different kinds of pinch interaction techniques. Related we have explored different interaction techniques for games on mobile devices and implemented and compared six different techniques for playing cards across mobile phones and tablets.

In the period 2011-15, this research activity has produced 11 publications of which 1 is an A-level journal article, and 8 are A-level conference papers.

## Changing Behaviour with Technology

The changing behaviour with technology activity has dealt with different aspects of addressing a growing interest in how we can utilize information technology in changing people's behaviour in different situations and contexts. In particular, we have had activities, funded research projects, and international collaboration on key focus areas of sustainability (electricity and water consumption) and smoking cessation. Both sustainability and smoking cessation have been important areas for research funding over the last year, and have received significant attention in the HCI research community.

The first focus within changing behaviour concerns sustainability research on electricity and water consumption. We have been part of the research project TotalFlex involving two other research units from our department, which focuses on flexible electricity consumption in private Danish households. A significant challenge for electricity distribution in the current Danish grid is the growing consumption and in particular high variances in peak demands. As a consequence, our role in this project has been to engage with private households (different kinds of families) and we have developed a number of technology probes in trying to investigate how these households can consume electricity in a more flexible manner - sometimes referred to as shifting. Our findings show that (1) people are generally unaware of when and how they consume electricity in their private homes, (2) people can learn about their own consumption and behaviour from visualizing electricity usage and forecasts, but they still find it challenging to shift consumption, thus change behaviour, and (3) eco-interaction might enable behaviour change as households can directly engage with e.g. heat pumps. Also and as part of this, we have focused on conducting longitudinal field studies (six months or more),

and we are currently finalizing a study on heat pump control. In addition to electricity consumption, we have had collaboration with the Interaction Design Laboratory group at the University of Melbourne, Australia, on water consumption, as water is a scarce and critical resource in Australia. This collaboration inspired and instructed our work on electricity consumption in Denmark, for example, how to apply mobile technology for sustainability.

Our second focus of changing behaviour concerns smoking cessation. Within this area, we have collaborated with people from Melbourne University, Australia, who have worked within this area for a number of years. As a part of this collaboration, researchers from the IS unit have visited Melbourne several times, and we have had master's thesis students visiting the lab as well. In particular, we have studied the use of apps for supporting smoking cessation, and achieved a honorable mention full paper at CHI 2015 where we illustrate how people are highly individual in their approach to quitting smoking and we identified a collection of empirically derived themes for tailoring smoking cessation apps.

In the period 2011-15, this research activity has produced 15 publications of which 7 are A-level conference papers.

## 5.5.2 Usability Engineering

The IS unit's research in usability engineering are in three streams: usability evaluation, redesign, and integration.

### Usability Evaluation

This stream emphasized studies on how to reduce resource demands for conducting usability evaluations. This entails new ways of analysing evaluation sessions and challenges of conducting usability evaluations for web development.

An emerging part of this activity deals with measuring UX continuously on the basis of psychophysiological measurements. This is done by, e.g. measuring emotional reactions by attaching a Galvanic Skin Response (GSR) sensor. A GSR sensor reveals peaks in emotional arousal, which indicates particular points of interest during an interaction sequence. This work has so far resulted

in one full-paper publication at Interact.

This activity also includes usability evaluation activities in relation to health informatics. We have several publications on the use of usability evaluation methods in healthcare settings and studies where we examine the use of such technology from the perspective of senior citizens.

In the period 2011-15, this research has produced 11 publications of which 1 is an A-level journal article and 3 are A-level conference papers.


## Redesign

This stream deals with the application of usability evaluation results for design and in particular redesign of a system that has been evaluated. There has been a number of activities, mostly in collaboration with IT development organizations. The typical approach has been that we have assisted IT developers in evaluating one of their systems and subsequently supported them in drawing up a redesign of the system.

In the period 2011-15, this research has produced 6 publications of which 2 are A-level conference papers.


## Integration

This stream focused on the conduction of usability evaluations in industrial practice. Particularly, the activity involved close collaboration with industry on adoption of methods and techniques for usability evaluation. The activity utilized a combination of field and laboratory experiments as well as action research and we applied both qualitative and quantitative measures.

One major externally funded project has contributed to this activity. The Web Portal Usability project (2009-2015) focused on the particular challenges of creating usable web portals through the creation of new methods for usability engineering. In addition, a number of externally and internally funded smaller projects have contributed to the activity.

During the period we coined the term "barefoot usability evaluations", which describes the situation where software development practitioners conduct usability testing and analyze results (as opposed to usability specialists). Previous research efforts have suggested to use developers as observers in evalu-

ation activities, but the barefoot approach takes this a step further as developers become more active.

In the period 2011-15, this research has led to 18 publications of which 2 are A-level journal papers and 2 are A-level conference papers. Two PhD students have graduated based on this activity.

### 5.5.3 Systems Development

The IS unit's research in systems development falls in three streams: agility, innovation, and management.

**Agility**

This research is directed at studying practice in software companies. Software companies are increasingly committing to agile methods and practices. This area of concern has been investigated through case studies and through action research. A series of case studies have addressed how usability work can be integrated into agile software development as the difficulties are theoretically hypothesised. This research aligns with usability engineering (section 5.3). The focus in this research activity is on agile development and the study of agile software teams, looking at usability work, how it can be integrated into the agile process, what the challenges are, and how teams and software organisations further or hinder its integration into agile practices. There is little in agile software development per se that hinders usability work in agile development, but large and late usability evaluation is not feasible.

Another series of case studies have looked at distributed software development – in particular when globally distributed teams are agile. The findings concern how control can be exercised over distance when teams depend on each other, the need to coordinate, and employ mechanisms for coordination. The findings also concern communication in globally distributed teams through virtual meetings and the breakdowns and repairs of such communication.

Action research on agile development has been conducted through one large and two small projects. The large action research project was conducted in collaboration with a large pharmaceutical company seeking to employ agile development practices in the software teams yet under the constraint that

other parts of the product development (hardware, mechanics, and clinical) were developed through a plan-driven approach. The findings show that an agile enclave can operate with agility, but is challenged by quality management requirements to documentation. They also show how agile practice can be transferred between teams, while concurrently facing severe barriers.

Two small action research studies have addressed: (1) maintenance in an agile software team in product line development; (2) agility at company-level in web development. Both studies contribute to a deeper understanding of practice, but also with particular methodical ideas and constructs to improve practice being introduced and evaluated.

In the period 2011-15, this research has produced 17 publications of which 4 are A-level journal articles and 2 are A-level conference papers. Three PhD students have graduated from this activity, two of which are through the collaboration with University of Agder.

**Innovation**

Globalization and outsourcing calls for software development in high-cost countries to produce high-value solutions. This observation forms the point of departure for the activity on software innovation. This activity focus on innovation and creativity at the level of the software developing team. The purpose was to develop methods, tools, and techniques as well as infrastructures and conceptual models to support such teams in producing valuable solutions throughout the span of a systems development project. The activity involves one research project on the Essence method, and this research is now documented in a textbook. Another textbook on creativity in software innovation has also been developed as part of this activity.

A particular line of software innovation research has investigated how to support creativity with tools that integrate idea evaluation. This line of research has led to a design framework that has both theoretical backing and has been experimented with. The resulting design theory claims that much creativity comes from idea evaluation and that should be designed into supporting tools.

The DISIMIT project was at an overall level directed at IT management in municipalities, but our part was specifically concerned with value-creation and business cases as a particular way of explicating the effects of digital innova-

tion in Danish municipalities. The action research with three municipalities and a major IT consultancy led to the development and evaluation of a minimalist method for business cases – a method that is now widely used in many municipalities. It also led to deeper understanding of benefits realisation as a process and how it integrates with business cases. A final contribution came through the theoretical development of a values framework for IT in public organisations linked to the values observed through the action research with the municipalities. Our findings show that it is possible under some circumstances to explicate benefits and values of digital innovation and to devise a process in which stakeholders can be involved and can understand the involved benefits, values, and their ramifications. The findings are generalisable and are not specific to municipalities.

The Pay-E-Safe project developed a token-based electronic micropayment system for East Africa with a focus on sanitation services. The solution had to be inexpensive and sensitive to the local context: low income, illiteracy, hygiene issues, unreliable power supplies, unstable communications, and a very different culture and economic structure. These challenges required the team to combine knowledge areas such as hardware design, anthropology, software innovation, business innovation, and interaction design. This project linked between this activity and the interaction design activity of the unit.

The unit has organised and chaired an IFIP WG 8.6 conference on information systems diffusion and value-creation for all. In the period 2011-15, this research has produced 24 publications of which 4 are A-level journal articles and 2 are A-level conference papers. One PhD student is about to graduate from this activity.

## Management

Management of software development as well as management of IT in general is directed at several aspects of research within the broader community of information systems research. Management of software development has been directed at knowledge management, risk management, and software process improvement; and management of IT in general has been directed at the company-level at outsourcing and e-government.

Knowledge management in software projects has been researched by the unit

for more than a decade and knowledge management theory has played a role in action research projects as well as in a few experimental studies. Tool support for knowledge management has been investigated in the KiWi project where a traditional application for project management has been compared and contrasted with a web prototype based on linked data to support the knowledge management activities.

Software process improvement is a research activity studying how software companies improve their development practices. This has led to findings through action research on how to change software organisations via organisational influence processes and on which tactics can have particular impact. It has also led to the realisation that since software companies espouse different organisational cultures tactics must be aligned with the culture in question. A related study has investigated through case studies how Danish software companies outsource some of their development activities to Eastern Europe, what the challenges are, and what they do to overcome these challenges.

E-Government research concerns the management of IT in public organisations, how they organise for IT, and manage large implementation projects. A particular focus in this period has been an investigation of trust in a large case study leading to a better understanding of prerequisites of trust between different stakeholders and how to manage relationships and trust.

In the period 2011-15, this research has produced 21 publications of which 4 are A-level journal articles. One PhD student has graduated from this activity.

## 5.6 Own Evaluation

We evaluate our research in the last five years against the plans we made five years ago and provide an overall evaluation of our research efforts in the period.

### 5.6.1 Quality and impact of publications

By far the most significant part of our own evaluation is the topic of quality and impact of publications. Five years ago we set the goal of improving both quality and impact. Quality of research publications can be measured in terms of how high ranking our outlets are and impact can be measured in

terms of how well cited our publications are. We first analyse the number of publications at A-, B-, and C-levels and compare this with our standing in 2010 (see the IS appendix for information about ranking of journals and conferences). We also analyse our h-index to see both overall production and citation, yet the most interesting part of the h- index is the impact that can be read into the numbers. To further analyse our quality and impact we analyse our co-author network as another perspective on both quality and impact.

**Production and ranking analysis**

Table 5.1 shows the research production in each of the ranking categories and compares this to the previous 5-year period. This shows a drop in numbers, and even calibrated for the total number of staff in the period there is a relative drop in production. There are two observations that are relevant to understand the production. Firstly, the effective number of senior staff has been lower in 2011-2015 and the proportion of PhD students has thus been higher in 2011-2015. PhD students productivity is generally (and expectedly) lower than that of senior staff, and that can in part explain it. Secondly, the publication strategy has been to publish in better outlets at the possible expense of higher production numbers. This can be seen directly in the data where 50 C-level conference papers for 2005-10 is now down to 27 for 2011-15. It is also worth noticing that PhD students often use C conferences as training grounds, and senior staff use workshops in connection with high ranking conferences for focused networking activity. So while the number of C conference papers could be lowered further, some targeted publishing in this category could reasonably still take place - even when pursuing the strategy of increasing quality and impact.

In order to assess if the publication strategy for 2011-15 has been effective, it is necessary to analyse the numbers further. Publications in A conferences is up from 21 (12%) to 40 (29%) for the unit, which is the result of a deliberate attempt to publish in the best outlets within the HCI research community where the highest ranking conference is CHI. The significant growth in A conferences is seen within Interaction Design where 45% of all publications fall in this ranking category. This is significantly higher than the unit average. This is partly at the expense of A journals. Systems Development research

| | **2006-2010** | **2011-2015** |
|---|---|---|
| A journals | 33 (18%) | 20 (14%) |
| B journals | 24 (14%) | 15 (11%) |
| C journals | 0 (0%) | 2 (1%) |
| A conferences | 21 (12%) | 40 (29%) |
| B conferences | 49 (28%) | 34 (25%) |
| C conferences | 50 (28%) | 27 (20%) |
| Sum | 177 | 138 |

Table 5.1: Number of publications ranked by outlet level for journals and conferences

is by large published within the research community of Information Systems where there has long been a move towards publishing more in journals at the expense of conferences. Here, the AIS Basket-of-8 journals signify the highest ranking A journals. Very few Systems Development publications are in A conferences while many more are in A journals (29%) and B journals (26%). This is significantly higher than the unit average for journals. Based on this it seems fair to conclude that the strategy of publishing better has, in part, been fulfilled. More details can be seen in the appendix where all publications are classified, ranked, and counted.

**H-index analysis**

In order to understand the impact of the IS unit's research, we have conducted a simple citation analysis based on h-index of the individual unit members. All senior staff members of the IS unit have a Google Scholar profile page (as part of the department policy) and we used these pages in our h-index analysis. As a part of this, we have calculated the h-index of the unit - that is $n$ published papers (by any member of the unit) that each has been cited at least $n$ times.

First, we see that the IS unit is characterized by having a relatively even distribution in the h-index for the individual members (see Table 5.2). Here we notice that four members of the unit have an h-index above 20 and that the unit covers a spectrum in h-index from 5 to 27. Further, we can also calculate (not in the table) that the total number of citations for the unit is currently

13796 where 7880 of these citations is from the current research evaluation period (since 2010). This implies that (with the same unit members) the unit had a total of 5916 citations five years ago and have had a growth of 133% in total number of citations during the period.

| | Citations | H-index | i10 | Citations since 2010 | H-index since 2010 | i51 |
|---|---|---|---|---|---|---|
| JK | 2909 | 27 | 49 | 1671 | 22 | 13 |
| PAN | 2857 | 18 | 33 | 1337 | 11 | 9 |
| JS | 2289 | 23 | 35 | 1200 | 16 | 12 |
| JR | 2019 | 23 | 37 | 1221 | 17 | 11 |
| MS | 1545 | 22 | 31 | 1027 | 18 | 9 |
| JP | 635 | 14 | 21 | 474 | 13 | 1 |
| IAA | 574 | 12 | 17 | 255 | 8 | 3 |
| JN | 458 | 12 | 15 | 270 | 9 | 3 |
| DR | 219 | 7 | 2 | 142 | 5 | 1 |
| JPe | 178 | 6 | 4 | 172 | 6 | 1 |
| AB | 113 | 5 | 2 | 111 | 5 | 0 |

Table 5.2: Citation analysis including total number of citations, h-index, i10-index, and h-index since 2010 (from Google Scholar). The rows signify individual senior members of the IS Unit. Also, we have included a column called i51, which refers to the number of papers that a given author has that has at least 51 citations (the IS Unit h-index).

Secondly, we calculated the combined h-index for the IS unit. This was done in a number of steps (1) printing, (2) sorting, (3) removing duplicates, and (4) manual count starting with the paper with the highest number of citations and then going downwards. The resulting h-index for the IS unit is 51. We can see from the table that all unit members except one is contributing to this number, by having at least one paper with 51 citations, and that the unit is divided into two groups namely one group (N=6) with 3 or less i51 publications and another group (N=5) with 9 or more i51 publications. This last group comprises those with the highest h-indices.

In relation to impact and citations, it should also be mentioned that members of the unit have received the "ACM Mobile HCI n-10 most influential paper award" for three consecutive years, (2013, 2014, 2015) based on citation
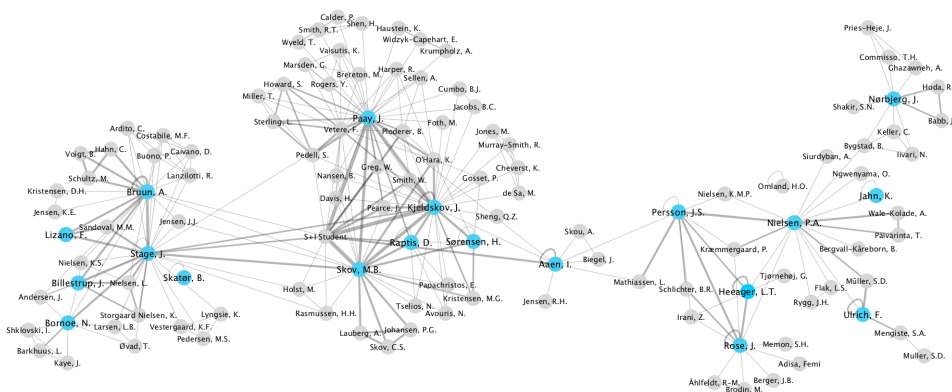
Figure 5.2: Co-author network of IS unit research publications

counts for papers published 10 years earlier. Although this award is based on research from before the current evaluation period, this is indicative of the impact that the unit's research is having.

**Co-author network**

Part of our effort to understand the publication strategies that we have implemented during the previous five years was to conduct a detailed social network analysis on our publications. Social network analysis is useful in visualizing the unit's publication data and provide insights in two areas. First, the relationships between the three research areas that the IS unit is involved in are clearly visible. Secondly, the degree of external collaboration the unit has in relation to publication production is significant. The network is presented in Figure 5.2 and each blue node represents a member of the unit, while a grey node is an external co-author in one of the unit's publications. The thickness of a connection between two nodes illustrates the number of papers by these authors in collaboration. The thicker the line, the higher the number of publications.

The structure of the network clearly depicts the three activities the unit is involved in. On the left side of the network we see a cluster of authors that are publishing on usability engineering, while in the center and on the right side we see clusters of authors that are publishing in interaction design and systems

development, respectively. What is evident from the network is that all three unit activity areas have involved a large number of external collaborators (co-authors of the unit's published research). What is also apparent from the network is that there is some, but not a lot, research collaboration between the three main activity areas of the unit.

It seems from this analysis fair to conclude that all the research activities are well connected externally with (mostly international) co-authors, and internally within the three research themes. The deliberate attempt by the unit to enhance collaboration within the unit on publishing research has only recently commenced as indicated by the relative weak co-author ties between the three research themes.

### 5.6.2 Strategic research project

At the last research evaluation in 2011 we defined, as one of our goals, that we wanted to establish a strategic research project on digital public service to span across the whole unit. The project had the tentative title of "digital municipality 2020". There are two main reasons why we have not done this. Firstly, the unit's research has evolved in other directions than those expected, for example toward smart cities, sustainability, and digital ecologies. Secondly, we have experienced great success in partnering with other units in the department and in the faculty in pursuing larger strategic research grants.

However, with the same aim of *creating a setting for collaboration across the unit and to exploit the individual research competences*, we have instead established the Research Center for Socio+Interactive Design (S+I) spanning the whole Information Systems research unit. This center was established in the summer of 2011 and has since been used to brand the unit's research, funding applications, publications, presentations, etc. The S+I research center has been successful in facilitating new collaborations in the research unit across its historical divide between Systems Development and Human-Computer Interaction. Specifically it has led to joint efforts in research leadership and grant applications, joint efforts on publication and research strategy, and joint publications with new constellations of authors. It has also led to a joint S+I seminar series (fortnightly), 4 scheduled S+I research meetings per semester, weekly S+I catch-ups in the coffee area, and joint teaching on the master's thesis course.

### 5.6.3 Extend international research collaboration

At the last research evaluation it was our goal to extend our collaboration with existing international and national partners, leading to closer research collaboration and joint research funding. This has been achieved. We have not only extended collaboration with existing international and national partners, but have added new international research collaborations as evidenced by the number of co-authors from international institutions in our publications during this period. The co-author network analysis shows 49 different collaborating partner institutions and companies, 35 of them international and 14 national. These partners are detailed in Figure 5.3. We have also increased the bandwidth on existing international collaborations with an increase in the number of joint research studies and projects undertaken, and an increase in sabbaticals, research visits and internships for research students to these institutions. In terms of joint research funding we have partnered with different institutions in the EU to apply for joint funding. While none of these applications have yet been successful, more are still under development, such as an EU application for Innovative Training Networks with, amongst others, Strathclyde (UK), Lund (Sweden), Lancaster (UK), and Aalto (Finland) for a network on Wearable and Mobile Evaluation.

### 5.6.4 Research education

The goal was to maintain a high level of involvement in research education. The unit has kept an ordinary level of involvement by supervising own PhD students and PhD students at University of Agder and by being involved in several PhD courses both at Aalborg University as well as in Norway. The national PhD umbrella programme DaRSIS ended in early 2011 and attempts to finance its continuation have not been successful. This led to considerable reduction in national PhD courses, workshops for supervisors, and summer schools. The collaboration has continued, but at a much lower level and based on informing PhD students and supervisors at the other universities of activities and courses. Other national supervisor workshops not specific to information systems research has been attended in small numbers.

There has, however, been a significant contribution from the unit to the IRIS seminar that now functions more as a doctoral consortium and a reasonable
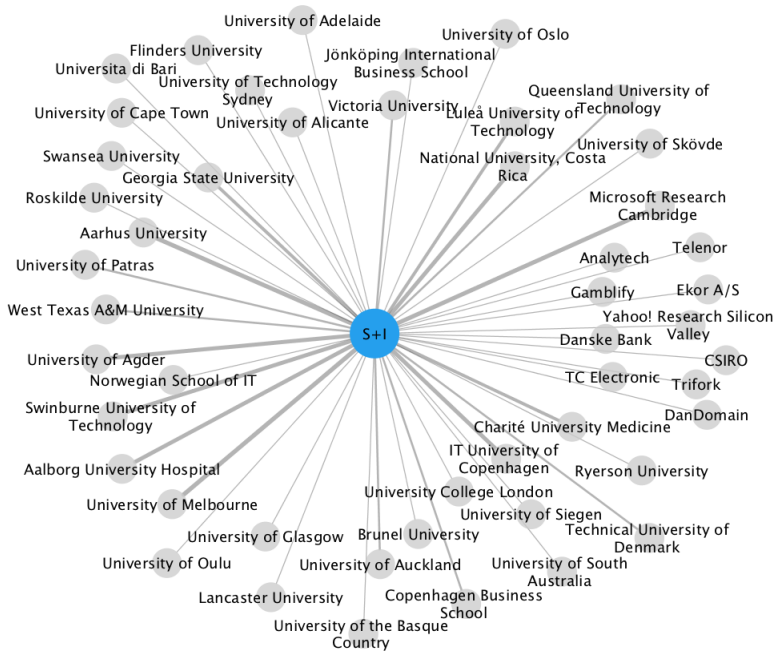
Figure 5.3: Co-author network analysis showing collaborating institutions/companies.

training ground for PhD students.

### 5.6.5 Recruit more staff

At the last research evaluation it was our goal to recruit more scientific staff, specifically at the lower levels of the career stages, including more PhD students. Although the foundation for increased recruitment, in terms of significantly expanded teaching tasks, has been achieved, we have not been able to achieve the goal of recruiting more staff. In fact the research unit has decreased in size rather than grown. We have only just been able to maintain the number of tenured staff, but we have not been able to maintain the number of assistant professors, postdocs and PhD students. The reason for this is that the IS unit was hit particularly hard during the faculty's economic crisis in 2013-14. During this period we were unable to recruit replacements for staff whose positions ran out, and for staff who left. In this sudden downturn this

meant that the unit lost two assistant professors and one associate professor. In the same period the unit has had too few external research funding to compensate for the diminishing internal funding. In addition to this, we have not been able to offer the expected number of PhD scholarships, which has resulted in a reduced rather than increased number of PhD students. On the positive side, as the unit has significantly expanded the volume of its teaching tasks, most notably from the establishment of the new BSc and MSc in Interaction Design, a solid foundation now exists on which future recruitment of staff can be done.

### 5.6.6 Research funding

Two externally funded projects, DISIMIT and WPU, started in 2010 or before were active to 2012 and 2015 respectively. In this research evaluation period, the unit has been less successful in attracting new external research funding compared to the previous period. The unit has submitted a much higher number of applications to the Danish Council for Independent Research (Technology and Production Sciences as well as Natural Sciences) and to the Strategic Research Council than in the previous period, for research projects as well as individual postdoc projects; but these applications have not been successful due to the very low rate of acceptance (as low as 3-12% for some calls). In the previous period (2006-10), the majority of projects funded were medium-sized with a few partners, which the unit had some success in attracting. In the current period (2011-15), the funding scheme for Danish research councils changed towards larger projects carried out by research consortia. The unit has been successful in being partner with two such large consortia, in the TotalFlex project, and the DiCyPS research centre, attracting 2.4 and 4.5 Million DKK to the unit respectively.

In addition to the large research consortia, a new type of research funding has become more common in this period, namely funding for small projects, with short durations, from various sources. Several of our research activities have been based on this type of funding.

## 5.7 Plans (2016-2020)

**Focused publication strategy:**  We are convinced that our publication strategy will pay off both short term and long term. Hence, we will continue to pursue A conferences within the HCI research community and A journals within the IS research community with vigilance. Spin-off workshops should still play a role in particular when in connection to A and B conferences; but these are primarily seen as training grounds for PhD students. It is also a goal to see an increase in joint publications across the unit as synergies are yet to be fully developed.

**Research funding:**  We shall pursue a two-legged strategy with one leg devoted to playing an active part in creating large research consortia and make our research attractive in these consortia. These large consortia are more and more likely to address societal concerns and grand challenges in which our contribution both will be investigating and theorizing on the context and the influence of the challenges and in interaction design of prototypes and their evaluation to address these challenges. Funding sources for these consortia are likely to be the Innovation Foundation and Horizon 2020.

The other leg will pursue smaller funding sources where the research can often be more focused and where the contribution is more direct, solution-oriented and in collaboration with companies. Funding sources for these smaller projects are likely to be the industrial PhD grants and the innovation platform InfinIT.

**Recruitment:**  Our fundamental need for more research staff cannot be separated from our teaching obligations where the unit is responsible for a new full 3+2 years education in interaction design and an international 2 years master in application design as well as shared responsibility for the full educations in computer science, software engineering, and informatics. The economic revenue from this effort will have to be turned into positions in the IS unit in a near future.

Fueled by both teaching revenue and the above strategy for research funding the unit will proactively seek to attract the right competences to its open positions by searching actively and internationally. As a supplement to this,

the unit will seek to hire research assistants later to be groomed into a research career through PhD studies.

Increase in PhD scholarships is necessary to feed the chain from below. An increase in number of PhD students depends on the success of the above funding strategy and to alleviate some of the risks associate with this we will explore and exploit the industrial PhD scholarships in collaboration with companies.

**Industry collaboration:**  Close collaboration with software companies has long been the primary mode of operation for our research on systems development and usability engineering, but we are seen increasing interests in interaction design from companies—not only software companies. Industry collaboration is not only an interest it is also an opportunity both to confront our research with what is relevant and useful to these companies and help us in providing solution-oriented research, but it is also a lever to many types of research funding. In many respects we see industry collaboration go hand in hand with funding and recruitment.

**Maintain international and national research collaborations:**  It is our plan to maintain collaboration with existing international and national research institutions and strengthen these relationships with increased activity in joint research projects, resulting in co-authored publications. We will pursue joint funding opportunities with existing international partners, especially those in the EU, to increase our success at achieving collaborative research funding. We will also build on existing partnerships with local Danish companies, and negotiate with additional companies, to get their commitment of resources toward joint research projects. We will maintain our connections with International collaborators through sabbaticals, research visits, collaborative writing projects, online communication and internships for research students.

**Research themes for 2016-2020:**  A few research themes seem very relevant for the coming period though some are very likely to change as the research areas evolve:

- *From usability to user experience:* Emotion is key in assessing UX of interactive systems, yet measuring emotions is challenging for several reasons. Within the next period we aim to study how UX can be measured on the basis of psychophysiological measurements of emotional reactions. This enables real-time continuous measures of UX.

- *Interaction design for data-intensive cyber-physical systems:* As part of our involvement with the DiCyPS research centre, we will be doing research into interaction design for digital ecosystems, where several systems, computerized as well as physical, are interconnected and used in concert. This will extend our previous work on ubiquitous and mobile computing. It will also intersect well with the interaction design activities on interaction techniques and behavioural change, as well as planned S+I activities within the broader theme of "Smart Cities".

- *Value-driven software development:* The most interesting values created by the design, development and use of new software-based systems are often intangible or wicked compared to the more traditional requirements and features. These values can be researched through business cases and business models at the intersection between the software company and its customers and users. They can also be investigated and innovated in more complex settings such as triple and quadruple helix between software companies, public procurers, knowledge providers, and citizens and users.

- *Smart city and smart society:* Driven in part by the EU interest in furthering smart city ideas and in Danish cities interests in becoming smarter cities and part of a smarter society we see the unit playing a central role in this movement. Our research competences lie in between the smart city technologies and the necessary understanding of users, interactions, benefits, and social processes. This has a clear intersection with both the research in DiCyPS research centre and the CLINES project on smart city collaboration with companies. It is also a research interest that has the potential to tie the three research area together in S+I and utilize their synergies.

## 5.8 Committee Evaluation

**Observations**

The information system unit has a unique composition of information systems and human computer interaction researchers with both a well-established tradition and an insight into the more recent development of the fields. One of the unique strengths of the unit is the action research tradition, well established in IS, but not so much used in HCI, even though very relevant even for the developments within HCI. In times when we are increasingly speaking about scientific and societal impact, a research methodology that clearly honours the change ambitions is becoming increasingly relevant for the society and both interaction design and information systems works with societal change, but in different ways and on different levels.

Since the last evaluation the group has worked towards increasing the quality of the scientific impact by publishing in more reputable venues and increasing the share of A-level Journals and conferences in a remarkable and impressive way. However, the same strategy has not been applied to the production of funding applications resulting in fewer and smaller grants being awarded. The group could clearly benefit by prioritizing which calls to submit to, by establishing an internal quality check and review process and by more carefully researching the funding organization's requirements and traditions.

Among the weaknesses we see a too low degree of PhD students and problems in landing external grant applications. In addition we sense a staff tension in the group where there could be clear benefits by addressing and resolving some of these issues separately. Currently the full potential of the staff cannot be exploited to the extent possible which causes disturbances in the organization.

One of the greater impacts that you can have is through the education and through the production of skills that are relevant for companies, municipalities or start-up entrepreneurial organisations. The IS unit is very much engaged in the Informatics education but also in the Interaction Design bachelor and subsequent Master education there are great capabilities to contribute well to the work force of the future. There is a great potential in using the education more strategically as an asset in the research. The problem-based education tradition of Aalborg university also fits very well with this opportunity.

Collaboration is key to future development, both international collaboration, where the unit has strong traditions already as well as collaborations within the university. This has not fully been exploited. Collaborations with the neighbouring units known for their high scientific esteem may benefit by including also an IS perspective in their projects to increase their knowledge about and collaboration on the user perspective of their research. More closely collaborating with the neighbouring units may eliminate the possible threat that these research units may cannibalize on the required and necessary growth within IS.

Leadership in the unit is of huge importance and with the right strategic leadership, acknowledged and endorsed by the rest of the unit and by the department, wonders can be achieved. This is important for the needs of development of common visions and missions, for the development of the publication and application strategy and for the important recruitment needs. With careful individual leadership and mentoring there are also several opportunities of lifting younger faculty, postdocs and PhD students above their current levels. Leadership, mentoring, strategy and organization are important keywords in this development.

Finally the important ongoing digitalization that is quickly changing the surrounding society and the conditions for working life and leisure provide vast opportunities for development. The digitalization puts big demands on IS skills and research from the society and this should be further explored and supported.

### Suggestions and recommendations

**Renaming the unit:** For historical reasons the group is named and branded as Information Systems. Information systems is a research field of its own, with journals, conferences and groups widespread all over the world. Human Computer Interaction is also nowadays an internationally established field, but not necessarily included in Information Systems. Therefore, our recommendation is to consider renaming the unit to better reflect the group composition ("Information Systems and Interaction Design" or "Information Systems and Human Computer Interaction")

**Vision and mission:** The Information Systems unit can easily be perceived

as a collection of individual researchers or research groups without a joint mutual direction. We therefore suggest that the unit engage in an activity to develop a joint vision and mission to better align the activities within the unit and possible to bring the different research interests closer together and maybe also produce new alliances and synergies.

**Impact:** The latest important quality to address in research and education can be labelled under impact. Scientific impact, the quality of the research, publication venues, conferences, citations etc. have been the focus of the IS unit lately. However, recently the issue of societal impact have come more into play as an important quality, not only addressing the importance of previously conducted research but also working with methods to strategically develop the societal impact of the research, by developing the communication and engagement with society, politicians, decision and policy makers, standardization organizations, industry and the start-up sector. By carefully planning for and developing future pathways to societal and scientific impact, the quality of the research and the likelihood of successful research applications increases.

**Improving funding applications:** Working with the understanding of the research funding organisations, targeting the right funding agencies for the appropriate applications, working with internal review and quality assurance, using scientific advisory boards for advice and review, and broadening participation from all staff are all mechanisms to increase the likelihood of applications being funded.

**Strategic recruitment:** Recruitment is one of the most important strategic considerations to be made. By working actively with search committees, headhunting, friendraising, inviting potential candidates to give seminars, posting your call for open positions in strategic forums, etc., you increase the likelihood of being able to recruit candidates with higher excellence. Try to avoid inbreeding and recruiting your own well known candidates if possible.

**PhD students:** Increase the number of PhD students to at least be as many as the existing faculty. Of course this relates to funding but there are

other mechanisms to increase the funding than to only acquire external grants. Use strategic partnerships to recruit industrial PhDs, be flexible in the contract procedures to allow shorter contracts, and investigate further international collaborations to allow for possible PhD students that are employed by other universities (without PhD granting rights) to collaborate with you and earn their degrees at Aalborg university.

**Multidisciplinary collaboration:** The digitalization increases the demand for broader multidisciplinary collaboration stretching over scientific areas and departments. There are many opportunities for collaborations within the university at other departments that are not sufficiently exploited and that can provide huge benefits to the department by broadening the abilities to acquire larger external projects addressing more complex problems requiring multidisciplinary skill combinations of yet unseen proportions. One way to start this is through collaboration within the education programs, but by seeing this as a first step to broaden the collaborations. Joint PhD students are other possible mechanisms (i.e. Media Technology, health sector, smart cities, etc.)

**Reworking strategy:** As said before, society is rapidly changing and the need for the skill sets that IS educates and promotes are increasingly needed to solve current problems in the society. By reflecting these problems and showing a greater interest in the societal issues and digitalization, one might prove to be more societally relevant and engageable by the society. Areas to be explored could possibly be: digitalization of the society, sustainable development goals, the changing landscape for higher education, etc.

**Increased ambition and self-esteem:** The IS unit has a greater potential and relevance than currently understood from within the unit and the department. Ramp up the ambitions when it comes to joint efforts (e.g. value-based development to bridge HCI and IS, Agile User-Centered Design, action research projects for societal change, interaction design for new start-ups, sustainability in a broader perspective)

**The strategic importance of education for research development:**
With the rapidly changing needs of higher education to digitalize and change focus, there are opportunities to use the education programs

more strategically for development of the research. This goes for Interaction Design, Informatics, and ICT both on BSc and subsequently on an MSc level.
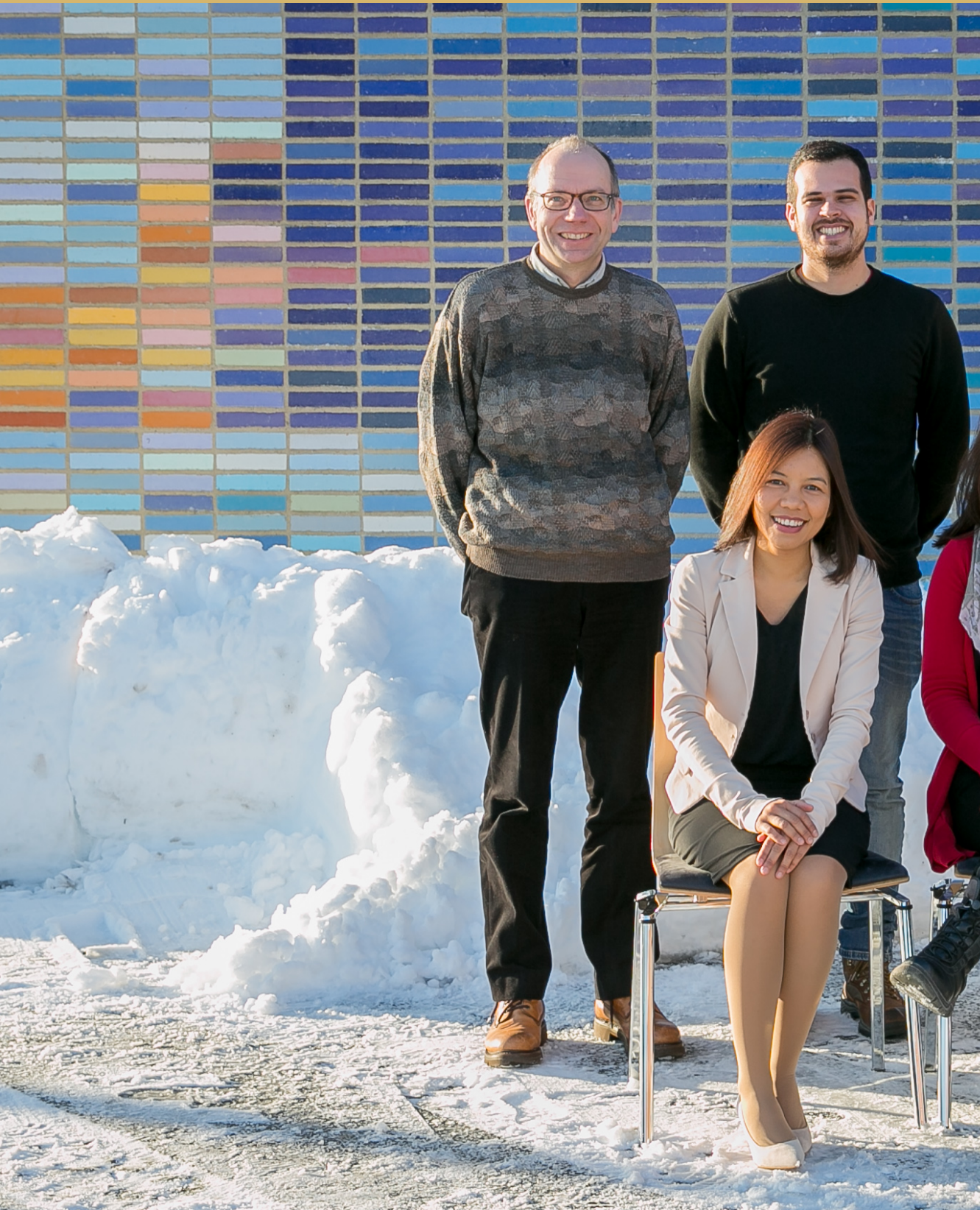
# Machine Intelligence

# 6 Machine Intelligence

## 6.1 Executive Summary

The Machine Intelligence unit is internationally well recognized for its research on probabilistic models, machine learning, and intelligent web and information systems. In spite of challenging conditions defined by a low number of staff and significantly increasing teaching and adminstration work loads, the unit has maintained during the 2011-2015 period a high level of research activities and engagement in the international research community. Visible indicators of these activities are a total of 124 journal and conference articles, leading roles in the organization of 7 international conferences and on editorial boards of major scientific journals, and the graduation of seven PhD students.

Over the report period the unit has been a partner in 3 European research projects, and has significantly increased its collaboration with other research units at the department.

## 6.2 Profile

The Machine Intelligence unit has a long established track record as one of the international leaders in research on probabilistic graphical models and decision support systems. A textbook on Bayesian networks authored by members of the unit has been the standard reference on this topic for a long time, and (over several editions) has received a total of over 7500 citations. This original profile has been broadened in several directions. During the previous reporting period (2006-2010), and for an initial part of the current period, the research area of autonomous agents has been pursued by the unit. However, due to personnel changes, this area became inactive. At

the beginning of the period, the intelligent web and information systems sub-group, formerly associated with the Information Systems research unit, joined the MI unit. As a result, the area of intelligent information systems has been added as a strong component to the profile of the MI unit. Going beyond the traditional focus on graphical models, the research in machine learning has been diversified, but retains its focus on probabilistic and statistical methods.

For most of the duration of the 2011 − 2015 period, and as the current status at the end of the period, the main research areas of the MI unit can be summarized as follows:

**Machine Learning Methods** We develop new types of probabilistic models and learning techniques for complex types of data, such as time-varying or highly structured data, or data that is modeled using unobservable (latent) variables.

**Machine Learning Applications** Probabilistic models and machine learning techniques are applied to a diverse range of application problems, often in the context of inter-disciplinary research projects.

**Intelligent Web and Information Systems** Methods, tools, and techniques for personalized and adaptive provision of information and services, especially in a web-based context, and leveraging social network data.

## 6.3   Staff

The current staff of the unit consists of:

**Associate Professors**: Peter Dolog, Manfred Jaeger, Thomas D. Nielsen

**Assistant Professors**: Nattiya Kanhabua

**Postdocs**: Hanen Borchani, Ana M. Martinez

**PhD students**: Felipe Soares da Costa

Also associated with the unit is Adjunct Professor Anders L. Madsen, CEO of HUGIN EXPERT A/S.

Members at the beginning of the period who left within the period are: Professor Finn V. Jensen (retired 01/14), (temporary) Associate Professor Yifeng Zeng (moved to a Readership at Teesside University, U.K.), Assistant Professor Paolo Viappiani (moved to CNRS position at Université Pierre et Marie

Curie, Paris, France), Postdoc Guandong Xu (moved to a research fellowship at Victoria University, Australia), and PhD students Yingke Chen, Hua Mao, Rong Pan, Fred Durao, Ricardo Gomes Lage, and Jorge Pablo Cordero.

Furthermore employed during part of the period were Associate Professor Bo Thiesson (01/13-08/15; moved to CTO position in industry), and PhD students Martin Leginus, Jiuchuan Jiang.

## 6.4 Goals (2011-2015)

Regarding the general, strategic plans for the unit, we wrote in the 2006-2010 evaluation report:

*The overall plan for the next years is to continue the successful broadening of the unit's research activities, and to achieve additional research capacity and impact through externally funded international and national research projects. A key component in this plan is a fruitful integration of the Intelligent Web and Information Systems sub-group with the Machine Intelligence unit. Areas of common interest, notably recommender systems, will form natural focus areas of research.*

Regarding staff development, the stated goals were:

*The MI unit currently has 4 permanent staff members, ... and one member in a non-permanent position. In 2011 a new assistant professor (Paolo Viappiani) will join the unit. His research (recommender systems) will strengthen the link between the traditional research areas in the MI unit, and the new intelligent web and information systems area. ... the unit will consist of 6 active researchers as of 2011. This is at the lower level needed to maintain and consolidate the research activities in the current four main areas, and the unit will therefore aim for a moderate growth to about 7 or 8 active researchers at the assistant, associate and full professor level.*

## 6.5 Activities and Results

In the period 2011-2015 the research in the MI unit has been focused on machine learning and intelligent web and information systems, where the unit builds on long established core competencies. Due to the departure

within this period of Associate Professor Yifeng Zeng and Professor Finn V. Jensen, formerly active research areas in autonomous agents and decision models moved out of focus. New colleagues who joined the unit for part of the report period (Assistant Professor Paolo Viappiani, Associate Professor Bo Thiesson) contributed further expertise and research activities in the wider area of machine learning.

### 6.5.1 Machine Learning

**Relational Learning**  Traditional statistical and machine learning techniques are made for data given in tabular form, where for a number of independent cases the observed values of certain attributes are recorded. This data model is not adequate for describing inter-connected entities whose properties show mutual dependencies, such as people connected in a social network, web-pages connected by links, or genes connected by co-regulation relationships. The increasing prevalence of such relational network data has given rise to a new area of machine learning, often referred to as (statistical) relational learning. Our work on relational learning pursues two main objectives: the development and implementation of general probabilistic modeling frameworks and learning techniques for relational data, and a better understanding of their theoretical statistical and complexity-theoretic properties.

Our main focus with regard to methodological developments has been the Type Extension Tree framework for the specification of complex relational features. In relational structures, the properties of one entity can depend on properties of other entities that it is directly or indirectly connected with by a chain of relations. In order to construct accurate probabilistic models for relational data, one has to be able to specify such dependencies in a formal representation language. Type Extension Trees are a logic-based representation language for complex features capturing the combinatorial structure of the relational neighborhood of an entity. We have developed a learning method for Type Extension Trees that automatically discovers combinatorial features that are informative for the prediction of a given target property of entities.

On the theoretical side, we have been concerned with the analysis of the computational complexity of inference for probabilistic models of relational data, such as computing the posterior probability of a target attribute value of

some entity, given evidence on the properties of other entities in the relational network. Extending our previous work on this subject, we have established that for the currently most popular types of probabilistic modeling languages, this inference problem does not scale polynomially in the size of the relational domain.

Our research on relational learning is conducted in close collaboration with the universities of Florence, Trento, and Leuven. Major outcomes of this line of research have been published in two of the highest ranking machine learning and artificial intelligence journals.

**Learning Probabilistic Automata**   Probabilistic automata models play an important role in the formal design and analysis of hard- and software systems. In this area of applications, one is often interested in formal model-checking procedures for verifying critical system properties. Traditionally, models used in model-checking are manually constructed, either in the development phase as system designs, or for existing hard- or software systems from known specifications and documentation. This can be both time-consuming and error-prone. On the other hand, for such already implemented systems, empirical observations of system behavior can be used as training data to infer automata models using machine learning approaches. We have adapted and extended existing automata learning techniques to address the specific challenges that arise in the context of formal system analysis. In particular, we have extended previous learning methods for purely probabilistic automata classes to richer classes including nondeterministic inputs, and time-delayed transitions. A focus of our work has been to analyze theoretically and empiricallly whether the learned models provide a good approximation of the true data-generating models with respect to model properties that can be expressed in linear time temporal logic.

This work has been conducted in an internal collaboration with the DES unit, and has resulted in two PhD theses.

**Probabilistic latent variable models**   A standard probabilistic inference task is classification. That is, to predict the (discrete) class label of an object based on attribute values describing that object. Typical application examples include spam filtering, credit scoring, and diagnostics. In previous projects

we have developed methods for learning classification-oriented probabilistic graphical models that rely on unobserved/latent variables for capturing probabilistic dependencies inherent in the domain being modeled. Not only did these methods demonstrate very good classification results, the latent structures included in the models also provided insight into the underlying data generating processes and thereby added to the interpretability of the models.

More recently, we have extended these models and learning algorithms to dynamic domains that evolve over time. The models have been evaluated in a reliability engineering setting (specifically activity recognition in offshore well-drilling) as well as for credit scoring of bank customers. For the latter scenario, particular emphasis have been put on the detection and adaptation to drift in the underlying distribution.

The use of latent variables to represent unobserved process dynamics have also been explored in a recommendation/collaborative filtering setting. In this context, all users and items in the domain are represented simultaneously in the model, where latent variables are used to mediate the dependencies between these entities. The latent variables also provide semantic insight into the underlying profiles of the users and items. For learning these general types of models, we have employed a scalable Bayesian learning framework, where the model parameters are treated on equal footing with the latent variables in the model. We have shown that the complexity of the learning algorithm grows linearly in the number of observations and that it supports an efficient use of the MapReduce framework.

The majority of this work has been conducted in close collaboration with the Norwegian University of Science and Technology.

**Learning and inference in complex domains**    Domains containing mixtures of discrete and continuous distributions, or simply complex distributions that fall outside standard manageable distribution families, are inherently difficult to model and reason with. Often one needs to resort to approximate inference algorithm, such as sampling-based methods, when dealing with these types of models. Alternatively, modeling frameworks based on either mixtures of polynomials or exponential functions have been proposed as flexible alternatives for both representation and inference. The advantages offered by these frameworks are that they can, in principle, represent any distribution

arbitrarily well while at the same time provide support for standard operations required for performing exact inference.

With the aim of trying to unify the developments of these frameworks, we have proposed a general framework, called mixtures of truncated basis functions (MoTBFs), which generalizes the previously proposed frameworks. Based on the MoTBF framework, we have developed algorithms for learning MoTBF-based models from data, and we have explored inference methods tailored to these types of models. In particular, we have consider coupling the learning phase with the inference phase in order to guide the learning algorithms towards models which support efficient inference.

The developments in this area has been achieved together with the Norwegian University of Science and Technology as well as the University of Almería. Some of the developments have also been realized as part of the project *Data mining with PGMs: New algorithms and applications* (TIN2010-2090 0-C04-02,03) supported by the Spanish Ministry of Economy and Competitiveness.

**Analysis of streaming data**  With the rapid growth of information technology, unbounded flows of observations are generated daily in a wide range of domains, such as network monitoring, social networks, fraud and intrusion detection, telecommunications, etc. These flows are known as data streams. Contrary to finite stationary data sets, data streams pose many challenges to computing systems, due to limited memory resources, processing time, and the possibility of concept drift (i.e., the underlying data-generating distribution changes over time).

As part of the EU FP7 AMIDST project, we have explored and developed scalable algorithms for learning probabilistic graphical models within a data streaming context. In particular, we have considered scalable Bayesian learning algorithms focusing on data streams exhibiting concept drift. This not only included gradual and abrupt changes in the generating distrubution, but also changes in the relevancy of the predictor variables. We have used latent variables for explicitly capturing and representing global types of concept drift. For demonstration, the models have been evaluated based on bank data supplied by Banco de Crédito Social Cooperativo S.A. Subsequent analysis have indicated that the latent concept drift variables included in the model reflect the general economic climate (unemployment rates in particular) in

Spain before, during, and after the start of the economic crisis.

The main developments have been realized as part of the EU FP7 AMIDST project. The models and algorithms have been implemented as part of the open source AMIDST toolbox (`http://amidst.github.io/toolbox/`), which is available under the Apache Software License version 2.0.



## Spotlight: Social Network Analysis
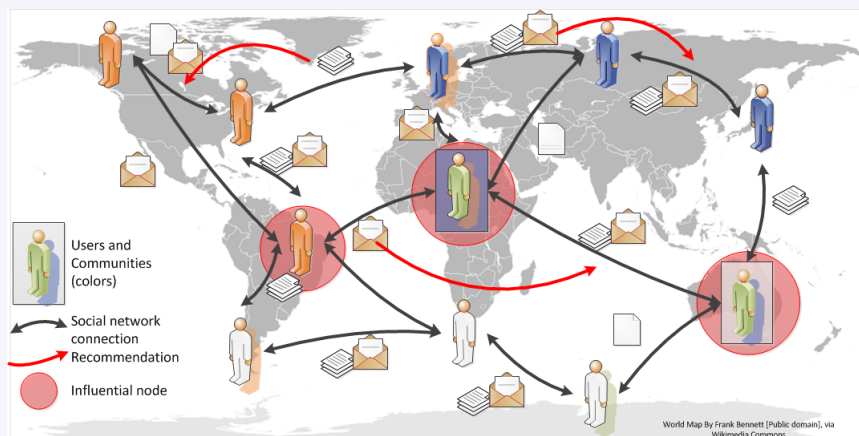
Social networks are an important application area for many of the techniques developed in the MI unit: our *relational learning* techniques can be used to learn probabilistic models for the community structure in the network. *Latent variables* model hidden properties of the network members, such as preference profiles that can be exploited for recommder systems. Observed user activities, such as *tagging*, *bookmarking*, or *message forwarding* can be exploited to design more powerful tools for personalized recommendation and information access.

**Preference Elicitation** Dealing and reasoning with preferences is an important topic for research on autonomous agents, recommender systems, and decision-support tools. Our interest is to design efficient methods for eliciting or learning preference information from preference statements, critiques to examples, observations of user's clicking behaviour, etc. On the foundation of utility theory and Bayesian belief updating, we have developed an interactive methods for the identification of a query that will be most informative for a user's unknown utility function. Related techniques have also been exploited to develop a method for information aggregation and information elicitation in crowdsourcing scenarios, which we model as the problem of finding an unknown Boolean concept from unreliable statements about the concept from different sources.

This research has been conducted in collaboration with the universities of Regina and Toronto, Canada.

## 6.5.2 Intelligent Web and Information Systems

We have concentrated on research of methods and algorithms which utilize information from so called web social media or web social network in this period. The algorithms help to generate *recommendations* or in *information access* for users. The social networks we consider can be grouped into three types: traditional web social networks such as facebook or twtitter, social annotation systems, and wikis. The algorithms and systems which we have developed are also empirically validated.

**Recommender Systems** In this area we have focused on two main approaches in research: *latent based models* and nearest neighbour based multifactor approaches.

We have developed a hybrid nearest neighbour based tag based recommender system based on both, individual as well as group based preferences taking into account content of the items, tagging activities of the users, tags, and social network of users and collaborators. We have applied this method in various contexts such as eppidemic surveilance of social media, collaborative problem solving on wikis in software engineering domain, in publishing and sharing messages on social media, and in collaborative creativity sessions

online. We have also experimentaly validated the approach in those contexts.

We have also developed various variations of topic models based on latent dirichlet analysis for recommendation in various contexts. The model was extended with clustering, KNN, and other methods to provide a hybrid which best suits the analysed task. Among the tasks which we have experimented with are advertisment recommendation, query recommendation in web search, and items recommendations in tagging systems. We have experimentally validated the approach in those contexts.

One of the components which is neccessary in any content based recommendation task is similarity calculation. We have improved a similarity calculation for short messages which can be found for example on twitter. Most of the traditional similarity measures require larger text corporas. We have shown, that with our similarity calculation on short messages we can effectively compute recommendations while our calculation also outperforms traditional similarity calculation approaches. For this work we have received a best student paper award at WebIST 2013 conference.

We have further improved a tensor factorization framework for movie recommendation with clustering.

We have also in collaboration with the University of Marie Curie in Paris developed a contextual bandit based approach for publication of message on social network optimizing the spread of the message based on group user preferences.

These activities led together to 3 PhD thesis and one MSc thesis. Further, some papers already attracted a good number of citations. For example a paper on personalized tag based recommenders from 2012 already received 43 citations. Further to this the activities resulted in many publications where several have been published at the top venues such as AAAI, WISE, WWW or WWW Journal.

**Personalized Information Access**  In this area we have focused mainly on *tag cloud generation* as a navigation and exploration user interface for information access and various computational methods to support that. In the recent period we have also developed a novel map based explorative user interface *BeoMaps* for querying trending topics in social media.

In tag cloud generation we have employed various random walk on graph

algorithms based on a graph constructed from co-occurances of tags and words. We have learned preferences for personalizing the tag cloud from queries and also from user activities on social media. We have also shown that by incorporating entities learned from the social media messages we also improve the tag cloud generation in comparison to word and tag based approaches. For this work we have received a best student paper award at WebIST 2015 conference.

The BeoMap is a user interface which provides two dimensions formed by two chosen parts of a user query. The underlying social network messages are then displayed in the two dimensional space according to a relevance metric chosen, and closer to the dimension they match more closely. The BeomMaps interface was developed in collaboration with University of Illinois at Urbana Champaign.

This activity produced so far one PhD thesis and produced papers at good conferences and journal such as HT, ICWE or JASSIST journal.

## 6.6  Own Evaluation

The unit has worked under challenging conditions in the period 2011-2015. Financial difficulties at the university severely limited the possibilities to hire new staff, and therefore the growth to 7 or 8 active researchers envisioned at the beginnning of the period proved to be unrealistic. For most of the period the unit consisted of only 4 active researchers at the assistant, associate and full professor level, and at the end of the period there are 3 permanent staff at the associate professor level, and 1 assistant professor, which is less than the staffing level at the beginning of the period. This low number of staff on regular academic positions was only partly compensated by the employment of two project-funded postdoctoral researchers. As a consequence both of the low staffing level, and of generally increasing teaching and administration loads, the work load of non research-productive tasks has increased significantly for the members of the unit.

Nevertheless, the unit has succeeded in maintaining a high international standard in its research activities.

*Publications:* Members of the MI unit have published a total of 124 journal and conference or workshop papers in the reporting period. Publication

venues are chosen based on considerations of impact, and the need of PhD students to obtain within a short time frame a significant publication record. Impact can be achieved through publication venues that have a generally high reputation and visibility, or venues at which a more specialized audience can be directly addressed. It is difficult to strictly categorize publication outlets according to their reputation or importance. In lieu of a better alternative, we examine our publication record using the Danish 'Autoritetsliste', which is a comprehensive list of scientific journals and conference series, both of which are classified into top level (level 2) and normal level (level 1) venues:

|             |  Level |    |    |        |
| ----------- | ------ | -- | -- | ------ |
|             | 2      | 1  | ∅  | Total  |
| Journals    | 16     | 16 | 3  | 35     |
| Conferences | 12     | 50 | 27 | 89     |

Publications at venues not contained in the Autoritetsliste are listed in the ∅ column. This can include venues which are not on the list because they are minor events (workshops), that are newly established conferences which have not been added to the list yet, or venues that simply have been overlooked during the compilation of the list. This breakdown shows for our journal publications a very satisfactory proportion of level 2 entries. In the conference category, the percentage of level 2 publications is significantly lower. The two main reasons for this are: level 1 conferences are extensively used for publications by PhD students, who need to obtain dissemination experience and feedback on their work right from the beginning of their study program, and who need to fulfill publication targets set up in their study plans. Secondly, level 1 conferences can be conferences that also senior members of the unit participate in due to their importance for maintaining (European) scientific collaboration networks. Overall, the publication record of the unit is very strong, both in terms of quantity, and in terms of quality.

*PhD student and postdoc training:* Within the period, 7 PhD students have successfully defended their theses. Out of these, three have continued to pursue an academic career, and within short periods after their graduation obtained (permanent) academic positions. Of the remaining four, three have found employments in industry in line with their academic qualifications, and one has started an own business based on innovative technology. Former

unit members at the postdoc, assistant professor, or temporary associate professor level also all have continued successful academic careers, and are now on permanent academic positions in Australia, France, and the United Kingdom.

*Collaboration:* The MI unit has maintained and further developed its strong connections and visibility in the international research community. Members of the unit have organized 3 international conferences, and are active as editors of major international journals. The unit has an extensive international collaboration network, and most of the key publications of the unit are the result of this international collaboration.

At the end of the previous evaluation period a collaboration between the MI and DES unit on learning probabilistic automata was initiated. Increased collaboration with other units at the department was both mentioned in the previous research evaluation report among the plans of the unit, and included among the recommendations by the evaluation panel. Internal collaborations create scientific synergies, and the MI unit can indirectly benefit from external grants obtained by other units. On the other hand, the machine learning expertise provided by the MI unit becomes increasingly relevant for many projects in quite different fields of research. During the current reporting period, the collaboration with the DES unit has been fully developed, and as a result two PhD students from the MI unit graduated within this collaboration. Further internal collaborations have been initiated, and currently 4 PhD students are co-supervised by members from the MI unit, and members from the DPT and DES units.

*Funding:* A lack of success in procuring external funding has been a major concern in previous reporting periods. Within the 2011-2015 period the unit has continued its efforts to obtain external funding, with a total of approximately 15 applications submitted to Danish and European funding agencies. One successful European application, together with two EU projects already running at the beginning of the period, meant that within 2011-2015 the unit has received funding from 3 different EU projects, which together provided a major contribution to supporting PhD students and postdocs in the unit. Still, the level of external funding is not quite satisfactory, and the unit continues to face challenges in procuring sufficient funding to support an adequate number of PhD students and postdocs.

*Staff development:* The unit has missed its ambitious targets for the growth in staff. A main reason for this has been the generally very difficult financial situation for the department, which allowed only very limited new hiring, especially in the second half of the reporting period. An increasing load of teaching and administrative duties, and limited access to resources such as funding for PhD students, have been factors that contributed to the decision of two members of the unit to accept alternative job offers in academia, respectively industry.

*Research areas:* Due to the dynamically changing staff situation, the main areas of research developed in a slightly different manner than originally anticipated. Activities in the area of Intelligent (Multi-) Agent Systems ran out, whereas research in machine learning methods and applications came more into focus. The integration of the original (pre-2011) Machine Intelligence unit with the Intelligent Web and Information Systems sub-group has been successful – especially on the educational side, where new courses that build on shared interests and complementary expertise have been designed.

## 6.7  Plans (2016-2020)

### 6.7.1 General Plans

The main objective for the next 5 year period will be to overcome the currently critically low staff level, and to re-establish a unit size at which the MI unit can fulfill its teaching and administrative obligations, while also pursuing new research initiatives. The short term goal is to find a replacement for the associate professor who recently resigned from his position. In the longer term, the target is a moderate growth by an additional 1-3 assistant and associate professors.

Future appointments should consolidate and strengthen the research areas in which the unit already has a strong profile. In particular, we aim to increase our activities in core areas of machine learning, and in intelligent web and information systems.

Apart from growth through regular appointments, we will try to enable additional research activities through continuing efforts to obtain external funding, and through collaboration in externally funded research projects of other

research units at the department.

## 6.7.2 Research Plans

### Machine Learning

**Relational and Network Data** We will continue our methodological research on learning from relational data. Currently recognized as an important research topic is the development of models and learning techniques for relational data containing both categorical and numeric attributes and relations. Extending our previous lines of research on relational learning, we will develop general, expressive modeling frameworks for such hybrid relational data. The modeling frameworks then are supported by generic learning techniques and software implementations, which can be applied to a wide variety of concrete tasks. An important application domain that we plan to pursue within a European research project is modeling, learning, and decision support for sustainable management of water and land resources.

**Machine Learning and Formal Methods** Our recent interdisciplinary work on automata learning and model checking has opened a new line of research that looks more broadly at the integration of machine learning and formal system analysis. In collaboration with the DES unit, we will work on practical and theoretical aspects of this integration. Practical aspects that we will consider include the use of reinforcement learning techniques for the synthesis of optimal controllers, and the use of Bayesian networks for system diagnostics. Theoretical challenges are posed by the fact that in the domain of formal system analysis usually strong correctness or optimality guarantees are required, whereas machine learning techniques often only are known to produce locally optimal solutions. We therefore aim to develop new theoretical tools that facilitate the derivation of quantitative quality guarntees for learned models.

**Structured latent variable models** We plan to continue our research on latent variable models, focusing on regular model structures inspired by the model classes that we have previously established for collaborative filtering. For this application domain, in particular, we will consider methods for integrating different types of data sources, including both structured and unstructured data as well as network data representing, e.g., social relations among customers. As part of this work we also plan to put particular em-

phasis on the interpretability of the learned latent model structures as a way of getting additional insight into the underlying data generating process and the behavior of the model.

From a methodological point of view, we plan to continue to pursue the learning task within a Bayesian learning context, where we also wish to explore non-parametric methods for estimating model complexity (i.e., the dimensionality of the space defined by the latent variables); in contrast to the more standard score-based learning techniques that are currently being used. From an application perspective, we wish to extend the developed model structures to other types of domains such as environmental modeling where we currently have some promising contacts.

**Streaming data** Our activities within streaming data was initiated in 2013/2014 with the start of the EU FP7 project AMIDST. We wish to extend and build on top of the current results achieved within this project. Specifically, we plan to consider algorithms for adapting both parameters and model structures within a streaming context, subject to both gradual and abrupt concept drift. Due to the nature of data streams, key issues in the design of these algorithms will be constraints on processing time and memory resources. Another particular challenge of this task is that the models under consideration explicitly represent and capture the temporal nature of the data. This aspect of the models makes straight-forward application of standard scalability techniques for learning/inference difficult as data-processing has to respect the temporal constraints encoded in the model.

## Intelligent Web and Information Systems

**Web Long Tail.** We would like to pursue research in *new methods for personalized ranking or recommendation methods for Web long tail*. The long tail is a large part of the Web (70 − 80%). This can be exemplified on web search engines queries where more than 55% of queries are issued only once and 97% queries are issued less than 10 times. Current methods do not treat this part of the Web very well for tasks like recommendation, ranking, task identification, ad recommendation, and similar. The reason is mainly because current methods are well suited for balanced data sets, where frequencies of items or events is distributed evenly. This is not the case for the Web data where large frequencies happen in the head and therefore methods

usually are biased to that part. In this area we would like to study:

- novel latent based methods,

- novel graph based methods,

- novel reinforcement learning methods, and

- their combination.

By doing so I would like to improve niche user preference learning and niche item recommendation and ranking to improve access to the items in the Web tail including social networks.

**Computational Methods for Web User Interfaces.**   Related to the area above, we would like to also develop *novel personalized computational methods for Web user interfaces*. This includes *explorative interfaces for news and messages on social networks* including those of niche interest not only trending topics. Further to that this includes *explorative interfaces for web querying and web search*. In this area, one of the problems is that if user issues multiple keyword query, he does not know well how returned results are related to parts of the query. This can be solved by innovative user interfaces where resources or labels representing resources or their groups are displayed so that this relevance become more obvious. Extending and visualizing long tail items in this contenxt is even more challanging. Therefore, new ways of displaying tag or keyword maps with interactive parts for users to click are needed to solve this problem and will make significant improvement to how user access large amount of information on the web.

**Understanding and Unlocking the Web**   We focus on an interdisciplinary approach to understanding and unlocking the data available on the Web, in which people play a central role. Analyzing these aspects must take account of various factors, including the diversity of data over time, with regard to content, and cultural diversity, as well as the users and their interactions. Our goal is to achieve a deeper understanding, from the standpoints of computer science and IT, social science, law, and business, of the content and structures of the Web and how these change over time, along with how they affect social

and economic processes. The expected outcome of our research includes intelligent analytical *tools* and *algorithms* that go far beyond simply observing the Web. For this purpose, we will significantly advance information filtering in data streams using a mixed crowd- and machine-based approach, which will dynamically adapt to evolving knowledge for identifying novel and relevant information. We will develop a time-aware ranking model that makes entities and events first-class citizens of retrieval in order to enhance search results. To this end, we will further support analysis and exploration using entity-centric timeline summarization based on a novel method of context-aware dynamic relationship extraction.

## 6.8 Committee Evaluation

**Observations**

The main themes of the group are machine learning and intelligent web and information systems. The work on machine learning is characterized by a long tradition in various probabilistic approaches to machine learning: the group has strong core competences in various types of probabilistic models based on graph formalisms. Natural extensions of this research has been the work on relational learning and on latent variable models. In the area of intelligent web and information systems the work has concentrated on using information from web social media and networks to obtain recommendations for the user or to make information access easier.

The strengths of the group include strong core competences in probabilistic approaches in machine learning. Over the years, the group has done interesting work in many subareas. The scope of the group is relatively wide, but to a large extent it builds on existing competences. The group has a good or even excellent publication record when compared to the number of people. The publications include very interesting papers with lots of good ideas. Also, the publication forums are good. The group has a good track record on PhD education, with seven graduated students during the period.

The clear weakness of the group is the small number of people. As the self-evaluation points out: "The unit has worked under challenging conditions in ... 2011-2015." Furthermore, the two parts of the group (machine learning

vs. intelligent web and information systems) seem to have relatively little to do with each other. Given the size of the group, there are fairly many research themes, and it is not always completely clear what the competitive advantage of the group is when compared to other groups working in the same area.

The future plans of the group continue building on the existing competences. There is a fairly long list of themes with many interesting ideas. However, the priorities of the group are not completely clear, and given the small number of people, it is not obvious how much width there should be. The opportunities of the group are many. The general theme of the group is excellent: machine intelligence is becoming more and more important. There is interesting theory to be done and great potential in applications and for societal impact. One could even say that the importance of this subarea has grown a lot in recent years.

However, the small size of the group causes problems. It can be very harmful in recruiting: it is not obvious how attractive the group is. What are the ways in which the group can attract excellent people from elsewhere? There is a clear danger that the group becomes marginal over time.

## Questions and Recommendations

The current small size of the machine intelligence group is a clear problem. More generally, one can ask whether in the department the group structure is a management structure, a structure for organizing teaching, or a research structure. There are many different possibilities, and the current situation seems to be a mixture of these three.

Machine intelligence can be viewed as a cross-cutting theme in many areas of computer science, and it is a strong point in the department. The current small group does not seem to be the best way of supporting the research in the area, and some organizational changes are needed.

Regarding the content of the work, the high quality research combining fundamental and use-inspired work in the area should be continued.

Even stronger and more influential collaborations should be sought, ones in which there are important problems where the skills of the group are useful, where there are world-class collaborators close by, where the problems are

such that computation can make a difference, and where the collaboration is fun.

There are lots of possibilities for this already inside the department, in other parts of the university and elsewhere.