

# InspiredResearch

RESEARCH NEWS FROM THE DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF OXFORD

## SECURITY: NEW LEVEL UNLOCKED

Oxford expands security research



### INVISIBLE INK IN THE COMPUTER AGE

The art and science of information hiding – p10



### PROTECTING PERSONAL BYTES

The EnCoRe cyber security project – p8



### KEEPING THE AIRWAYS CLEAR

Modelling chronic obstructive pulmonary disease – p4



DEPARTMENT OF  
**COMPUTER  
SCIENCE**

## Welcome

InSpired Research is a twice-yearly newsletter published by the Department of Computer Science at Oxford University.

If you would like to learn more about anything you read in these pages please get in touch: Department of Computer Science, University of Oxford, Wolfson Building Parks Road Oxford, OX1 3QD

editorial@cs.ox.ac.uk

## Editorial Board

Wendy Adams  
Bob Coecke  
Ivan Flechais  
Jeremy Gibbons  
Max Goulding  
Thorsten Hauler  
Muzammil Hussain  
Pete Jeavons  
Sharon Lloyd  
John Lugton  
Thomas Lukasiewicz  
Suzanna Marsh  
Alexandru Mereacre  
James Osborne  
Bogdan Panait  
Maureen York

## Contributors

Kelly Burrowes  
Stephen Cameron  
Leanne Carveth  
Sadie Creese  
Georg Gottlob  
Andrew Ker  
Sophie Kershaw  
Luke Ong  
Mehrnoosh Sadrzadeh  
Niluka Satharasinghe  
Andrew Scull  
Jamie Vicary  
Jennifer Wilkinson  
Standa Zivny

Printed on recycled paper.

# Students win prestigious international programming contest

**Four doctoral students have won first prize in the programming competition at the International Conference on Functional Programming (ICFP).**

The ICFP Programming Contest is a well-established annual competition. This year's challenge saw competitors write a program that plays 'Lambda: The Gathering', a card game that was specifically invented for the competition.

Back in June, participants had 72 hours to complete and submit their entry over the internet, with the winners being announced at the conference in Tokyo, Japan. The team – 'Eta-LONG Normal Form' – fought off competition from 198 teams from all over the world, to be crowned the winners. The team was

made up of Chris Broadbent, Martin Lester, Robin Neatherway and Steven Ramsay, who are all undertaking DPhils under the supervision of Luke Ong at the Department.

Steven commented: 'The ICFP programming competition was originally conceived to help demonstrate the effectiveness of functional programming, but achieving that goal is contingent on functional programmers actually winning the contest. In recent years, the winners have almost all used more conventional, imperative programming languages, so we're most proud of being able to reclaim the title on behalf of functional programming!'

[www.icfpcontest.org](http://www.icfpcontest.org)

## Goal! Oxford scores at Robot World Cup

**A team from Oxford recently competed in the RoboCup in Istanbul – the 'world cup' of robot football. For the first time Oxford brought real robots to the competitions – and scored the first goal in their opening match.**

The Oxford participants partnered with Newport University from Wales and Kouretes University from Crete to form the 'Noxious-Kouretes' team. Despite their early lead, their form sadly didn't continue, with the team taking 20th place in the challenge – not bad for a first attempt.

In previous years Oxford has only taken

part in leagues that used simulated robots, which allows them to develop control programs for them without worrying about the practical issues in running real robots.

This year they continued their participation in the virtual leagues: the first of which was another football league, namely '3D simulation' with the team – called 'OxBlue' – gaining ninth place.

The third league is rather different and is called the 'Virtual Rescue Robots'. This team – a joint effort with the University of Amsterdam called 'Amsterdam-Oxford Joint Rescue Force' – came in fifth.

## Career Acceleration Fellowship won

**Mehrnoosh Sadrzadeh has been awarded a five-year EPSRC Career Acceleration Fellowship. These fellowships provide support for outstanding researchers at an early stage of their career.**

They are highly sought after: each year 450 applications are received by the EPSRC for approximately 25 career acceleration fellowship awards. Mehrnoosh's

work on 'Foundational Structures for Compositional Meaning' will see her exploiting the theoretical and experimental underpinnings of recent research breakthroughs on modelling the meanings of sentences, inspired by the diagrammatic methods of modelling quantum protocols. Mehrnoosh will be working with three project partners – the Universities of Utrecht and Cambridge, and Google.

# Security: new level unlocked

**Oxford has significantly increased its research activities within the security domain, with the addition to the Department of Sadie Creese and Michael Goldsmith, formerly of the University of Warwick. Ivan Martinovic, an expert on wireless security, will be joining us from Berkeley in January.**

Their arrival coincides with the launch of the newest of the Department's seven research themes: security.

The increased use of internet-based applications, the rise in computer based crime, together with the impact of social media applications, has increased and changed the nature of security risks. Security considerations pervade personal, corporate, military, governmental, and national infrastructure systems and networks. The theme's research activities, with a particular interest in the field of cyber security, will involve protecting information, services and property

from theft, corruption, and natural disaster, whilst ensuring it remains accessible to its rightful users.

The Department is already at the forefront of security research in the UK, with research spanning areas such as intrusion detection, information flow & non-interference, pervasive (ubiquitous) computing, security protocols, trusted computing, access control and steganography.

See pages 8-11 for more on our research.

# Academies honour Oxford professors

**Head of Department Bill Roscoe has been newly elected as a Royal Academy of Engineering Fellow.**

The organisation is the UK's national academy for engineering, bringing together the most successful and talented engineers from across the engineering sectors for a shared purpose: to advance and promote excellence in the field.

Bill commented: 'It is wonderful to be recognised as an engineer, since throughout my career I have worked closely with industry at the same time as doing theoretical work. I believe

I am the first ever FEng in this department: there cannot be many university departments in the country that simultaneously have staff who are FEng, FRS and FBA, as we now do.'

Congratulations also go to Jim Woodcock. Jim now holds the Anniversary Chair in Software Engineering at the University of York, but previously founded (along with Tony Hoare) Oxford's Software Engineering Programme. John Lazar, who did his MSc and DPhil with the Department, was also elected.

**Marta Kwiatkowska and Bill Roscoe have both been elected to join Academia Europæa, an eminent organisation for European scholars.**

Academia Europæa members are scientists and scholars who collectively aim to promote learning, education and research. Membership of the organisation recognises an individual as one of the top researchers in Europe. Samson Abramsky and Georg Gottlob from the Department are already members.

# DPhil student wins Google Scholarship for Students with Disabilities

**DPhil student Sophie Kershaw has been awarded a Google Scholarship for Students with Disabilities, as one of only ten recipients across Europe in 2011-12.**

Established in 2010, the Google award 'gives recognition to outstanding scientific contributions from students with disabilities in the field of Computer Science ... to help break barriers that keep students with disabilities from entering computing and encouraging them to excel in their studies and become active role models and leaders in creating technology'.

As well as additional funding, the scholarship also includes an all-expenses-paid trip to the Scholars' Retreat at Google's Zurich headquarters, which Sophie attended in the summer. She presented her current research as part of the event, as well as attending a series of seminars on machine learning, Android phone technology and career development.

A member of the Computational Biology group, Sophie is currently in the third year of her DPhil. Her research focuses upon mathematical modelling of cancer, developing dynamical systems models for key biochemical pathways implicated in colorectal cancer and exploring the consequences



Photo credit: Alastair Kay

for tissue-level disruption. The resulting computational framework will ultimately be applied to *in silico* translation experiments, investigating the impact of tissue geometry upon the expression of subcellular biochemistry and the success of pharmaceutical drug targets.



## Keeping the airways clear: modelling chronic obstructive pulmonary disease

Oxford scientists are developing computer models to help doctors improve the diagnosis and treatment of respiratory diseases that can obstruct our airways. Their new computer simulation will amalgamate vast amounts of existing, but disparate medical knowledge of chronic obstructive pulmonary disease, as conditions such as bronchitis and emphysema are collectively known, into one place. But the real challenge comes in taking this data and translating it into information that's truly useful to doctors at the point of care: personalised models that allow medical practitioners to accurately anticipate the progress of the disease, and impact of different types of treatment on their particular patient, based on the experiences of patients with a similar profile.

### Project partners:

- Barcelona Digital Centre Tecnològic (Spain)
- Biomax Informatics AG (Germany)
- Linkcare Health Services, S.L (Spain)
- IDIBAPS, Consorci Institut d'Investigacions Biomèdiques August Pi i Sunyer (Spain)
- Karolinska Institutet (Sweden)
- The University of Birmingham (UK)
- Infermed, Ltd. (UK)
- Technical University of Budapest (Hungary)

Chronic obstructive pulmonary disease (COPD) is a narrowing or obstruction of the airways, and includes conditions such as chronic bronchitis and emphysema. COPD is common: about three million people in the UK have it, and it is estimated that another half million have the condition but have not yet been diagnosed. COPD accounts for more time off work than any other illness, and is one of the most common reasons for admission to hospital.

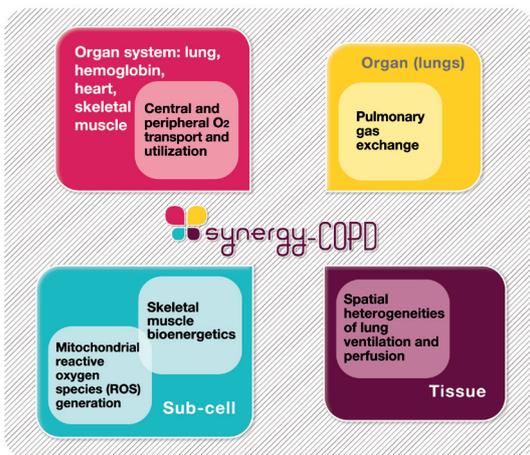
Computer scientists from Oxford, led by Kelly Burrowes,

are involved in the Synergy-COPD project: looking into the underlying mechanisms of COPD phenotypes associated with poor patient prognosis and recovery. (A phenotype is an organism's observable characteristics or traits: such as its development, biochemical or physiological properties and behaviour).

The project aims to produce a more complete computer model of the mechanisms of COPD than currently exists. Synergy-COPD's simulation will pull together a vast compendium of medical knowledge: drawing from

existing epidemiological data, clinical trials, physician interviews, and specific physiological models; and bringing it together in the form of a computer program.

This program replicates and simulates the human physiology (the complex workings of human biology), allowing doctors and medical researchers to dynamically simulate the medical problems, and (side-)effects of treatments. This is done by simulating different functions of the human body, using computer models at different physiological levels: not just modelling the organ system as a whole, but modelling at organ, tissue, cellular and sub-cellular levels. The behaviour of each physiological network can then be observed separately and in an integrated way. This represents the world's most cutting-edge technology at the interface between computer science, mathematics and the biomedical sciences.



Sitting alongside the simulation, a decision support system (DSS) will allow clinicians to anticipate the progression of the disease, using the mass of background data on the condition, but taking into account specific patient profiles – individuals' responses to the condition, and treatments. Thanks to this system of predictive and personalised medicine, patients will receive the most suitable treatment according to their personal characteristics, based on the experiences of previous patients with a similar profile.

The Synergy-COPD platform will therefore tackle the complexity of the vast amount of clinical, physiological, genetic and experimental data currently available, making it manageable and useful at the point of care. The project goes beyond conventional care and promotes the link between systems medicine and integrated care, which represent two holistic approaches that should generate efficiencies in future strategies aiming at personalised health for patients in the early stages of chronic diseases.

As well as aiding physicians, Synergy-COPD will be hugely beneficial to researchers working to improve and extend our knowledge about the physiology of the human body. Ultimately the project will also contribute towards the establishment of shorter and cheaper virtual trials of new treatment. The project will also contribute to biological research by fostering the convergence between basic and clinical sciences

## Case study

Daniela is a 55-year-old suffering from COPD. She has had periodic examinations, and the disease keeps progressing. Daniela recently had two severe exacerbations and now Doctor Norton is thinking about prescribing her with inhaled corticosteroids. He knows that they are not always effective and they often have side-effects. He decides to feed the Synergy-COPD system with Daniela's data. The system compares patterns and, following one of the rules contained in the knowledge base, checks the concentrations of some proteins in Daniela's blood. The results confirm that corticosteroids would indeed be effective for Daniela, and the Synergy-COPD system communicates so to the doctor.

by contributing to shape systems medicine and, consequently, a new paradigm in translational research.

It is hoped that the project will boost the European industry of medical knowledge management, medical simulation and decision support, and integrated healthcare, in a market currently dominated by Asian and American companies.

The Synergy-COPD project will run until 2014, and has received funding from the European Commission of €4.5 million (£3.9 million).

## And in other news...

- **Fujitsu researchers using PRISM** won the best paper award at the IEEE International Conference on Cloud Computing (CLOUD 2011) in Washington. Shinji Kikuchi and Yasuhide Matsumoto from Fujitsu Laboratories used PRISM to model and analyse performance of live migration in cloud computing systems. PRISM is a probabilistic model checker developed by the Department's Marta Kwiatkowska

and her group. It has been applied to analysing the performance, reliability, security and dependability of a wide range of computer systems and protocols.

- Thomas Lukasiwicz and Gerardo Simari have been given an \$80k (£50k) **Google Research Award** to carry out initial research on probabilistic semantic query answering on the web.

- Germany's Karlsruhe Institute of Technology (KIT) has awarded a KIT-Doktorandenpreis (roughly translated as a '**KIT Doctoral Award**') to Markus

Kröttsch. The awards recognise outstanding dissertations. Each year, there is one award in each of KIT's six main 'areas of competence'. Markus was the joint winner in the Information, Communication and Organisation category, taking away €1000 (£860) in prize money.

- Professor of Computer Science Gavin Lowe gave the keynote address at **Communicating Process Architectures 2011**, the 33rd WoTUG conference on concurrent and parallel programming, at the University of Limerick, Ireland. He spoke on 'Implementing Generalised Alt'.

## Feeling constrained? Oxford research group has the answers

**How do you switch the flow of oil around a pipeline network to meet all the demands in different places? Which ten blogs give the best overall coverage of the biggest stories on the web? Where's the best place to put sensors to quickly detect contamination in a drinking water distribution network? These are all examples of real-world problems that are now being solved by constraint satisfaction techniques.**

The Oxford Constraints Research Group, led by Peter Jeavons, is now the biggest in the UK. Their work focuses on fundamental theoretical issues, such as the complexity and expressive power of constraints, but they also work on industrial configuration problems, and finding the most efficient ways to use software tools.



Photo credit: H.Simonis

In September five members of the group made a big splash at the 17th annual International Conference on Principles and Practice of Constraint Programming in the beautiful Italian hilltop city of Perugia. Out of the 51 papers presented at the conference, five were from the Oxford group.

Georg Gottlob won the best paper award for his paper 'On Minimal Constraint Networks'. This settled a theoretical problem about the complexity of constraint solving that had been open since 1995.

Standa Zivny won the Doctoral Research award for his Oxford DPhil thesis, which deals with optimisation problems. Standa is now a Junior Research Fellow at University College, Oxford and spent part of last year as an intern at Microsoft Research. He co-authored three papers at the conference, with collaborators from France and Ireland.

Third-year DPhil student Markus Aschinger presented his work in the doctoral programme; postdoctoral researcher Conrad Drescher co-organised a workshop on configuration; fourth-year DPhil student Justyna Petke presented a poster on her work; and associated member David Cohen and former member Paidi Creed also gave research talks.



Photo credit: H.Simonis

### So what is constraints research?

Constraint satisfaction is a powerful paradigm for solving combinatorial search problems which draws on a wide range of techniques from artificial intelligence, operational research, databases, algorithms, graph theory and elsewhere. It pulls together all those computational problems from many different areas where the aim is to find values for variables satisfying certain specified restrictions and optimising certain specified criteria.

The success of constraint satisfaction lies in its generality and wide applicability. As the framework captures so many different problems in a clean mathematical framework, rather than dealing with numerous specific problems separately, they can all be treated in the same way. This allows for using general methods, structural results, and powerful mathematical tools.

## And in other news...

● Four papers co-authored by Oxford computer scientists were given at CONCUR 2011 – the 22nd International **Conference on Concurrency Theory**, in Aachen, Germany. The conference brought together researchers, developers, and students in order to advance the theory of concurrency, and promote its application...

● ...and 'Static Livelock Analysis in CSP' written by Joel Ouaknine, Hristina Palikareva, Bill Roscoe and James Worrell achieved the **best paper award** at CONCUR 2011. The paper develops methods of analysing the structure of recursions and networks to find proofs of the absence of divergence, without calculating the state space of the target system. These are implemented as a tool, called SLAP, which searches for these proofs

using either binary decision diagrams (BDDs) or SAT tools (tools for finding a satisfying solution for a Boolean formula). This type of analysis typically scales much better than state enumeration as used in FDR.

● A paper entitled 'S-Match: an Algorithm and an Implementation of Semantic Matching' co-authored by the Department's Mikalai Yatskevich has been awarded the first ever **ESWC: 7-Year Award** for its 'significant and lasting impact on the semantic web community'.

# 2020 Science: training the next generation of cutting-edge scientists

**The 2020 Science programme will train a new generation of first-rate scientists to think, work and lead across the boundaries of traditional scientific disciplines and cultures. These will be scientists who are able to apply and develop computational and mathematical modelling approaches to advance our understanding of complex natural systems – ranging from how cells work and why they go wrong, to how the interaction of climate and ecosystems regulates the planet's life support system.**

Scientists are commonly faced with the challenge of modelling, predicting and controlling multiscale systems that cross scientific disciplines. Entirely new kinds of exploratory and predictive models and research strategies are needed to address these types of scientific challenges.

A novel computational approach and environment is needed for doing this kind of science which does not exist today. Progress requires both a cultural and technological change in the way in which mathematical

and computational models, tools and software are developed, and a related change in the way in which groups of scientists are trained to develop and use these approaches.

Currently the majority of software-based research applications for the natural sciences continue to be developed for individual, single-purpose research use with little thought given to long-term sustainability, utility or extensibility outside of the immediate application domain. There has been little cross-over from software development methods and infrastructures used in industry to those used in academia.

The research aims to initiate a radical alteration in this situation making it easier for researchers to utilise experimental data sources to build and validate mathematical and computational models, and to compare competing modelling paradigms, within a consistent computational framework.

The 2020 Science programme is committed to supporting and

mentoring this new generation of scientists as they make the transition to leading their own research in either academia or industry.

The University of Oxford, University College London and Microsoft Research, Cambridge have recently been awarded £3.9 million by the EPSRC Cross-Disciplinary Interfaces Programme. At least 17 post-doctoral researchers will benefit directly from the training they receive over the lifetime of the programme.

The programme will provide training in state-of-the-art approaches to computational science whilst also conducting cutting-edge life sciences research across disciplines and institutions as well as across academia and industry.

Those who embrace this cross-disciplinary ethos are encouraged to get involved by applying for one of the research or visiting fellowships, attending their training courses, or subscribing to the e-newsletter at: [www.2020science.net](http://www.2020science.net)

## Uncovering the mysteries of quantum nanoscience

**What can philosophy tell us about the structure of quantum information? How can the foundations of quantum mechanics be tested experimentally? A new multi-disciplinary collaboration between the Computer Science, Materials, Philosophy and Physics departments at Oxford, and the Centre for Quantum Technologies in Singapore, has recently been launched to answer these questions, and more.**

The computing effort is led by Bob Coecke, along with Jamie Vicary and Raymond Lal.

The foundations of quantum computing is a subject without regard for traditional barriers between disciplines. When the

experimentalists are continually pushing our understanding of quantum materials and the quantum-classical interface, the computer scientists are revolutionising the mathematical foundations of quantum algorithmics, and the philosophers are reaching fresh insights into the meaning of quantum theory, it's abundantly clear that much progress can be made by reaching out between departments.

Unfortunately, broad collaborations of this kind can find it difficult to attract funding, too easily falling outside of the traditional project templates expected by standard funding agencies. This project is jointly funded by the John Templeton Foundation and the Centre for Quantum Technologies in Singapore. These are organisations which together support a broad

range of research in the foundations of quantum theory.

On all fronts, there are important goals in mind. From the Computer Science perspective, it is key to understand how physical properties of the world – such as causal separation, or classicality, or even the direction of time – can be represented mathematically. Such an understanding allows us to design quantum protocols that take advantage of these features of physical reality, and prove that these protocols work as they should. From an experimental perspective, we still have much to learn about how to control quantum materials, essential for the successful future implementation of a quantum computer. These investigations inform our philosophy, as we seek better ways to characterise and understand physical reality.



## Protecting personal bytes: the EnCoRe project

**The EnCoRe project is aimed at delivering revolutionary new controls for managing the risks associated with exposing personal data to the internet.**

EnCoRe stands for Ensuring Consent and Revocation and is a collaborative project funded by the Technology Strategy Board (TSB), the Engineering and Physical Sciences Research Council (EPSRC) and the Economic and Social Research Council (ESRC).

For many years we have been digitising our personal data and sharing it on the internet, sometimes to access commercial or citizen services, other times to socialise.

There is no doubt that enterprises have an ever-growing incentive and capability to collect, store and process huge quantities of personal data; understanding more about us enables the personalisation of online services, the targeting of advertising, and in some cases ensures that we have the appropriate credentials

for access. But when we share we do so without having any practical control over how our data is handled.

Once we have handed over our data it resides on technology beyond our physical and logical reach, and the small print can often mean that the service provider is committed to very little in terms of privacy and data protection.

Mechanisms to enable users to control the handling of their data once it has been shared are generally

missing, and many have simply accepted that once it is 'out there' it is effectively beyond reach.

Historically, enterprises have often been unwilling to implement such mechanisms in their databases due to the cost and the constraints that these would impose on enterprise data-handling practices. Privacy controls have only recently been introduced in large-scale information systems, and the use of privacy-impact assessments and provision of usable and practical privacy controls is still a maturing discipline.

Social networking sites include embedded mechanisms to capture users' preferences regarding their consent, which do offer some semblance of control, although the use of privacy settings in such contexts is aimed at accessibility to other social network users, not the service provider. Whilst users may consent explicitly to the service provider sharing, storing and processing data on such sites, they cannot so easily revoke (permissions to hold or process) data that they may already have disclosed.

However, it is not only individuals who are concerned with the risks associated with sharing personal data in cyberspace, the organisations

who handle such data also face risks. Enterprises manage and administer huge sets of personal data which are collected as part of normal business practice, and in many cases are also essential to their ability to deliver the expected service. Such enterprises must comply with the relevant data protection and privacy law, which if found lacking can lead to financial penalties and, possibly worse, the loss of trust in the brand.

It is likely that forthcoming revisions to the EU data protection directive will endorse the use of privacy-by-design. Although still a relatively informal concept, this can be generally interpreted to mean that systems and technology must consider the privacy of personal data as a core function, and ensure that due care and attention is given to the potential negative impacts upon such privacy that use of such systems and technology might have.

There is a clear business and societal need to increase the control and transparency surrounding the sharing and handling of personal data in cyberspace.

## Research approach

The EnCoRe project is delivering real solutions to these problems. It is a truly interdisciplinary project which is developing an approach for providing individuals with fine-grained consent and revocation controls. These controls allow the user to express their preferences through selection of options from easy-to-use interfaces. The interfaces enable the automatic creation of policies which are then used to control what the service provider can do.

Options can be used to control onward sharing to third parties, so often practiced but resulting in personal data pervading the Internet without the knowledge of the individual whose data it originally was.

The controls can also be used to oblige data handlers to notify whenever data is processed or shared, and, where the individual changes their mind and revokes their consent, the means of putting data out-of-reach can also be



selected. Through the use of 'sticky policies' such controls can be propagated down through supply chains. The options offered to individuals are chosen by the service provider, to ensure compatibility with legal, regulatory and business requirements. Individuals can then take informed choices as to which service providers they use.

The project has defined a set of principles that service providers must adhere to in order to be considered a true implementation of the EnCoRe method. Involved in this is the definition of the minimal set of options which must be offered, to ensure that business needs cannot override the core EnCoRe philosophy (and still be considered an EnCoRe service). So, for the first time, individuals will be able to engage in management of their personal data throughout the lifetime of its exposure to the internet.

## Logics and compliance

We are now in the final year of the project. The Oxford Computer Science Department's team are focusing on the final phases of research into consent and revocation logic, and the run-time compliance monitoring technology.

Our consent and revocation logic was originally conceived as a formal mechanism for understanding the requirements and complexities involved in offering a truly dynamic model which captures the range of handling rights and obligations an individual and service may need to express.

**Continued...**

## An interdisciplinary partnership

The EnCoRe project is a collaboration between Oxford's Computer Science Department and:

- HP Labs Bristol
- HW Communications Ltd
- QinetiQ
- Oxford Helex Centre
- The London School of Economics (LSE),
- and previously the University of Warwick.

The project partners provide expertise in technology, Computer Science, information assurance, privacy, law, ethics and regulation. More information can be found at: [www.encore-project.info](http://www.encore-project.info)

## Continued...

However, that logic is now underpinning a test methodology used to check the integrity of the systems being implemented. Specifically, we can use the logic to capture the requirements of such a system, and then automatically derive test suites, which can then be deployed both within the design stage and throughout the systems' lifetime.

The latter point is particularly important as all systems evolve as new functionality and services are offered, and new partners are utilised in the supply chain. Hence, an ability to check that changes do not harm the integrity of the systems' ability to deliver against the wishes of an individual is crucial.

No doubt there may be cases where requirements can no longer be met, and promised controls no longer delivered. In this case the EnCoRe methodology requires that individuals must be given the chance to opt-out, so revocating previously given consents, and the service must respond appropriately including ensuring that any revocations are propagated to third parties who have been in receipt of the affected data.

An additional control for ensuring the integrity of EnCoRe systems is the use of run-time compliance. Our research is aimed at defining the control-flows that are related, in particular, to data-sharing with third parties and obligations associated with implementing revocation choices.

The system will then be monitored in real-time in order to detect any relevant data-flows which might violate the wishes of the individual. In this way we hope to detect accidental or malicious attempts to compromise the system, particularly pertinent in complex systems where mistakes may be made.

The run-time compliance test regimes are part of the EnCoRe Compliance Framework, which comprises a range of new controls, processes and technology, and is based upon existing best practice in information risk management.



## Invisible ink in the computer age: the art and science of information hiding

**Information hiding is the art and science of making invisible alterations to an item of digital media – a picture, movie, or audio file – to embed secret messages in it.**

Enormous amounts of information can be embedded without significantly damaging the quality of the file which contains it; this is so because the human eye, or ear, has much less resolution than typically stored by a computer.

When the aim is to embed information to avoid detection it is called steganography; finding ways to detect embedded information is called steganalysis and this is related to computer forensics. When the aim is to embed information which cannot be removed without destroying the container medium, it is known as digital watermarking.

There are many applications for these technologies. Media companies are keen to embed digital watermarks in their audio and video files in attempt to deter piracy (for example to trace the originator

when a movie is leaked to the internet in advance of its release).

Medical images can be invisibly tagged with the name of the patient they come from, as a way to safeguard against errors. But most applications of pure steganography are subversive: the infamous 'Russian spies' used steganography to smuggle their communications in digital photographs.

A recent jihadist training manual includes an article on using it to conceal stored information. This makes it imperative to find better methods for detection of hidden data.

Work by the Department's Andrew Ker is based on this topic of detection. We can detect whether a photograph contains hidden data by using tools of machine learning and statistics. But the focus of our theoretical research is the theory of detection, and in particular exactly how much data can be hidden before the risk of detection becomes too great.



Andrew has formulated and proved a class of results called the square root law of steganographic capacity, which state that the asymptotic capacity of a cover is proportional only to the square root of its size.

This surprising result is in sharp contrast to capacity results for noisy channels and demonstrates that the theory of hidden information is quite different from that of Shannon, the World War II cryptographer who founded information theory. Andrew's results have been extended to embedding using adaptive source codes, with the result that true steganographic capacity can be extended to  $O(n^{1/2} \log n)$ , and no further.

These results are the first steps towards a fundamental theorem of steganography. This line of research has wide-ranging implications: for example, even the terminology with which researchers describe payload should be revisited, and standard measures such as bits per pixel or bits per transform coefficient should be reconsidered.

Another innovation has been to examine hiding in multiple objects: until now, researchers have only considered embedding and detection in single pictures, movies, and so on, but it is much more realistic to suppose that a secret payload is split between a collection of covers. The questions are how best to divide the payload between the covers, and how a detector can pool the evidence from multiple objects.

Recent work has proposed that clustering techniques could be used to identify steganographers operating within a large network of innocent users: perhaps future Russian spies will be identified when their covert messages are hidden in photographs and uploaded to Facebook.

## And in other news...

● Six members of Oxford's Department of Computer Science had four papers between them accepted to the 16th ACM SIGPLAN **International Conference on Functional Programming**, in Tokyo, Japan. The event provided a forum for researchers and developers to hear about the latest work on the design, implementations, principles, and uses of functional programming.

● 'The complexity of conservative valued constraint satisfaction problems', written by Vladimir Kolmogorov (from IST, Austria's Institute of Science and Technology) and Oxford's Standa Živný, has been accepted for publication to the 23rd Annual ACM-SIAM **Symposium on Discrete Algorithms (SODA)**, next year in Kyoto, Japan. SODA is the most prestigious conference on discrete algorithms in theoretical Computer Science.

● Michael Benedikt will be chairing PODS 2012, the top international conference on **theoretical foundations of database systems**, which will be held in Arizona, USA in May 2012.

● Thomas Lukasiewicz will be co-chairing the Seventh International Symposium on **Foundations of Information and Knowledge Systems (FoIKS)**, taking place in Kiel, Germany in March 2012.

● Markus Krötzsch, a postdoctoral researcher in the Information Systems Group, has been appointed co-chair of the 'reasoning track' for the 2012 **Extended Semantic Web Conference**, to be held in Crete in May 2012. He will also be one of two Programme Committee Chairs for the 6th International **Conference on Web Reasoning and Rule Systems** in Vienna, Austria, in September 2012.

● Final-year DPhil student Aadya Shukla won the best poster award at the recent **Cybersecurity Summit** in London. She received the award for her work entitled 'Establishing the Baseline: A Framework for Organising National Cyber Strategies'.

● Marina Jirotko has been invited by The Institute of Systems Science (ISS) of the **National University of Singapore** to be a member of a Visiting Committee of international experts to carry out an assessment of the Master of Technology (degree programme with specialisations in Software Engineering and Knowledge Engineering).

● In the last edition of *InspiredResearch* we brought you news of Edward Grefenstette and Mehrnoosh Sadrzadeh's work on quantum-inspired linguistics. They have since presented their work at the Conference on **Empirical Methods in Natural Language Processing (EMNLP11)** in Edinburgh.

## Nuffield Science Bursary student contributes to research

Sixth form student Andrew Scull joined the Department of Computer Science for four weeks earlier this year, to contribute to the Department's research activities, via a Nuffield Science Bursary.

The national initiative enables students in their first year of an advanced or higher-level STEM course to carry out a four to six week project during the summer holidays in an area of science or technology that interests them.

Andrew Scull who is currently studying for his A Levels at The Cotswold School (a Gloucestershire state school) was awarded a placement working with the group led by Daniel Kroening that is conducting research into automated techniques for formal software verification.

Andrew said: 'This is a great opportunity to learn about some of the areas of current Computer Science research directly from the people carrying it out.

Not only have I been learning about the projects but I have been able to help with some of the testing of the tools being developed. ... I will definitely be applying to study Computer Science at university.'

## Oxford team - joint winners of SVC2UK app challenge - meet PM

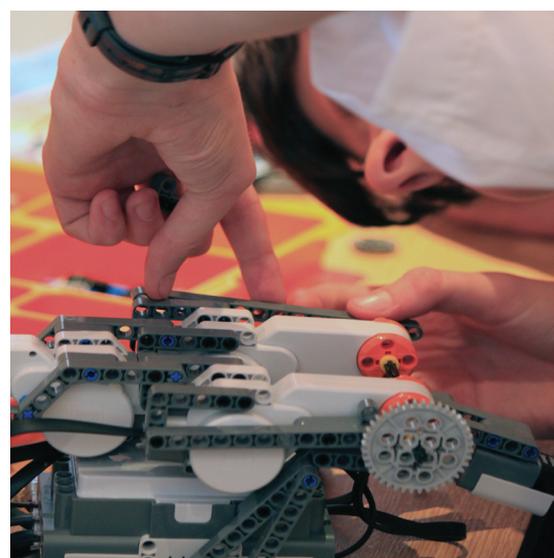
David Cameron has presented an award to the Oxford winners of the Silicon Valley comes to the UK challenge: a competition that sees university students using government data to hack together innovative apps for the public good.

The winning app from Oxford, CourseHorse, provides its users with aggregated information about local vocational training opportunities, as well as the funding is available, and enrolment information.

The team was made up of Niluka Satharasinghe (Computer Science), KC Lee (Physics), Giovanni Milandri (Engineering Science), and Fatima Sabar (International Development), who were joined by Nikhil Adhia of the London School of Economics.

The award was presented by Prime Minister David Cameron, Reid Hoffman (co-founder LinkedIn) and Sherry Coutu (TechCrunch Top CEO Mentor - Europe) at Number 10 Downing Street.

Further information about the competition: <http://svc2uk.com/>  
A YouTube video of the winning app: <http://www.youtube.com/watch?v=OLIsFyrfznQ>.



## RobotGames launched

The Department has been welcoming school teams of Year 9 and 10 pupils (age 13-15) from UK state schools to the inaugural Oxford RobotGames events.

These one-day events see students attending an introductory lecture on the science behind robots, then designing and building a wheeled robot using Lego Mindstorms, under the guidance of university staff and students. The teams then have the chance to pit their designs against the Oxford University machines in two robot Olympics-style challenges: a sprint and a hurdle event.

The winner of the first Oxford RobotGames day was The *John Warner School* from Hertfordshire.

[www.cs.ox.ac.uk/OxfordRobotGames](http://www.cs.ox.ac.uk/OxfordRobotGames)

## Make it 'appen! bada student developer challenge

Oxford's Department of Computer Science has signed up with Samsung and Steele Media to take part in their bada Student Developer Challenge 2011.

Samsung has selected ten leading UK universities to participate in this nationwide challenge, giving Computer Science students, from

first year to Masters, the opportunity to team up and develop an app for the bada platform.

The challenge is a nationwide competition for universities to develop an application for Samsung smartphones using the bada platform, and win a share of the £15,000 prize fund. The challenge

was officially kicked off in Oxford with a briefing session by Samsung's media partners, Steele Media, on 13 October. This was followed by a 24-hour codeathon in Oxford where the teams battled it out against each other and teams from across the country to develop the best apps. Other codeathons around the country continue.

# UNIQ: welcoming tomorrow's computer scientists

Over the summer the Department played host to 20 young people participating in the UNIQ Computer Science summer school. The scheme provides an opportunity for academically gifted state school students with a passion for learning to discover what it is like to be an undergraduate at the University of Oxford.

Students undertook a week-long intensive programme of academic study here in Oxford. The Computer Science stream, led by Geraint Jones, with assistance from Bernard Sufrin

and Andrew Ker, introduced some of the material that a Computer Science student would study as part of a degree at Oxford, via functional programming. The course drew on the links between sixth form mathematics and the popular science of cryptography. Half the time was given to lectures, alternating with practical sessions building up a collection of intellectual tools leading to some cryptanalysis programs.

Throughout the week there were also workshops on the application process, interviews and much more.

One hundred per cent of attendees surveyed said that the experience had made them more confident and enthusiastic about the subject, and indicated that they wanted to go on to study Computer Science at university.

The summer school is free to attend. It is open to high-performing UK state school students studying in the lower sixth form at the time of applying. Further information, including how to apply for the 2012 UNIQ summer schools: [www.ox.ac.uk/uniq](http://www.ox.ac.uk/uniq)

## Undergraduate Group Design Practicals: winners announced

The winners of the inaugural 'Group Design Practical' prizes have been announced. The IBM prize was awarded to Team Six for their smart phone application and the Ocado prize was awarded to Team One for their 'Tower Tunes' Microsoft Surface project. Each team received a cheque for £400.

The Group Design Practicals are a new element to the undergraduate Computer Science degrees at Oxford. They give second-year students the opportunity to experience working on a group development project.

The students were split into six groups of five or six people and given the choice of a project from a list of design briefs. Two teams worked with the Microsoft Surface,

and implemented an application for collaborative music-making. Two teams worked with Lego Mindstorms Robots, and constructed robots that could navigate a maze and retrieve a tin can. Two teams worked with smart phones, and implemented applications to enable software developers to work more effectively. The briefs were both open-ended and challenging leaving plenty of opportunity for students to decide how best to approach the project. Teams were allocated an academic sponsor who held scheduled meetings to discuss progress and obstacles. Students were also given seminars on team-working and how to give presentations, which were presented by IBM, Microsoft and departmental staff.

Photo credit: S Marsh



Photo credit: Son Pham

The group design practical culminated with a demonstration at the Department's Industry Showcase Day, on 1 June 2011. The teams gave presentations on their projects and also the experience of team-working, after which a small panel of judges (Oxford's Stephen Pulman, Michael Coleman of IBM and Kieren Johnson from Ocado) awarded two prizes of £400 each to the winning teams. Samsung and Microsoft also supported the Group Design Practicals.

The members of the winning groups were:

- Team Six: Lukas Bosko, Mark Brown, Ben Gazzard, Max Goulding and Stephen Parish.
- Team One: Christopher Hydon, Daniel Nichol, Kristoforos Joanidis, Ian Jones and Bogdan-Alexandru Panait.

Companies interested in getting involved in future Group Design Practicals should contact [leanne.carveth@cs.ox.ac.uk](mailto:leanne.carveth@cs.ox.ac.uk)



## Web data extraction: an interview with Georg Gottlob

Georg Gottlob joined Oxford University in 2006 as Professor of Computer Science. Before that, he was a Professor at TU Vienna in Austria, where he also co-founded the Lixto web data extraction company that provides semi-automatic tools and services for online market intelligence and web-based business process automation.



In Oxford, Georg has been continuing his work on web data extraction, and has recently been awarded a 2.4 million euro (£2 million) Advanced Grant from the European Research Council for his DIADEM project. The basic funding for DIADEM has been topped up by substantial grants from the James Martin 21st Century School, and from Yahoo! and Google. The goal of DIADEM is the full automation of data extraction from domain-specific web pages. The aim of the following interview is to better understand what web data extraction is and how it could be useful to our industrial partners. Georg is interviewed by Thomas Lukasiewicz.

**Thomas: What is web data extraction? The data on web pages are already in digital format, so why do we need to extract them?**

**Georg:** Web data is digital, but in a highly unstructured format. While a human observer can easily understand the information, a computer program cannot interpret it. In order to make web data accessible to automated analysis and to further processing, we need to extract information from the web and put it into a structured data format.

For example, the data could be mapped to schema and stored into a relational database, or it could be transformed into the XML data format and annotated by meaningful labels.

Data extraction is especially useful when large amounts of web data, for example an online catalogue with varying price data, needs to be monitored.

**Thomas: But don't most such web data actually stem from a database?**

**Georg:** Yes, you are right. But it is often impossible to access that database. Why would competitors give me access to their databases? So we need to extract the relevant data from the web. This is, in a certain sense, a process of reverse engineering.

**Thomas: In which areas is web data extraction particularly useful?**

**Georg:** The main application is in the area of online market intelligence (OMI). This covers all aspects of market intelligence that are related to online information sources. Currently, almost every large company has OMI needs for marketing and pricing.

**Thomas: Can you give me some examples of data extraction applications for OMI?**

**Georg:** Consider, for example, an electronics retailer who would like to get a comprehensive overview of the market in the form of a dashboard displaying daily information on price developments including shipping costs, pricing trends, and product mix changes by segment, product, geographical region, or competitor.

Lixto, for example, extracts such data for Fujitsu. Alternatively, consider an online travel agency offering a best price guarantee. Such a company needs to know on a daily basis, at what price similar packages are being sold by their competitors.

Moreover, they wish to be informed about the average market price of each travel product they feature. Similarly, a hotel chain may want to be informed in real-time about the prices of their own hotel rooms, and rooms of competing chains that are

marketed over the web. Lixto, for example does this currently for Best Western. But there are many types of businesses that can profit from a competitive analysis of web data: supermarket chains, car suppliers, to name just two.

There are also completely different uses of web data extraction in the business intelligence area. For example, think of a road construction corporation that would like to be informed about new public tenders in dozens of different countries or states. Such information can be extracted from known governmental web pages.

Or consider a hedge fund interested in house price index variations. Once a week, the house price index of a country is updated on the website of the country's national statistics agency. Variations of this house price index immediately trigger changes in stock prices related to the housing market. The hedge fund wishes to anticipate the house price variations by one day.

To achieve this, the hedge fund needs to obtain online market intelligence by aggregating data such as asking price, number of listings and average time on market, from the web pages of estate agents. There are hundreds of further application examples.

**Thomas: But web data extraction is also used for B2B business process automation. How does that work, and where is it applied?**

**Georg:** In supply chain management there are various web-based business processes across companies. For example, automotive original equipment manufacturers (OEMs) usually communicate with their suppliers over web pages. They put information such as quality norms on their websites, but also information for single suppliers, such as requests, notices, bills or complaints.

Of course it is vital for suppliers to obtain and manage this information as it arises, and to answer requests timely. Modern web data extraction tools help a supplier to automate this process. They continually monitor the OEM websites and send relevant changes to an appropriate employee of the supplier firm, or input the relevant information into the supplier's SAP system. In addition, they can automatically fill in OEM web forms where necessary, for example when a requested lot is shipped.

**Thomas: What is currently state-of-the art in data extraction? Which tools are used?**

**Georg:** Currently, only supervised semi-automatic tools exist. These are tools that need to be trained to go to specific websites, fill in forms, and extract specific data from the result pages. We thus need to know all websites of interest in advance and train the tool on each single website. The current tools are, however, much more powerful than simple screen scrapers. They can automatically sift through complex menus, fill in forms, push buttons and perform conditional actions. They can identify complex data structures, and automatically harvest data from thousands of pages belonging to a website.

These tools have sophisticated visual interfaces that allow a programmer to quickly train them on a specific website. In fact, only a few hours are needed for teaching the system how to extract data from a specific site, and for testing the extraction process. These tools can, moreover, connect to databases, SAP systems, data warehouses, analytics packages and the like, and deliver the extracted data into them.

**Thomas: And what is your current research in the DIADEM project about?**

**Georg:** In many applications there is only a limited number of websites of interest, and thus supervised tools suffice. For example, there are not

that many supermarket chains or internet travel agencies in the UK. So, it is not too hard to train the tool on each such site. However, for other applications there is a huge number of websites. There are, for instance, about 15,000 estate agents in the UK, some selling only a handful of properties, and there are more than 30,000 restaurants. It would be impossible, or at least extremely expensive to train a data extraction tool on each of these sites.

What would be needed is a system that can automatically navigate websites from a certain application domain and extract the relevant data fully automatically.

Take, for example, the housing market domain. We would need a system that is able to identify search forms on the homepages of estate agents, fill in these forms so to get to the pages where the properties are actually listed, recognise and extract the property descriptions on that pages, and extract and decompose each property for sale into its components such as location, price, number of bedrooms, and so on. A human has enough common sense to do this.

For a computer, this is, of course, a very difficult task. However, we think that this is possible, and this is what we want to prove in the DIADEM project. Our first results are very encouraging, and we will soon share them with the public.

**Thomas: I wish you much success. But coming back to existing technology, how can you help our industrial partners?**

**Georg:** Well, we can share our knowledge on data extraction through consulting. Moreover, we can build proof-of-concept prototypes and advise on whether a desired extraction task is feasible, and which resources they would require.

**Thomas:** Thank you for this interview.

Georg Gottlob can be contacted at [georg.gottlob@cs.ox.ac.uk](mailto:georg.gottlob@cs.ox.ac.uk)  
Further information: [www.lixto.com](http://www.lixto.com)





## Fighting childhood pneumonia in Nepal

**Nearly a million children die from pneumonia each year, making it a leading cause of death and the single most important health issue facing children under the age of five. The standard vaccination schedule calls for three doses of pneumonia vaccine given at 6 weeks, 10 weeks, and 14 weeks of age. But administering the vaccine at such an early age also reduces how long the vaccine protects the child.**

The Oxford Vaccine Group is conducting a trial in Nepal to determine whether shifting the vaccination schedule can extend childhood immunity until the critical five-year point. For the trial, the team is scheduling the first 2 doses to be given at 6 weeks and 14 weeks, but the third dose is given much later, at 8 months of age.

The team is hopeful that delaying the final vaccination will protect children for much longer, thus reducing the mortality rate from this serious disease.

One of the biggest problems in medical informatics is keeping track of the data during a trial. Researchers must meticulously log who collected each piece of information, how it was collected, and any associated details.

Manually inputting this data takes time away from actual research and is prone to error, while incomplete entries may cause problems for other researchers who refer to the material later.

That's where a team from Oxford University's Department of Computer Science, led by Jim Davies, come in. They are working on software support for medical informatics and finding ways to simplify the data collection process and reduce the risk of errors.

With support from Microsoft Research, the team developed CancerGrid, a system to manage all the diverse data that are associated with a clinical trial. Each data item to be collected is associated with a clearly-defined semantic label so that the precise meaning will be clear to clinical staff, and researchers can be certain that any two trials that use the same semantic label for an item of data are recording exactly the same thing. This makes it possible to reuse and combine data, making each trial far more valuable to researchers.

Tools integrated with Windows Azure and Microsoft Excel, SharePoint and InfoPath are used to collect and organise the data, providing easy and intuitive access to data and

implementing rules to ensure that critical data is recorded consistently and accurately.

Forms, databases, and the associated infrastructure for each new trial can be generated at the touch of a button, permitting the deployment of trial support infrastructure in a fraction of the time and a fraction of the cost of conventional methods.

It is this flexibility and automation that made it possible for CancerGrid to meet the needs of the Oxford Vaccine Group, rapidly generating full document management support for the Nepalese pneumonia vaccine trial.

By using a secure internet connection, researchers in Nepal now transmit data back to the University of Oxford, where it can be analysed and the effectiveness of the new vaccination regime assessed.

Microsoft Research have made a Research Connections film about this project entitled 'Adjusting Childhood Pneumonia Vaccination Periods May Save Lives in Nepal'.

The full video, including interview clips with Jim Davies can be seen at: <http://research.microsoft.com/apps/video/default.aspx?id=154165>