# 2016/17 Vacation Scholarships

|  |  |
| --- | --- |
| **Job Title:** | CSIRO Undergraduate Vacation Scholarships – Data61 |
| **Reference No:** | 24344 |
| **Classification:** | CSOF1.1  |
| **Stipend:** | $1462.77 per fortnight (before tax) |
| **Location:** | Please refer to the list of ***Projects*** at the end of this document |
| **Tenure:** | 8 to 12 weeks from November 2016 to February 2017 |
| **Role Purpose:** | The 2016/17 Vacation Scholarship Program is designed to provide students with the opportunity to work on real-world problems in a leading R&D organisation.Participation in the Vacation Scholarship Program has influenced previous scholarship holders in their choice of further study and future career options. Many have gone on to pursue a PhD in CSIRO or to build a successful research career within CSIRO, a university or industry. |
| **Project Description:** | Please refer to the list of ***Projects*** at the end of this document. *If you require more information please contact the person listed for the project.* |
| **Eligibility/** **Pre-Requisites:** | To be eligible to apply you must be an Australian or New Zealand Citizen, Australian Permanent Resident or an international student who has full work rights for the 8 to 12 weeks duration (does not require visa sponsorship).Vacation scholarships are for students who:* are currently enrolled at an Australian university;
* have completed at least three years of a full-time undergraduate course (however exceptional second year students may be considered);
* have a strong academic record (credit average or higher); and
* intend to go on to honours and/or postgraduate study.
 |
| **How to Apply:**  | **You will be required to:**1. select your **top 2 research projects** in order of preference;
2. submit a **resume/cover letter** (as one document) which includes:
* the reasons why the research project/s you have selected are of interest to you; and how your previous skills/knowledge and experience meets the project requirements; and
* an outline of your longer-term career aspirations and detail how this program will help you achieve them.
1. upload your **academic results** in the ‘***Requested Information’*** field.

**Referees:** If you would like to include referees (either work or university lecturers/ tutors)in your application, please add their name and contact details into your resume**.** If you experience difficulties applying online call 1300 984 220 and someone will be able to assist you. Outside business hours please email: csiro-careers@csiro.au. *Please do not email your application. Applications received via this method may not be considered.* |

|  |  |  |
| --- | --- | --- |
| **Project No.** | **Location** | **Project Title (see the following pages for more information)** |
| [**Data61 1**](#_Data61_1) | Eveleigh, Sydney | Smart Infrastructure Systems  |
| [**Data61 2**](#_Data61_2) | Eveleigh, Sydney | Automated machine learning testing framework. |
| [**Data61 3**](#_Data61_3) | Eveleigh, Sydney | Data visualisation and transformation framework for machine learning  |
| [**Data61 4**](#_Data61_4) | Eveleigh, Sydney | Behavioral analysis of personality traits |
| [**Data61 5**](#_Data61_5) | Eveleigh, Sydney | Understanding Fitness App Users – Log Analysis and User Modelling |
| [**Data61 6**](#_Data61_6) | Eveleigh, Sydney | GIS |
| [**Data61 7**](#_Data61_7) | Dutton Park, QLD | Immersive Visualisation for Science and eLearning |
| [**Data61 8**](#_Data61_8) | Canberra City, ACT | Diagnosis from Microscopy Imaging |
| [**Data61 9**](#_Data61_9) | Kensington, NSW | Automatic Translation of Routing Protocol Specifications |
| [**Data61 10**](#_Data61_10) | Kensington, NSW | Automating Formal Proofs |
| [**Data61 11**](#_Data61_11) | Kensington, NSW | Formalising and Analysing Blockchain Protocols |
| [**Data61 12**](#_Data61_12) | Kensington, NSW | Build the world's first secure network stack. |
| [**Data61 13**](#_Data61_13) | Kensington, NSW | CAmkES on Linux |
| [**Data61 14**](#_Data61__14) | Kensington, NSW | eChronos Art Project |
| [**Data61 15**](#_Data61__15) | Kensington, NSW | Formal Verification of multi-threaded embedded application software |
| [**Data61 16**](#_Data61__16) | Kensington, NSW | Fuzz testing a new language and compiler |
| [**Data61 17**](#_Data61__17) | Kensington, NSW | Graphical Editor for Building Componentised Operating Systems |
| [**Data61 18**](#_Data61__18) | Kensington, NSW | Implement and Verify a CakeML Compiler Optimisation |
| [**Data61 19**](#_Data61_19) | Kensington, NSW | Implement and Verify Enhancements to CakeML |
| [**Data61 20**](#_Data61__20) | Kensington, NSW | Improving automation in concurrent software verification |
| [**Data61 21**](#_Data61_21) | Kensington, NSW | Linear type inference in Cogent language |
| [**Data61 22**](#_Data61__22) | Kensington, NSW | Model Checking of Mesh Network Routing Protocols |
| [**Data61 23**](#_Data61_23) | Kensington, NSW | Modelling Routing Protocols |
| [**Data61 24**](#_Data61_24) | Kensington, NSW | POSIX environment for the seL4 microkernel |
| [**Data61 25**](#_Data61_25) | Kensington, NSW | Protected-Mode eChronos |
| [**Data61 26**](#_Data61_26) | Kensington, NSW | ROS native on seL4 |
| [**Data61 27**](#_Data61_27) | Kensington, NSW | Sloth vs eChronos |
| [**Data61 28**](#_Data61_28) | Melbourne | Near Real-Time OC-SVM for detecting high dimensional anomalies |
| [**Data61 29**](#_Data61_29) | Parkville | In silico Design of Bimetallic Nanoparticles and Their Catalytic Applications |
| [**Data61 30**](#_Data61_30) | Parkville | Convolutional Neural Networks Models of Nanomaterials Performance |
| [**Data61 31**](#_Data61_31) | Parkville | Intermolecular Interactions with Polarisable Quantum Monte Carlo-Molecular Mechanics (QMC/MM) Method. |
| [**Data61 32**](#_Data61_32) | Parkville | Molecular geometry optimisation with Quantum Monte Carlo. |
| [**Data61 33**](#_Data61_33) | Spring Hill, Queensland | Business Process Data Compliance Verification |
| [**Data61 34**](#_Data61_34) | Spring Hill, Queensland | Business Process Management Workflows in Blockchain Systems |
| [**Data61 35**](#_Data61_35) | Canberra City, ACT | 3D Web Tools |
| [**Data61 36**](#_Data61_36) | Canberra City, ACT | 3D Interactive Techniques |
| [**Data61 37**](#_Data61_37) | Canberra City, ACT | Novel Urban Visualisation |
| [**Data61 38**](#_Data61_38) | Marsfield, Sydney | Deep learning based object detection and classification |
| [**Data61 39**](#_Data61_39) | Marsfield, Sydney | Blockchain-based B2B Collaboration |
| [**Data61 40**](#_Data61_41) | Eveleigh, Sydney | Decentralizing big data processing |
| [**Data61 41**](#_Data61_41) | Eveleigh, Sydney | Performance Analysis of Blockchain-based Systems |
| [**Data61 42**](#_Data61_42) | Eveleigh, Sydney | Reputation Mechanism on Blockchain-based Decentralised Systems |
| [**Data61 43**](#_Data61_43) | Eveleigh, Sydney | Bitcoin/Blockchain-driven Systems |
| [**Data61 44**](#_Data61_44) | Eveleigh, Sydney | Big Data Provenance |
| [**Data61 45**](#_Data61_45) | Eveleigh, Sydney | Dependable Auditing on Operations of in-Cloud Applications |
| [**Data61 46**](#_Data61_46) | Eveleigh, Sydney | Continuous Deployment for Big Data Analytics Applications |
| [**Data61 47**](#_Data61_47) | Eveleigh, Sydney | Dependable Blockchain Crowd-Funding Application |
| [**Data61 48**](#_Data61_48) | Eveleigh, Sydney | Exploring the risks of software monoculture |
| [**Data61 49**](#_Data61_49) | Canberra City, ACT | Block chain on Cloud |
| [**Data61 50**](#_Data61_50) | Marsfield, Sydney | Self-adaptive IoT |
| [**Data61 51**](#_Data61_51) | Canberra City, ACT | Home IoT and Security |
| [**Data61 52**](#_Data61_52) | Spring Hill, Queensland | Rule-Based Reporting System (RuleRS) |
| [**Data61 53**](#_Data61_53) | Canberra City, ACT | Pig+ |
| [**Data61 54**](#_Data61_54) | Marsfield, Sydney | Cryptographically protected Cloud Data |
| [**Data61 55**](#_Data61_55) | Marsfield, Sydney | Fighting Ransomware on Mobile Devices with Document Randomization and Encryption |
| [**Data61 56**](#_Data61_56) | Floreat | Real forecasting with consideration of data uncertainty |
| [**Data61 57**](#_Data61_57) | Hobart, Tasmania | Workload Analysis Toolkit |
| [**Data61 58**](#_Data61_58) | Clayton, VIC | Skeletal motion capture using Microsoft Kinect™ for sports and rehabilitation simulations |
| [**Data61 59**](#_Data61_59) | Clayton, VIC | Hedging FX risk LSV-style! |
| [**Data61 60**](#_Data61_60) | Clayton, VIC | Virtual and Augmented Reality Visual Analytics for Computational Modelling and Simulation  |
| [**Data61 61**](#_Data61_61) | Clayton, VIC | UX Design for Graphical Workflow Software |
| [**Data61 62**](#_Data61_62) | Clayton, VIC | OurClimate |
| [**Data61 63**](#_Data61_63) | Clayton, VIC | Evaluation of a Microstructure Model of Titanium for Additive Manufacturing |
| [**Data61 64**](#_Data61_64) | Kensington, NSW | Fair Allocation of Chores |
| [**Data61 65**](#_Data61_65) | Marsfield, Sydney | Text Mining to Assist Physicians in Patient Diagnosis and Treatment. |
| [**Data61 66**](#_Data61_66) | Marsfield, Sydney | Natural Language Queries to Structured Data  |
| [**Data61 67**](#_Data61_67) | Sandy Bay, TAS | Augmented Human-Bee Interaction |
| [**Data61 68**](#_Data61_68) | Clayton, VIC | Mobile IDE for OpenIoT platform |
| [**Data61 69**](#_Data61_69) | Clayton, VIC | Applying Machine Learning to the Design of New High Performance Granular Materials  |
| [**Data61 70**](#_Data61_70) | Clayton, VIC | Exploring high dimensional data sets in Virtual Reality Environments |
| [**Data61 71**](#_Data61_71) | Sandy Bay, Tasmania | 3D Data Management (VoxelNet) for Intelligent Mining |
| [**Data61 72**](#_Data61_72) | Eveleigh, Sydney | Radically Transparent Data Logging |
| [**Data61 73**](#_Data61_73) | Eveleigh, Sydney | Privacy Preserving Voting |
| [**Data61 74**](#_Data61_74) | Melbourne (CBD or Clayton) or Geelong | ASPIRE to engage with industry |
| [**Data61 75**](#_Data61_75) | Marsfield, NSW or Clayton, Vic | Capability extraction |
| [**Data61 76**](#_Data61_76) | Eveleigh, Sydney | Automatic RDB2RDF schema mapping  |

Select the **Project Numbers** above to take you directly to the project details. Pease read though these and decide **which 2 projects are your preferred choices** as you will need to enter these into your application. If you require more information please contact the person listed for each project.

**Note**: CSIRO are advertising vacation scholarships by the different business units we have. You can apply for more than one CSIRO business unit, but your application for Data61 should only refer to Data61 projects, such as Data61 1, Data61 2, etc.

|  |  |
| --- | --- |
| Project No. | **Data61 - Vacation Scholarships Project Details** |
| Data61 1 | **Project Title**Smart Infrastructure Systems**Project Description**Data61 has instrumented the Sydney Harbour Bridge with a custom sensing platform, using vibration measurements to determine structural health. The deployment consists of 800 embedded linux sensing nodes (each with 3 accelerometers). Signal processing and machine learning classification is performed in real-time on each node, with a network connection to backend software at Data61 providing the asset owner (NSW Roads and Maritime Services) with a dashboard to monitor the bridge. A new research project is also underway to instrument two more bridges in NSW with both accelerometers and strain gauges, connected to commercial data logging hardware. This will feed data to a new cloud-based backend with flexible signal processing and machine learning. Based on experience with both the Data61-developed hardware on the Sydney Harbour Bridge, and also commercial data acquisition equipment, a need for a new platform has been identified. While this is likely to require full embedded linux to provide high levels of processing and flexibility for research, there's also commercial advantages to deploying simpler and lower cost hardware. This is a pilot project to connect low-cost sensors to Data61 's cloud-based stream processing system using MOTT communication to a 'gateway' node.**Project Duties/Tasks*** Create a network of low-cost, resource-constrained sensors using a lightweight communication protocol (preferably MOTT but potentially CoAP or similar). The sensors will be of various types (i.e. temperature, acceleration, strain).
* Collect the data produced by these sensors on a 'gateway' machine and demonstrate some basic preprocessing (i.e. sliding window filtering, dimensionality reduction).
* Push the preprocessed data into our stream processing system.

The network must have the following properties:* Robustness. If one or many of the sensors fail, the network must continue streaming information from all other sensors. If the gateway loses internet connectivity, it must cache data until connectivity is restored.
* Flexibility. It must be possible to add or remove sensors to/from the network without taking the system offline or manually altering its configuration.

The following properties are optional but desirable:* Configurability - the gateway machine should be able to start/stop data collection from any attached sensor.

**Relevant Fields of Study*** Computer Systems Engineering
* Computer Science
* Electronics Engineering

**Location:** Eveleigh, NSW**Contact:** Ben Barnes via phone on (02) 9490 5642 or email ben.barnes@data61.csiro.au  |
| Data61 2 | **Project Title**Automated machine learning testing framework.**Project Description**When a new machine learning algorithm is developed, its performance needs to be checked and guaranteed in some way. Typically this validation is a manual process where the algorithm is run on a standard dataset and compared against a suite of other standard machine learning algorithms. We would like you to automate this process, much in the same way as a unit testing framework automates code testing. That is, it should be easy for a user to quickly script up a test of their algorithm on pre-canned problems (i.e. for regression, clustering, etc). The output of this test will be the rank of the algorithm's performance against other standard algorithms, and a warning if the algorithm has significantly under-performed itself on a previous run of the test (e.g. regression testing). Once the basic framework is established, a UI could be created for monitoring test status, and/or it could be integrated with existing continuous integration tools (e.g. travis). Ideally the candidate will be familiar with, or wishing to gain familiarity with tools such as:* A variety of machine learning algorithms
* Python
* Numpy/Scipy
* Scikit Learn Algorithms and Pipelines
* Plotly, MPLD3 or some other browser based visualisation library

**Project Duties/Tasks*** Establish benchmark machine learning algorithms and datasets for a variety of common machine learning problems.
* Automate the application of these algorithms, and user algorithms, to these problems, and report results for a variety of scoring metrics.
* Cache previous runs of this framework to make sure performance of user algorithms does not significantly degrade

**Relevant Fields of Study*** Computer Science
* Engineering (Software, Mechatronic)
* Science (Physics)

**Location:** Eveleigh, NSW**Contact:** Dr Daniel Steinberg via phone on 02 9490 5520 or email daniel.steinberg@data61.csiro.au  |
| Data61 3 | **Project Title**Data visualisation and transformation framework for machine learning.**Project Description**Machine learning can be a powerful tool for pattern recognition and making predictions from data. However, the success of many machine learning algorithms can be strongly tied to the types of assumptions they make about the data. For instance, many regression algorithms assume the training targets have zero mean (Gaussian Process regression), and many clustering algorithms assume the each of the dimensions of the data to be clustered have similar scales (kmeans). So naturally, a large part of a machine learning practitioner's time is consumed by transforming "raw-data" to try and achieve the best performance possible from a machine learning algorithm. In fact, knowing which transformations are optimal in which circumstances is one of the "tricks of the trade" in data science. In this project you would construct a framework in which a user can quickly visualise the statistics of each dimension of an input dataset independently of, or jointly with, other dimensions. The framework that you create will also contain many transformation classes, so that a user of the framework can quickly transform the data they are analysing. The user will be able to save the subsequent workflow from this framework and apply it to new data in much the same way as scikit-learn pipelines operate. Preferably the framework will include a browser-based interface that could be interoperable with Jupyter Notebooks. Ideally the candidate will be familiar with, or wishing to gain familiarity with tools such as:* Python
* Numpy/Scipy
* Matplotlib
* Scikit Learn Pipelines
* Plotly, MPLD3 or some other browser based visualisation library
* Some javascript

**Project Duties/Tasks*** Make a framework for visualising aspects of high dimensional data
* Incorporate the ability to rapidly and selectively apply transformations to this data
* Allow the workflows built from this framework to be saved and automatically called by machine learning pipelines

**Relevant Fields of Study*** Computer Science
* Engineering (Software, Mechatronic)
* Science (Physics)

**Location:** Eveleigh, NSW**Contact:** Dr Daniel Steinberg via phone on 02 9490 5520 or email daniel.steinberg@data61.csiro.au |
| Data61 4 | **Project Title**Behavioural analysis of personality traits.**Project Description**Personality questionnaires backed by psychology theories have long been used to classify people into broad psychological groups. This topic explores the feasibility to determine a person’s personality without the need for a dedicated questionnaire, relying on physiological and behavioral indicators instead. Capturing signals such as pupil dilation, skin conductance or simply finger activity on a mobile device during specific tasks such as gaming or looking at photos may be enough to produce an accurate picture of the personality.The topic will involve experiment design and conduct, data collection, and analysis using machine learning techniques. The successful candidate for this project will work with CSIRO researchers in Data 61 at the Australian Technology Park in Sydney, who have expertise in human-computer interaction, machine learning and psychology. The student should be proficient in a programing language such as Python, Java, or C#.**Project Duties/Tasks*** Experiment design, including task design
* Experiment conduit, in Data61’s lab
* Data acquisition using physiological and behavioral sensors
* Data analysis using statistical and machine learning techniques
* Production of a report and presentation
* Ideally, publication of a scientific paper

**Relevant Fields of Study**Electrical Engineering, etc* Computer Science
* Data Science/Analytics
* Machine Learning
* Statistics

**Location:** Eveleigh, Sydney**Contact:** Ronnie Taib via email ronnie.taib@data61.csiro.au |
| Data61 5 | **Project Title**Understanding Fitness App Users – Log Analysis and User Modelling.**Project Description**Mobile fitness apps have flooded the market in recent years, and the uptake of innovative services aimed at delivering behavior change is increasing. While app download numbers are high, many services fail to engage users sufficiently to see true behavior change and fitness changes. Sustained engagement with fitness services is not only vital in ensuring users achieve their fitness goals, but is often a key component in the business model of the service provider. The key step in delivering engaging services is to understand you user group, and their interaction with the service. This understanding can then be taken to provide personalised, more engaging services.The successful candidate for this project will work with CSIRO researchers in the eHealth Research Center and Data 61, in collaboration with an Australian app company to analyse statistics, data, and interaction logs gathered through a real-life fitness app with over 200,000 downloads to date. The analysis will investigate app usage trends, user retention and user behaviours to provide a better understanding of current usage and to inform strategies to sustain user engagement.**Project Duties/Tasks*** Define research scope with CSIRO and industry partner
* Data cleansing and preparation for analysis
* Conduct analysis using behavioural data analytics tools
* Produce analysis insights and deliver to stakeholders
* Design strategies to increase user engagement

**Relevant Fields of Study*** Computer Science
* Data Science/Analytics
* Machine Learning
* Statistics

**Location:** Eveleigh, Sydney**Contact:** Shlomo Berkovsky via phone on 02 9490 5743 or email shlomo.berkovsky@csiro.au |
| Data61 6 | **Project Title**GIS.**Project Description**The supply of clean and safe drinking water is paramount to human development, and comes with many issues ranging from long-term supply planning, to infrastructure construction, chemical processing and demand management. Data61 works with Sydney Water to optimise water treatment and pumping using machine learning analytics.The student will design and implement a framework allowing data and machine learning analysis output to be visualised and interacted with in a browser (or an app). The data is stored in a geographic information system (GIS) database accessed by web clients through a GeoServer instance. The student will work with CSIRO researchers in Data 61 at the Australian Technology Park in Sydney and should be proficient in a Python and Javascript. Some experience with GIS systems is highly desirable.**Project Duties/Tasks*** Convert research questions into code design allowing to answer them
* Implement and test the design
* Propose improvements to current water distribution based on data analysis
* Produce a report and presentation, possibly a scientific paper

**Relevant Fields of Study*** Computer Science
* Data Science/Analytics
* Machine Learning
* Civil and environmental engineering

**Location:** Eveleigh, Sydney**Contact:** Ronnie Taib via email ronnie.taib@data61.csiro.au |
| Data61 7 | **Project Title**Immersive Visualisation for Science and eLearning.**Project Description**Visualisation is an important tool that aids in understanding the results of experiments, simulations, and is indispensable to both the verification process and the deep comprehension and communication of the data. This project will aim to focus on the immersive data visualisation, to explore and interact with the underlying data sets coming from (Big Data) analytical engines. In the project, we will also investigate various method to improve the communication of complex science to general public through interactive techniques. We will use Virtual Reality techniques as a medium to achieve all those goals. Emphasis will be also put on developing e-learning demonstration to check old, well know hypothesis: “if you read a book you remember 10% of it, if you attend a lecture you remember 20% of it, if you experience something yourself you remember 80-95% of the experiments”.**Project Duties/Tasks*** Literature analysis, design and prototyping
* Creating interactive Virtual Reality visualisation demonstraction
* Small-scale HCI evaluations

**Relevant Fields of Study*** Electrical Engineering
* Computer Science
* Software Engineering
* Computer Vision
* Predictive Analytics

Basic experience with C/C++, OpenGL, Unity 3D, or OpenScenegraph is required.**Location:** Dutton Park, QLD**Contact:** A/Professor Tomasz Bednarz via phone on (07) 3833 5544 or email tomasz.bednarz@csiro.au |
| Data61 8 | **Project Title**Diagnosis from Microscopy Imaging.**Project Description**Pathology plays a vital role in Australia’s health care system as over 70% of diagnoses are made using pathology. The primary goal of this project is to develop software to identify and classify various cells from Indirect Immunofluorescence Images (IIF). In IIF, images are obtained by “staining” a biological tissue with antibodies that are tied to a fluorescent chemical compound. When illuminated producing a fluorescence image observable through a microscope that represents the distribution of the target antibodies inside the tissue. Automatic classification of IIF images will greatly improve the reliability and accuracy of current pathology methods as well as markedly shortening the time from pathology sampling to diagnosis.**Project Duties/Tasks*** **Cell detection.** The goal here is to tailor an object detection algorithm to detect cells in an image
* **Cell classification.** Once cells inside an image are detected, a robust representation for the purpose of classification must be attained. One possibility is to exploit recent advances in texture analysis (in particular deep networks) to address our problem

**Relevant Fields of Study*** Computer Science
* Electrical Engineering

**Location:** Canberra City, ACT**Contact:** Mehrtash Harandi via phone on (02) 6218 3752 or email mehrtash.harandi@data61.csiro.au |
| Data61 9 | **Project Title** Automatic Translation of Routing Protocol Specifications **Project Description**Wireless Mesh Networks (WMNs) are a promising technology that is currently being used in a wide range of application areas, including Public Safety, Transportation, Mining, etc. Typically, these networks do not have a central component (router), but each node in the network acts as an independent router, regardless of whether it is connected to another node or not. They allow reconfiguration around broken or blocked paths by "hopping" from node to node until the destination is reached. Unfortunately, the performance of current systems often does not live up to the expectations of end users in terms of performance and reliability, as well as ease of deployment and management. We explore and develop adaptive network protocols and mechanisms for Wireless Mesh Networks that can overcome the major performance and reliability limitations of current systems. Model checking is a technique for automatically verifying correctness properties of finite-state systems in general and WMNs in particular: given a model of a WMN routing protocol, a model checker like UPPAAL can test automatically whether this model satisfies a given specification. In order to solve such a problem, both the model and the specification are formulated in some precise mathematical language. But even if a precise description of a protocol is given e.g., in process algebra, this description is usually not accepted by a model checker. This is due to incompatibility of different languages. **Expected Outcome:** Prior work involved the implementation of translation software from a process algebra language to a model description for UPPAAL. Although large parts of routing protocol can now be automatically translated, some difficult language constructs so far require human intervention. The goal of this project is to further automatise the translation effort, ultimately making it fully automatic **References:** UPPAAL: <http://www.uppaal.org/>**Project Duties/Tasks*** Analyse existing translation
* Analyse missing cases
* Implement missing steps of the translation
* Use developed software on several case studies
* Present results to the group at the end of the project; write project report

**Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer Science;
* Experience in Java as well as the capability of understanding formal specifications is necessary

**Location:** Kensington, NSW**Contact:** peter.hoefner@data61.csiro.au |
| Data61 10 | **Project Title** Automating Formal Proofs**Project Description**Isabelle is an interactive theorem prover that combines automated and human reasoning through the use of proof methods. These methods allow users to make high-level decisions on how to progress in a proof without worrying about the formal details. Powerful proof methods are required for large proof undertakings, such as the L4.verified project at Data61. Eisbach is an in-development prototype of a high-level language for writing proof methods in Isabelle, which is historically done in Standard ML. It influenced by Coq’s Ltac, but distinguishes itself by leveraging Isabelle’s existing automation and backtracking infrastructure. The aim of this project is to investigate useful applications for Eisbach, with several possible domains to explore: An existing suite of Data61- developed separation logic tactics could be re-implemented and extended, or a similar investigation can be done against the verification-condition generator used in the L4.verified proof; Or a novel proof method can be developed solve, for example, word arithmetic proofs that appear frequently in the verification of C programs. **Expected Outcome:**Automated proof tactics for previously manual proofs in one or more of the specified application domains, ideally reducing proof size and time for those applications. **References:** Trustworthy Systems: <http://trustworthy.systems>seL4: <http://seL4.systems>Isabelle: <http://mirror.cse.unsw.edu.au/pub/isabelle/>**Project Duties/Tasks*** Design and implement Isabelle/HOL one or more proof methods in the Eisbach language.
* Evaluate these proofs methods in a research context.
* Present results to the group at the end of the project; write project report.

**Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer Science in particular functional programming and interactive theorem proving in higher-order logic.

**Location:** Kensington, NSW**Contact:** gerwin.klein@data61.csiro.au |
| Data61 11 | **Project Title**Formalising and Analysing Blockchain Protocols **Project Description**Blockchains form a new technology supporting distributed payment systems such as Bitcoin. Several alternatives to Bitcoin have emerged in the last few years. These systems use different (versions of) protocols to ensure consensus on which are valid payments and which are not. There is a public debate on the correctness of some of these protocols. The company Stellar for instance claimed that the consensus protocols of Ripple is flawed, and came up with an improvement; but Ripple says their protocol is fine and that Stellar's protocol is not an improvement at all. The goal of this project is to formalise such claims and determine who is right. Formal models of concurrency and distributed systems play a crucial role here. **Expected Outcome:** More insight into consensus protocols for cryptocurrencies and related blockchain technology. Possibly a verdict on whether certain protocols live up to their promised correctness properties. **References:** The Ripple Protocol Consensus Algorithm: <https://ripple.com/files/ripple_consensus_whitepaper.pdf>One of the articles on the ripple-stellar debate: <http://bravenewcoin.com/news/why-the-stellar-forking-issue-does-not-affect-ripple/>**Project Duties/Tasks*** Literature Review, summarising the state of the art
* Translation of concepts gathered into a uniform, formal framework
* Relate translated results to known (in)possibility results about distributed systems
* Present results to the group at the end of the project; write project report

**Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer Science, in particular theoretical computer science.
* The candidate should have experience in applying formal methods and/or logic-based calculi.

**Location:** Kensington, NSW**Contact:** Robert.VanGlabbeek@data61.csiro.au |
| Data61 12 | **Project Title** Build the world's first secure network stack.**Project Description**Build the world's first secure network stack by writing it in our special language. At Data61, we developed a new language called Cogent for writing verified by construction software. The Cogent compiler generates C code, a formal specification describing what this code does, and a mathematical machine checked proof that the generated C code corresponds to the generated formal specification. We implemented verified by construction file systems by writing them in this language. The goal of this project is to test the feasibility of implementing a simplified network stack in Cogent. This may contribute to enriching Cogent and you would be encouraged to ask for feature requests to make implementing in this language easier. Novelty: This will be the first use of the Cogent language in the domain of network stacks. **Project Duties/Tasks*** Under direction, identify the initial scope of key areas of a network stack implementation.
* Learn the Cogent language by implementing proof-of concept networking code.
* Take a critical look at the resulting implementation to identify where Cogen might be improved.
* Present results to the group at the end of the project; write project report

**Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer Science, in particular operating systems and networking.
* Experience with low-level TCP/IP would be advantageous.

**Location:** Kensington, NSW**Contact:** Peter.Chubb@data61.csiro.au |
| Data61 13 | **Project Title**CAmkES on Linux**Project Description**The Trustworthy Systems Research group at Data61 has developed a component platform (CAmkES) for developing microkernel-based systems on seL4 (our formally verified microkernel). While CAmkES helps to ease the difficulty of developing systems on seL4, developing significantly complex systems is still hard, due to the need to develop almost everything (e.g. device drivers, network stacks, display systems, etc.) from scratch. The goal of this project is to develop a version of CAmkES for Linux. This will provide a much richer environment on which to prototype and test CAmkES systems, before porting them to run on seL4.**Expected Outcome:** A version of CAmkES targeted to Linux, and sample systems to test it. **References:** CAmkES: <http://trustworthy.systems/camkes.pml>Trustworthy Systems: <http://trustworthy.systems/>**Project Duties/Tasks*** Develop tools and infrastructure required to build Linux-based CAmkES systems
* Develop example systems to show the capabilities of CAmkES on Linux
* Develop software to link Linux-based and seL4-based CAmkES systems

**Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer Science, in particular the candidate should have completed an Operating Systems course with excellent marks and have experience with systems programming on Linux and with Python and C programming.

**Location:** Kensington, NSW**Contact:** ihor.kuz@data61.csiro.au |
| Data61 14 | **Project Title**eChronos Art Project**Project Description**eChronos is a formally-verified RTOS designed for deeply-embedded systems with no memory protection and single-mode execution. It is fairly simple to drive sensors, actuators and relays from very cheap boards, and get interesting results. This project is to take simple M4-based development boards such as the Stellaris Launchpad, hook them up to ultrasonic sensors, motors and LEDs to produce an artistic installation that shows off the real-time and embedded capabilities of the platform, while being aesthetically interesting. **Expected Outcome:** eChronos-based art installation suitable for open days and demonstrations. **Project Duties/Tasks*** Design an art installation; order parts; build it. This may involve soldering,
* Program the MCUs using eChronos to have interesting interactive behaviours.
* As features/bugs are found and fixed in eChronos, prepare patches for upstream.
* Present results to the group at the end of the project; write project report

**Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer science, in particular embedded systems, or
* Electrical Engineering/mechatronics

**Location:** Kensington, NSW**Contact:** Peter.Chubb@data61.csiro.au |
| Data61 15 | **Project Title**Formal Verification of multi-threaded embedded application software**Project Description**The eChronos RTOS [0] is a small embedded OS for micro-controllers. It is commercially used in medical devices and is embedded in high-assurance autonomous flying vehicles (quadcopters) demonstrated in HACMS, a large DARPA-funded project, in collaboration with industry and university partners from the US. In Data61's eChronos verification project, we aim at proving strong guarantees about eChronos correct behavior, by means of formal (mathematical) proofs, machine-checked in the Isabelle/HOL theorem prover. The long-term goal is to provide developers of embedded applications using eChronos with a formal and verified specification of the OS API functions used for synchronisation (semaphores, mutexes, etc). The challenge lies in providing the right abstraction level for the formal API and provide a usable framework for application code verification. This project will investigate eChronos-based application verification, via small case-studies, which could be derived or inspired from application code from the HACMS project, namely the SMACCMPilot open-source autopilot software from Galois Inc [3]. [0] <http://trustworthy.systems/echronos/>[1] "Controlled Owicki-Gries concurrency: reasoning about thepreemptible eChronos embedded operating system". MARS'15. [2] "Proof of OS scheduling behavior in the presence of interrupt-induced concurrency". ITP'16. [3] <http://smaccmpilot.org/>**Expected Outcome:** The expected outcome of the project is the experimental verification of a case-study application running on eChronos, exhibiting the required formalised API from eChronos**Project Duties/Tasks*** Learn about the eChronos RTOS and Owicki-Gries concurrency reasoning
* Verification of a case-study application running on the eChronos RTOS
* Present results to the group at the end of the project; write project

**Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer Science, in particular interactive theorem proving in higher-order logic

**Location:** Kensington, NSW**Contact:** june.andronick@data61.csiro.au |
| Data61 16 | **Project Title**Fuzz testing a new language and compiler**Project Description**When developing software, the compiler that translates source code to machine instructions is generally assumed to be correct. Incorrect software is more likely to be a product of bugs in the source code than bugs in the translation. However, when high assurance techniques like formal verification are applied to the source code, this relationship is reversed. The compiler becomes the weaker link. Subtle compiler bugs may lurk undiscovered, and may only be triggered when compiling particular code. **Expected Outcome:** The result of this work will be a reusable tool for generating valid Cogent programs. It is expected that this will in turn lead to a more robust Cogent compiler. **References:** Cogent: <http://trustworthy.systems/cogent.pml>Trustworthy Systems: <http://trustworthy.systems/>**Project Duties/Tasks*** Develop ‘fuzz’ tests and run them against the compiler
* Explore any problems uncovered and determine their severity
* Present results to the group at the end of the project; write project report

**Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer Science, in particular compilers, programming languages and security

**Location:** Kensington, NSW**Contact:** ihor.kuz@data61.csiro.au |
| Data61 17 | **Project Title**Graphical Editor for Building Componentised Operating Systems**Project Description**Data61's CAmkES (a component-based platform for developing microkernel-based systems on seL4) uses an Architecture Description Language (ADL) to describe the software architecture of an operating system. While the ADL helps to ease the difficulty of designing and building such a system, ADL documents quickly become too complicated to read and manipulate (in a text format) when the operating system becomes nontrivial. The goal of this project is to develop a graphical editor to design such componentised operating systems: allowing users to draw new components and connections and manipulate existing ones, then generate the code that represents their drawn system. **Expected Outcome:** A graphical editor for designing and developing component-based operating systems. **References:** CAmkES: <http://trustworthy.systems/camkes.pml>Trustworthy Systems: <http://trustworthy.systems/>**Project Duties/Tasks*** Design and develop graphical editor software
* Test and validate software
* Present results to the group at the end of the project; write project report

**Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer Science, in particular experience with GUI programming and Python and C languages

**Location:** Kensington, NSW**Contact:** ihor.kuz@data61.csiro.au |
| Data61 18 | **Project Title**Implement and Verify a CakeML Compiler Optimisation **Project Description**CakeML is a functional programming language (similar to OCaml or Standard ML) with a formal specification (a definition in logic) and a mechanically verified (proven correct) compiler and runtime system. The CakeML compiler is at the forefront of research on verified compilers for functional programming languages, aiming for end-to-end correctness with good performance and reasonable cost. It is the basis for an approach to building high-assurance computer software using interactive theorem provers. The design of the CakeML compiler allows for various optimisation passes to be implemented as local source-to-source transformations over one of the compiler's intermediate languages. Implementing and verifying one such optimisation is a good way to improve the CakeML compiler while gaining experience with theorem proving and compiler writing. There are many specific examples of optimisations to add (including: pointer equality testing, pattern-match exhaustiveness checking, splitting large functions, simplifying redundant if-expressions, improving instruction selection for conditional jumps, and encoding stack-layout bitmap indices in return addresses). Do one or more of those, or design and verify your own optimisation! **Expected Outcome:** A new compiler optimisation added to the verified CakeML compiler. Benchmarks demonstrating an improvement in the compiled code. **References:** CakeML: <https://cakeml.org/>HOL: <https://hol-theorem-prover.org>More detailed CakeML project ideas: <https://cakeml.org/projects.html>CompCert: <http://compcert.inria.fr>**Project Duties/Tasks*** Design and implement a new optimisation pass for the CakeML compiler
* Run benchmarks to assess your new optimisation's effect on performance
* Verify your new pass and update the existing CakeML compiler proofPresent results to the group at the end of the project;
* Present results to the group at the end of the project; write project report

**Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer Science, in particular functional programming
* Any familiarity with interactive theorem proving in higher-order logic, or compiler implementation would be an advantage

**Location:** Kensington, NSW**Contact:** ramana.kumar@data61.csiro.au |
| Data61 19 | **Project Title**Implement and Verify Enhancements to CakeML **Project Description**CakeML is a functional programming language (similar to OCaml or Standard ML) with a formal specification (a definition in logic) and a mechanically verified (proven correct) compiler and runtime system. CakeML is the basis for an approach to building high-assurance computer software using interactive theorem provers. The aim of the project is end-to-end correctness with good performance and reasonable cost, and ultimately scaling formal methods up to large-scale applications. Formally specifying a programming language is a serious endeavour, and the CakeML language currently lacks many features of full-blown (but underspecified) programming languages. There are opportunities for enhancements at many levels, including the concrete syntax, the type system, the built-in primitive operations, and the standard library. Examples of features we would like to add include: explicit type annotations, expression-level let-polymorphism, user-defined infix operators, nested modules, named module interfaces (signatures), and records. Specify and verify one of these features, or design your own enhancement to CakeML! **Expected Outcome:** An extension to CakeML's formal specification, and, ideally,a corresponding extension to CakeML's verified compiler. **References:** CakeML: <https://cakeml.org>HOL: <https://hol-theorem-prover.org>More detailed CakeML project ideas: <https://cakeml.org/projects.html>**Project Duties/Tasks*** Design and implement a new feature for the CakeML semantics
* Update proofs about the semantics (e.g., type soundness) to incorporate your new feature
* (Optionally, if appropriate) extend the verified CakeML compiler to support your new feature
* Present results to the group at the end of the project; write project report

**Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer Science, in particular functional programming.
* Any familiarity with interactive theorem proving in higher-order logic, or programming language semantics would be an advantage

**Location:** Kensington, NSW**Contact:** ramana.kumar@data61.csiro.au |
| Data61 20 | **Project Title**Improving automation in concurrent software verification **Project Description**The eChronos RTOS [0] is a small embedded OS for micro-controllers. It is commercially used in medical devices and is embedded in high-assurance autonomous flying vehicles (quadcopters) demonstrated in HACMS, a large DARPA-funded project, in collaboration with industry and university partners from the US. In Data61's eChronos verification project, we aim at proving strong guarantees about the correct behaviour of eChronos by means of formal (mathematical) proofs, machine-checked in the Isabelle/HOL theorem prover. The challenge lies in the concurrency caused by eChronos running with interrupts enabled, including during scheduling operations. We have successfully proved, at a model level, the correctness of eChronos scheduling behaviour in presence of interrupt-induced concurrency [1,2]. This project will look at increasing the automation and scalability of the framework. Opportunities for such improvements include increased reuse of already proven facts, investigation of more modular (but less fine-grained) approaches, more use of Isabelle automation, increased compositionality of the proof process, etc. [0] <http://trustworthy.systems/echronos/>[1] "Controlled owicki-gries concurrency: reasoning about the preemptible eChronos embedded operating system" MARS'15. [2] "Proof of OS scheduling behavior in the presence of interrupt-induced concurrency". ITP'16. **Expected Outcome:** Your work will directly impact the efficiency of the framework and proofs developed for the verification of eChronos. **Project Duties/Tasks*** Learn about the eChronos RTOS and the framework we have developed for its verification
* Explore and implement improvements to the automation and scalability of our framework
* Present results to the group at the end of the project; write project report

**Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer Science, in particular interactive theorem proving in higher-order logic

**Location:** Kensington, NSW**Contact:** june.andronick@data61.csiro.au |
|  Data61 21 | **Project Title**Linear type inference in Cogent language**Project Description**Cogent is a programming language we at Data61 designed and developed for writing provably correct file systems. The Cogent compiler, given a type-correct source program, produces an efficient C implementation, a high-level language specification in the theorem prover Isabelle/HOL, and a series of intermediate language descriptions and proofs to show that the C code is correct with respect to the high-level spec. Cogent, a purely functional language, features a linear type system, which is the key to this approach. On the downside, however, our linear type system requires many user annotations, which makes programs verbose. The goal of this project is to experiment on how linear types, together with effectful computations like memory (de)allocation, can be inferred by the compiler, potentially with the aid of model checking technologies. **Expected Outcome:** The result of this work will be a formal type inference system (possibly with compromises) and its prototypical implementation in the Cogent compiler. **Project Duties/Tasks*** Review the relevant literature.
* Formalise the type inference algorithm, and implement a prototype.
* Present results to the group at the end of the project; write documentations

**Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer Science, in particular type theory, formal methods, and functional programming.

The candidate should have experience in Haskell (or alike). **Location:** Kensington, NSW**Contact:** gabriele.keller@data61.csiro.au |
| Data61 22 | **Project Title**Model Checking of Mesh Network Routing Protocols**Project Description**As utilisation of wireless networks becomes increasingly multimedia in nature (i.e., consisting of data and video), issues around bandwidth availability and Quality of Service (QoS) in general become increasingly important. Current wireless mesh solutions, however, do not consistently meet those requirements. It is a common belief that the failure of existing wireless mesh network systems are to a large extent due to the limitations of current network protocols. As an effort to improve the performance of wireless networks, we model, analyse, verify routing protocols for wireless networks. If errors or shortcomings are found we also try to fix them and report back to the developers of the protocols. Model checking is a technique for automatically verifying correctness properties of finite-state systems in general and WMNs in particular: given a model of a routing protocol, a model checker like UPPAAL can test automatically whether this model satisfies a given specification. **Expected Outcome:** The expected outcome will be a user-friendly model checking environment; the environment should be adaptable for the specification of classes of network topologies as well routing protocol formalisations. **References:** UPPAAL: <http://www.uppaal.org/> **Project Duties/Tasks*** Set up model-checking environment for UPPAAL, suitable for streamlining and simplifying the model checking of routing protocols on certain network topologies.
* Present results to the group at the end of the project; write project report

**Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer Science, interest in formal methods appreciated

**Location:** Kensington, NSW**Contact:** peter.hoefner@data61.csiro.au |
| Data61 23 | **Project Title**Modelling Routing Protocols**Project Description**Wireless Mesh Networks (WMNs) are a promising technology that is currently being used in a wide range of application areas, including Public Safety, Transportation, Mining, etc. Typically, these networks do not have a central component (router), but each node in the network acts as an independent router, regardless of whether it is connected to another node or not. They allow reconfiguration around broken or blocked paths by "hopping" from node to node until the destination is reached. Unfortunately, the performance of current systems often does not live up to the expectations of end users in terms of performance and reliability, as well as ease of deployment and management. We explore and develop adaptive network protocols and mechanisms for Wireless Mesh Networks that can overcome the major performance and reliability limitations of current systems. To support the development of these new protocols, the project also aims at new Formal Methods based techniques, which can provide powerful new tools for the design and evaluation of protocols and can provide critical assurance about protocol correctness and performance. Close collaboration with industry partners ensures the use-inspired nature of the project. **Expected Outcome:** So far we have modelled one of the standard protocols using process algebra, namely AODV, as well as a draft successor protocol that is currently being discussed by the Internet Engineering Task Force (IETF). The project's work should include the formalisation of a second standard protocol such as OSLR or HWMP. After a faithful specification has been given, the work could include the verification of basic properties of the routing protocol: packet delivery for example guarantees that a packet, which is injected into a network, is finally delivered at the destination (if the destination can be reached). **References:** <http://en.wikipedia.org/wiki/Optimized_Link_State_Routing_Protocol><http://en.wikipedia.org/wiki/Ad_hoc_On-Demand_Distance_Vector_Routing>**Project Duties/Tasks*** Read and understand the standard of the protocol
* Model the protocol using formal methods
* Prove properties about the protocol, such as correctness
* Present results to the group at the end of the project; write project report

**Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer Science, in particular theoretical computer science.

The candidate should have interest in applying Formal Methods and logic-based calculi **Location:** Kensington, NSW**Contact:** Robert.VanGlabbeek@data61.csiro.au |
| Data61 24 | **Project Title**POSIX environment for the seL4 microkernel**Project Description**[POSIX](https://en.wikipedia.org/wiki/POSIX) is a widely supported OS interface standard, provided by Linux and many other OSes. Much real-world software can be ported to asystem that supports POSIX. The are POSIX profiles that require limited functionality, up to full Unix compliance. A good start to a POSIX environment for seL4 was made with the [Robigalia](https://robigalia.org/) project, which uses Rust as the implementation language. This project will use Robigalia as a starting point and design, implement and evaluate a complete POSIX environment for at least a minimal profile. **Expected Outcome:** A complete POSIX implementation suitable for high-assurance use. Performance evaluation against Linux. **References:** seL4: <http://sel4.systems/>Trustworthy Systems: <http://trustworthy.systems/>Robigalia: <https://robigalia.org>**Project Duties/Tasks*** Study seL4 and Robigalia
* Design and develop further POSIX features extending Robigalia
* Evaluate the implemented features
* Present results to the group at the end of the project; write project report

**Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer Science in particular the candidate should have completed an Operating Systems course with excellent results and should already have experience programming for seL4. The candidate should be able to program in C and Rust programming languages

**Location:** Kensington, NSW**Contact:** gernot.heiser@data61.csiro.au |
| Data61 25 | **Project Title**Protected-Mode eChronos**Project Description**eChronos is an RTOS that is in the process of being formally verified. It is designed for deeply-embedded systems with no memory protection and single-mode execution. However, there are interesting use cases for a verified kernel on mid-range processors that feature a simple memory-protection unit (MPU). A particularly interesting case is the ARM Cortex-M4, which eChronos already supports, albeit withoututilising the MPU. This project is to design a protected-mode version of eChronos, implement and evaluate it. **Expected Outcome:** A version of eChronos that uses memory protection**Project Duties/Tasks*** Implement memory protection primitives in eChronos
* Design a framework for assigning per-task protected memory regions during compilation and enforcing this protection at runtime.
* Evaluate the performance and effectiveness of memory protection in the eChronos RTOS environment.
* Present results to the group at the end of the project; write project report

**Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer Science, in particular embedded operating systems.

**Prerequisites*** Students must have a strong background in Operating Systems and have at least completed an Operating Systems course with excellent marks. Preferably students will have experience developing OS kernel code as well.

**Location:** Kensington, NSW**Contact:** gernot.heiser@data61.csiro.au |
| Data61 26 | **Project Title**ROS native on seL4**Project Description**[**ROS**](http://www.ros.org/)(robotics operating system) is a communication middleware that is widely used for programming robots. It typically runs on a fully-fledged OS, such as Linux, using sockets for communication. This makes it readily accessible, but from the security and safety point of view is a nightmare. The purpose of this project is to produce a native ROS on the seL4 microkernel, depending on a minimal trusted computing base. It involved an assessment of the OS services required by ROS, and design, implementation and evaluation of ROS/seL4. **Expected Outcome:** A seL4-based ROS implementation that can support the high-assurance autonomous trucks developed under the DARPA HACMS program. Performance evaluation against a Linux-based implementation**Project Duties/Tasks*** Find out the core components for building a minimal ROS
* Implement required system functions and services on seL4
* Port required libraries to seL4
* Write device drivers for the target robot
* Performance evaluation of ROS on seL4 and compare with the Linux equivalent
* Security and safety analysis of ROS on seL4
* Present results to the group at the end of the project; write project report

**Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer Science, in particular Operating System
* Electrical Engineering, in particular Embedded System
* Strong C programming skills, especially with low-level software
* Knowledge of C++, Python programming

**Location:** Kensington, NSW**Contact:** gernot.heiser@data61.csiro.au |
| Data61 27 | **Project Title**Sloth vs eChronos**Project Description**eChronos is a RTOS designed for deeply-embedded systems with no memory protection and single-mode execution. Sloth is a system for a similar application domain, which takes the unusual approach of leaving all scheduling to hardware, by running everything in an interrupt context. This limits the use of Sloth to processors where interrupt mode can be entered by software. This project is to evaluate and quantify the performance advantage of Sloth over eChronos. Novelty: Sloth is presently the world's fastest RTOS. eChronos, which has the advantage of formal verification and less dependence on hardware features, is a more traditionally-designed RTOS. This project will determine whether the performance advantage of Sloth is significant enough to justify the different (and more limiting) design. The results are eminently publishable. **Project Duties/Tasks*** Create a testbed for both systems by installing and running the OSes on relevant hardware.
* Develop and run benchmarks to analyse the performance of the systems under test.
* Present results to the group at the end of the project; write project report

This is an open research area where the student can actively contribute to better understanding RTOS design tradeoffs. **Relevant Fields of Study**This project is most suitable for students in the following field/s of study: * Computer Science, in particular experience operating systems, embedded systems, or real-time systems.

**Location:** Kensington, NSW**Contact:**  gernot.heiser@data61.csiro.au |
| Data61 28 | **Project Title**Near Real-Time OC-SVM for detecting high dimensional anomalies.**Project Description**For decades, anomaly detection has drawn tremendous attentions in diverse areas and application domains such as fraud detection for credit cards, intrusion detection for cyber-security and fault detection in safety critical systems. The state of the art techniques based on k-means, one-class SVM (OC-SVM) and neural network are very time-consuming in training detection models with high dimensional data, which cannot achieve real-time anomaly detection. This project aims to develop an OC-SVM based anomaly detection system offering near real-time detection capability with low false positive rate and high accuracy. We have done considerable work about recognising anomalies (spam) on total 600 million tweets. With the presence of over 12 million labelled spam and non-spam tweets, customised evaluations will be performed in a demo system to examine whether our new OC-SVM meets the expectations of anomaly detection in big data scenario.**Project Duties/Tasks*** Task 1: Feature Selection: to reduce the dimensionality of the data by selecting features with larger Pearson correlation coefficient.
* Task 2: Parallelization: to implement a distributed SVM training algorithm with parallelized leave one out cross-validation.
* Task 3: Qualified Boundary Points Selection: to reduce the search space of support vectors by using hierarchical clustering to identifies the qualified boundary points

**Relevant Fields of Study*** Cybersecurity
* Machine learning
* Data science

**Location:** Melbourne**Contact:** Dr. Surya Nepal via phone on (02) 9372 4256 or email surya.nepal@csiro.au or Dr Jun Zhang via phone on (03) 5227 1378 or email jun.zhang@deakin.edu.au  |
| Data61 29 | **Project Title**In silico Design of Bimetallic Nanoparticles and Their Catalytic Applications.**Project Description**The project aims to find the structural relationship of alloy, core-shell, and phase-separated bimetallic nanocatalysts with the atomic deposition rate to demonstrate highly catalytic activity. In view of the complexity of the nucleation process the formation of structures with different morphologies is one of the major issues in current research of industrial catalysts. Through molecular dynamic (MD) simulations we are planning to controllably produce bimetallic arrangements to investigate their catalytic activity in terms of their morphology and surface defects. This research will address not only some of the fundamental issues such as the effect of surface defects in catalytic efficiency but also highlight the possible industrial applications. The ultimate goal will be to understand the structure activity relationship of nanocatalysts and lay a solid foundation for the further development of novel nanocatalysts in industrial applications.The majority of catalysts used in industry involve metal oxides and precious metals as components. Nanocrystals of noble metals are very attractive for use as catalysts because of their high surface are-to-volume ratios and high surface energies, which make their surface atoms to be highly active [1]. Bimetallic nanoparticles have attracted a great interest in scientific research and industrial applications such as catalysts. From both scientific and technological point of view, bimetallic nanoparticles composed of two different elements but arranged in a wide variety of forms including alloys, core-shell, phase-separated spatial structure, are more promising than monometallic nanoparticles, because synergetic effect is expected [2]. Furthermore, bimetallic nanoparticles have shown novel catalytic behaviours based on the effect of second metal element added, which can be often explained in terms of an ensemble and/or a ligand effect in catalysis [3]. Generally, catalytic performance of nanocrystals can be finely tuned not only by their composition, which mediates electronic structure, but also by their shape, which determines surface atomic arrangement and surface coordination [4]. Metal nanoparticles exhibit a wide variety of shapes, facets, fraction of surface atoms, which makes it critically important to study the effect of metal nanoparticles shape on the catalytic activity of various organic and inorganic reactions. However, little attention has been paid on the understanding of the cooperative effect of composition and shape of bimetallic nanocatalysts to their catalytic properties [5].We propose controllable nucleation of bimetallic Au, Pt and Pd nanoparticles with non seed-mediated and seed-mediated nucleation by using classical MD computational simulations. The size, composition and structure of these nanoparticles will be characterized by the simulation of their high-resolution transmission electron microscopy (HRTEM) images to compare with experimental evidence. In addition, we will investigate the features of specific atomic sites at the surface, which improves the catalytic activity properties of the particles, and further uncover the structure-activity relationship of bimetallic nanocatalysts. Finally, we will also explore the mechanisms under which the initial conditions lead the formation of different atomic arrangements such like alloys, core-shell, and phase-separated bimetallic nanostructures.[1] Z. M and F. Zaera, Heterogeneous Catalysis by Metals, “Encyclopedia of Inorganic Chemistry”, 2nd edition, ed. R. B. King, John Wiley & Sons, Chichester, 2005, p. 1768.[2] R. Narayanan and M. A. El-Sayed, Catalysis with Transition Metal Nanoparticles in Colloidal Solution: Nanoparticle Shape Dependence and Stability, J. Phys. Chem. B 2005, 109, 12663-12676.[3] M. A. El-Sayed, Some Interesting Properties of Metals Confined in Time and Nanometer Space of Different Shapes, Acc. Chem. Res. 2001, 34, 257-264. [4] C. J. Murphy, Spatial Control of Chemistry on the Inside and Outside of Inorganic Nanocrystals, ACS Nano 2009, 3, 770-774.[5] B. Corain, G. Schmid and N. Toshima, Metal Nanoclusters in Catalysis and Materials Science: The Issue of Size Control, first edition, Elsevier, 2008, p.49.**Project Duties/Tasks*** Creation of the input files for LAMMPS and learn how to send jobs to Bragg.
* Test different values of atomic deposition rates and temperatures to produce different atomic arrays of bimetallic nanoparticles.
* Try different compositions of bimetallic nanoparticles (Au, Pt and Pd).
* Characterization of the size, composition and structure by HRTEM simulated images and NC Code.
* Analysis of the surface defects and morphology to determine the catalytic efficiency.
* Perform simulations of the nucleation process with an initial seed.
* Comparison of the catalytic efficiency for nucleation without initial seed and with initial seed.
* Write report/paper.

**Relevant Fields of Study*** Computational Chemistry
* Materials Science
* Physics

**Location:** Parkville, Melbourne**Contact:** Dr. Hector Barron Escobar via phone on (03) 9662 7105 or email hector.barronescobar@data61.csiro.au |
| Data61 30 | **Project Title**Convolutional Neural Networks Models of Nanomaterials Performance.**Project Description**Big data analytics and machine learning techniques are having a tremendous impact across different research fields, from social sciences and marketing to climate change and medicine. Particularly in nanosciences, the combination of powerful computational chemistry methods and machine learning techniques can undoubtedly accelerate novel nanomaterials discovery. At the CSIRO Molecular and Materials Modelling laboratory, we are developing novel hybrid computational approaches that combine high-throughput (HT) simulations with data mining and machine learning techniques for the *in silico* characterization of large nanomaterials data sets.Our in-house computer python library for materials analytics transforms structural information data into a virtual representation suitable for the calculation of a large variety of structural descriptors, which are then analysed by statistics, machine learning and visualization modules. Recent we have added support for deep learning capabilities with the addition of convoluted neural networks (ConvNets) interface. However, we have found that the successful implementation of ConvNets to nanomaterials analysis is extremely challenging because of the high complexity of the network architecture and the large number of hyper-parameters that demands simultaneous adjustment.In order to improve the integration of ConvNets into nanomaterials modelling, we are exploring genetic algorithm (GA) routines for the optimization of the net architecture and hyper-parameters of ConvNets. The GA optimization simultaneously explores different regions of the topology and parameter spaces, as well as selecting the best representation of the nanomaterial features to be presented to the nets. In this way, the GA will yield optimum ConvNet predictors of nanomaterials performance by correlating sophisticated representations of the nanomaterials to its performance.In this project, the student will test the GA-optimization of ConvNet (GA- ConvNet) on 2D array image representations of silver nanoparticles derived from self-organizing maps (SOMs). Firstly, the image fingerprints will be constructed by training similarity SOM maps of the dataset and posterior systematic projection of each nanoparticles on the SOM map. Secondly, the data set of images will be assigned class labels according to electronic properties thresholds, (i.e. band gap > 1 eV); and GA-ConvNets will trained to identify the different class of nanomaterials. Finally, the student will evaluate the performance of different optimization configurations. This will be a great opportunity to interact with deep learning modelling environments, developing high-performing computing skills and analytic thinking to effectively report scientific findings.**Project Duties/Tasks*** Unsupervised and semi-supervised training of self-organizing maps (SOMs) of silver nanoparticles data set
* Generation of 2D-array image fingerprints by systematic projection of each sample on the generated SOMs
* Assignment of class labels to the each silver nanoparticles using different thresholds based on electronic properties.
* Preliminary evaluation of the existing deep learning tools like Nvidia Digits and Caffe software.
* Perform test run of the GA-CovNets code with different optimization configurations.
* Analysis of the GA-CovNets runs using different performance criteria, i.e. accuracy vs. precision, accuracy vs. recall, etc

**Relevant Fields of Study*** Information and Computer Sciences
* Machine learning and soft-computing
* Computational chemistry

**Location:** Parkville, Melbourne **Contact:** Dr. Michael Fernandez Llamosa via phone on (03) 9662 7151 or email michael.fernandezllamosa@data61.csiro.au |
| Data61 31 | **Project Title**Intermolecular Interactions with Polarisable Quantum Monte Carlo-Molecular Mechanics (QMC/MM) Method.**Project Description**Solvation and environment effects in biological systems are important, as they influence biological reactions and processes. Calculating these effects is very useful not only to understand and compare reaction mechanisms but also to design experiments. Quantum mechanics (QM) methods are popular in such calculations because of their accuracy. Quantum Monte Carlo (QMC) method is a very accurate QM method that uses random sampling of electronic space to calculate molecular energies.In calculations on biological or solute-solvent systems it is necessary to include a large number of molecules to make sure a reasonable model is considered. But the downside of this is the increase of the computational cost. To overcome this problem QM methods are coupled with simpler but efficient methods such as molecular mechanics (MM) methods.“CMQMC” is a computer program developed at CSIRO that can do QMC calculations using parallel supercomputers. Recently, we extended its capability to carry out polarisable QMC/MM (a hybrid computational method which allows us to handle systems with large number of atoms) calculations. At this point it is very important to test the new features of CMQMC on simple molecular systems first and compare with either experimental results or results obtained by other QM methods. Such comparisons will prove the validity of the implementation as well as point towards the direction of improvements either in accuracy or efficiency.In this project, first the student will construct a set of molecular clusters (~100 clusters) as the test cases and generate their XYZ coordinate files. These clusters will essentially consist of a solute molecule like H2O, CH3OH etc., which will be treated by QMC and one or more solvent molecules, which will be treated by a polarisable MM method. Then the student will use the new polarisable QMC/MM method to calculate the solute-solvent interactions. The student will also calculate these interactions with CCSD(T) and DFT methods and compare with the QMC/MM results.During the course of this project the student will learn to design and carry out quantum mechanics, quantum Monte Carlo, and QM/MM calculations. Also, this will be an ideal opportunity to become familiar with the supercomputer environments. The student will also learn to analyse data and effectively report findings.**Project Duties/Tasks*** Construct a test set of small molecular clusters (~100 systems) representing n-intermolecular molecular interactions. (n = 1, 2, 3, 4

….) (Example systems are: H2O-H2O, CH3OH-H2O etc..) These systems should be targeted for polarisable QMC-MM calculations with 1 QMC molecule and one or many MM molecules.* Collect and list the force filed parameters of the atoms of the MM region.
* Obtain the trial wave functions for these systems using GAMESS program and QuanPol package.
* Calculate the interaction between QMC and MM regions using the newly developed polarisable QMC-MM procedure in CMQMC program.
* Repeat the same calculations with CCSD(T) and B3LYP methods.
* Analyse and plot results to compare the accuracy of polarizable QMC- MM method with CCSD(T) and B3LYP methods.
* Write a report explaining the results and findings and present.

**Relevant Fields of Study*** Computational Chemistry
* Theoretical Chemistry
* Quantum Mechanics
* Theoretical Physics
* Physical Chemistry
* Chemical Physics
* Applied Mathematics

**Location:** Parkville, Melbourne**Contact:** Nandun Thellamurge via phone on (03) 9662 7308 or email nandun.thellamurege@data61.csiro.au |
| Data61 32 | **Project Title**Molecular geometry optimisation with Quantum Monte Carlo.**Project Description**The physical and chemical properties of a molecule are highly dependent on the geometrical arrangement of its atoms. Computational prediction of molecular structures from quantum mechanical principles is thus an important area of research towards *in silico* development of high- performance materials, improvements in the energy efficiency of chemical reactions, and the understanding of biological processes at the molecular level.Within the Molecular and Materials Modelling group, we have been working on developing a novel stochastic method for calculating molecular forces and performing geometry optimisations using Quantum Monte Carlo (QMC) techniques. QMC methods have been shown to produce reliably accurate energies for a range of molecules, with inherent parallelisability well suited to high performance supercomputing. Therefore, QMC methods are predicted to efficiently obtain more consistently accurate molecular geometries than the techniques currently available.Initial testing of our stochastic geometry optimisation (SGO) method has shown promise for a small set of systems. The focus of this vacation scholarship would be to apply our new method to a larger range of applications, and molecules with a greater number of geometry dimensions.This project will give the student a unique opportunity to gain experience in state of the art quantum chemistry methods, using large high-performance supercomputers. By carrying out an investigation into the optimal choice of parameters for use within the SGO algorithm, they will also obtain practical experience in data analysis and identifying trends. Finally, the student will have the chance to compare their results to experimental values and alternative theoretical methods, by extracting the relevant data from the literature.This project is designed to give the student experience completing a real- world computational research project, by working in a team that is actively developing these methods, and making a valuable contribution towards the development of our new SGO technique.**Project Duties/Tasks*** Gain a basic understanding of the principles behind QMC, the stochastic algorithms implemented for optimisation of molecular geometries, and the scripts used to perform these calculations through both independent background reading and one-on-one tuition.
* Apply the QMC geometry optimisation software to a selection of different molecules.
* Make changes to the parameters used in the stochastic algorithm to find the most reliable, consistent and accurate set of values.
* Compare our best theoretical results with experimental data and geometries obtained using alternative computational methods.
* Deliver a report/presentation on the project.

**Relevant Fields of Study*** Computational Chemistry
* Quantum Chemistry
* Theoretical Chemistry
* Physical Chemistry
* Chemical Physics
* Applied Mathematics

**Location:** Parkville, Melbourne**Contact:** Deidre Cleland via phone on (03) 9662 7275 or email deidre.cleland@data61.csiro.au |
| Data61 33 | **Project Title**Business Process Data Compliance Verification.**Project Description**Most approaches to business process compliance are restricted to the analysis of structure of processes. However, it has been argued that full regulatory compliance should also include the understanding of what the tasks have been doing in a business process. In the light of this, the objective of this project is to implement a business process data verification application based on the idea of control-tags (e.g., data, resources, etc.), which provide the metadata for compliance as input, to annotate a business process for design-time compliance checking.**Project Duties/Tasks*** Conception of an integration of business process model annotations with database schema;
* Implementation of a proof-of-concept integration;
* Test and assessment of the implemented integration.

Students can have an idea of the possible outcomes by reading “Business Process Data Compliance”, Mustafa Hashmi, Guido Governatori and Moe Thandar Wynn, RuleML 2012, pp. 32-46, 2012.**Relevant Fields of Study*** Computer Science
* Information Technology
* Information Systems

This project will involve working on business process and database modelling, norms, and coding in Java.To apply, please send a brief cover note with a copy of your CV and academic transcript to the email address stated below.**Location:** Spring Hill, Queensland**Contact:** Dr. Mustafa Hashmi via phone on (07) 3214 2761 or email mustafa.hashmi@data61.csiro.au or Dr. Brian Lam via phone on (07) 3214 2765 or email brian.lam@data61.csiro.au |
| Data61 34 | **Project Title**Business Process Management Workflows in Blockchain Systems.**Project Description**The integration of **business processes** across organizations is typically beneficial for all involved parties. However, the lack of trust is often a roadblock.**Blockchain** (the underlying architecture of the **Bitcoin** platform) is an emerging technology for decentralized and transactional data sharing across a network of untrusted participants. It can be used to find agreement about the shared state of collaborating parties without trusting a central authority or any particular participant.Some blockchain networks also provide a computational infrastructure to run autonomous programs called **smart contracts**. Smart contracts are small programs that can automatically apply contractual clauses. These programs are often proposed to promote security and impartiality in the execution of contractual clauses.The student(s) will investigate and implement a proof-of-concept solution where business process workflows are executed or monitored with or as smart contracts in a blockchain platform (e.g. Ethereum, [www.ethereum.org](http://www.ethereum.org)).**Project Duties/Tasks*** Conception of an integration of business process workflows with smart contracts and blockchain systems;
* Implementation of a proof-of-concept integration;
* Test and assessment of the implemented integration.

Students can have an idea of the possible outcomes by reading “Untrusted Business Process Monitoring and Execution Using Blockchain”, Ingo Weber, Xiwei Xu, Regis Riveret, Guido Governatori, Alexander Ponomarev, Jan Mendling, BPM 2016, or watching *https://*[*www.youtube.com/watch?v=1SNn9c5HHQs.*](http://www.youtube.com/watch?v=1SNn9c5HHQs)**Relevant Fields of Study*** Computer Sciences
* Information Technology

Skills required: Autonomy, business process modelling, and good programming skills.Application Requirement: Academic transcripts and C.V.**Location:** Spring Hill, Queensland**Contact:** Prof. Guido Governatori via phone on (07) 3214 2768 or email on guido.governatori@data61.csiro.au or Dr. Regis Riveret via phone on (07) 3214 2760 or email on regis.riveret@data61.csiro.au |
| Data61 35 | **Project Title**3D Web Tools.**Project Description**This project aims to develop new 3D Web Graphics tools for application across a wide range of sciences and industries. The successful applicant will use WebGL and related technologies to construct novel interactive prototypes.Web based Augmented Reality, and the use of hardware such as HoloLens is likely.**Project Duties/Tasks*** Learn about CSIRO’s existing Web3D prototypes and science areas
* Develop new WebGL based prototype tools
* Integrate tools within CSIRO’s Web3D system

**Relevant Fields of Study*** Computer Science
* Engineering

(Experience or interest in Computer Graphics is desirable)**Location:** Canberra**Contact:** Matt Adcock via phone on (02) 6216 7098 or email matt.adcock@csiro.au |
| Data61 36 | **Project Title**3D Interactive Techniques.**Project Description**This project will explore new ways of interacting with 3D data. The new techniques will be relevant to Abstract Data (such as immersive big data analytics) and acquired 3D data (such as using HoloLens to explore 3D scans).This project will also involve the combination of 3D Printing and Augmented Reality technologies.**Project Duties/Tasks*** Learn about CSIRO Data61’s existing 3D Interaction systems
* Develop new 3D Interaction prototypes
* Integrate CSIRO’s existing systems to create an illustrative demo

**Relevant Fields of Study*** Computer Science
* Engineering

(Experience or interest in Computer Graphics is desirable)**Location:** Canberra**Contact:** Matt Adcock via phone on (02) 6216 7098 or email matt.adcock@csiro.au |
| Data61 37 | **Project Title**Novel Urban Visualisation.**Project Description**CSIRO is investigating new tools for urban decision-making and new ways for stakeholders to engage with urban scenario models.This project will provide opportunities for the vacation scholar to explore Virtual Reality, Augmented Reality, 3D Web Graphics, Data Visualisation, Geospatial Systems and/or Interactive Simulation.**Project Duties/Tasks*** Learn about existing CSIRO simulation and visualisation technologies
* Review of existing solutions/standards
* Development of novel prototypes and demonstrators

**Relevant Fields of Study*** Computer Science
* Engineering

(Experience or interest in Computer Graphics is desirable)**Location:** Canberra**Contact:** Matt Adcock via phone on (02) 6216 7098 or email matt.adcock@csiro.au |
| Data61 38 | **Project Title**Deep learning based object detection and classification.**Project Description**With the rapid advancement of deep learning and high performance computing technologies, data scientists can now use GPUs, GPU clusters or purpose-built hardware systems for deep learning to develop and deploy machine learning based applications such as massive image classification and video analytics.The aim of this project is to develop a deep learning based image classifier using the latest high performance platform for visual computing. The work will include assessing the existing deep learning software packages, deploy a selected package using the platform, and implement the image classification application using the deep learning package. The practical outcome from this work will be increased capabilities for automated fishery monitoring to support sustainable fisheries.The project has close links to computer science, mathematics and engineering, and can help bridge the gap between your courses and developing a real scientific application. The suitable candidate will have opportunities to have access to world class facilities and work alongside CSIRO senior scientists while you are enjoying generous personal development and learning opportunities.**Project Duties/Tasks*** Software Engineering (C, C++)
* Implementation of deep learning applications
* Software testing and documentation

**Relevant Fields of Study*** Computer Science, Software Engineering, Information Technology
* Engineering
* Mathematics and Statistics

**Location:** Marsfield, Sydney**Contact:** Dr Dadong Wang via phone on (02) 9325 3223 or email dadong.wang@csiro.au |
| Data61 39 | **Project Title**Blockchain-based B2B Collaboration.**Project Description**With the increasing business to business (B2B) collaborations, there is a strong demand for de-centralized trusted platform to support the core services required for the B2B collaborations, such service discovery, contract negotiation, contract monitoring and enforcement. This project is to evaluate the feasibility and performance of bockchain-based collaborations, esp. its smart contract technology. The expected outcomes from this project would be:1. A testing suit of a blockchain performance benchmark
2. A technical report on the design and testing results
3. The student is expected to learn blockchain technologies and software testing skills via this project.

**Project Duties/Tasks*** Design and implement a simple smart contract for blockchain-based B2B collaborations.
* Design a set of evaluation criteria for performance testing of key smart contract operations with the blockchain.
* Conduct the performance testing using the prototype and evaluation criteria.
* Write a technical report to collect and analysis the testing results.

**Relevant Fields of Study*** Software Engineering
* Information Technology

**Location:** Marsfield, Sydney **Contact:** Dr. Shiping Chen via phone on (02) 9372 4663 or email shiping.chen@csiro.au |
| Data61 40 | **Project Title**Decentralizing big data processing.**Project Description**Traditional Hadoop data processing framework uses HDFS (Hadoop Distributed File System) to store very large data sets reliably in a large cluster, and to stream those data sets at high bandwidth to user applications.Decentralizing the data processing framework can support applications at the scale of the Web. Our first step of decentralizing big data processing framework is to replace the HDFS with IPFS (InterPlanetary File System), which is a peer-to-peer distributed file system that seeks to connect all computing devices with the same system of files. IPFS is a new hypermedia distribution protocol, addressed by content and identities. IPFS enables the creation of completely decentralized applications.**Project Duties/Tasks*** Task 1: Explore decentralized data stores and decentralized applications in general
* Task 2: Replace HDFS of Hadoop data processing frameworks with one decentralized data store (IPFS could be a candidate)
* Task 3: Evaluate performance

**Relevant Fields of Study**Software Engineering, Computer science**Location:** Eveleigh, Sydney**Contact:** Xiwei Xu via email Xiwei.Xu@data61.csiro.au |
| Data61 41 | **Project Title**Performance Analysis of Blockchain-based Systems.**Project Description**Blockchain [1], the underlying and separable technology behind Bitcoin, is increasingly being used as a public or community ledger (shared database) with programmable contracts running on Blockchain [2] and emergent consensus for many interesting uses. Examples include posting jobs and demands for other parties to bid and then verify the completion of the job, completing business transactions and registering digital assets (such as data or art) for copyrights and per-use micropayment. It can potentially disrupt a number of industries. Using blockchain in different scenarios requires comparison of blockchains with different configurations. However, in practice, the lack of reliable technology evaluation source makes this comparison a hard task. This summer project is aimed to compare and analyse blockchain- based systems in terms of performance through simulation or queuing models [4].[1][https://en.wikipedia.org/wiki/Block\_chain\_(database)](https://en.wikipedia.org/wiki/Block_chain_%28database%29) [2] <https://en.bitcoin.it/wiki/Contracts> [3] <https://www.ethereum.org/> [4] <https://en.wikipedia.org/wiki/Queueing_theory> **Project Duties/Tasks*** Task 1: Propose a performance model of blockchain-based systems
* Task 2: Analyse some existing proof-of-concepts using the proposed model

**Relevant Fields of Study**Software Engineering, Computer science**Location:** Eveleigh, Sydney**Contact:** Xiwei Xu via email Xiwei.Xu@data61.csiro.au or Mark Staples via email Mark.Staples@data61.csiro.au  |
| Data61 42 | **Project Title**Reputation Mechanism on Blockchain-based Decentralised Systems.**Project Description**Blockchain [1] is an emerging technology for decentralized and transactional data sharing across a large network of untrusted participants. It enables new forms of decentralised (peer-to-peer) systems, where components can find agreements on their shared states without trusting a central integration point or any particular participating components. Blockchain is increasingly being used as a public or community ledger (shared database) with programmable contracts running on blockchain [2] and emergent consensus for many interesting uses, for example, prediction market, trading market, or sharing data. Such a decentralised systems relies on reputation to represent the trustworthyness of a peer, which measures how much the community trusts the peer [4] [5]. This summer project is aimed to explore existing reputation mechanisms on peer-to-peer systems and proposes a generic reputation mechanism for blockchain-based decentralised systems.[1][https://en.wikipedia.org/wiki/Block\_chain\_(database)](https://en.wikipedia.org/wiki/Block_chain_%28database%29) [2] <https://en.bitcoin.it/wiki/Contracts> [3] https://[www.ethereum.org](http://www.ethereum.org) [4] Surya Nepal, Zaki Malik and Athman Bouguettaya, “Reputation Propagation in Composite Services”, ICWS2009**Project Duties/Tasks*** Task 1: Propose a generalized reputation mechanism that can be used on any blockchain-based decentralised system
* Task 2: Implement of the proposed reputation mechanism on one of our proof-of-concepts/prototypes

**Relevant Fields of Study**Software Engineering, Computer science**Location:** Eveleigh, Sydney**Contact:** Xiwei Xu via email Xiwei.Xu@data61.csiro.au or Surya Nepal via email Surya.Nepal@csiro.au |
| Data61 43 | **Project Title**Bitcoin/Blockchain-driven Systems.**Project Description**Blockchain [1], the underlying and separable technology behind Bitcoin, is increasingly being used as a public or community ledger (shared database) with programmable contracts [2] and emergent consensus for many interesting uses. Examples include posting jobs and demands for other parties to bid and then verify the completion of the job, completing business transactions and registering digital assets (such as data or art) for copyrights and per-use micropayment . It can potentially disrupt a number of industries. The research group you will be joining have already build systems that run blockchains (specifically, Etherum) for a number of scenarios including smart contracts. This project will investigate more ways of using blockchain for different purposes and evaluate them in terms of architecture designs, scalability and security.[1] [https://en.wikipedia.org/wiki/Block\_chain\_(database)](https://en.wikipedia.org/wiki/Block_chain_%28database%29) [2] <https://en.bitcoin.it/wiki/Contracts> [3] https://[www.ethereum.org](http://www.ethereum.org) **Project Duties/Tasks*** Task 1: Learn different blockchain products
* Task 2: Build proof-of-concept on blockchain for different purposes
* Task 3: Evaluate the implemtned systems

**Relevant Fields of Study**Software Engineering, Computer science**Location:** Eveleigh, Sydney**Contact:** Xiwei Xu via email Xiwei.Xu@data61.csiro.au or Liming Zhu via email Liming.Zhu@data61.csiro.au  |
| Data61 44 | **Project Title**Big Data Provenance.**Project Description**Data provenance is metadata which keeps track of where a piece of data comes from, and what operations have been done to it. This metadata provides a way to assess authenticity, enable trust in the data itself and the analysis results and allows reproducibility. This is particularly important when the data end-users, data analysts, data producers, and data wranglers are different groups of people across exploration and production stages.However, existing data provenance features are often embedded in vendor- specific tools or have strong assumptions of the types of analytics systems or operations and their tracking. There are also no good mechanisms of publishing/sharing data with usable and verifiable provenance attached. Big data is exacerbating the problem in terms of mechanism efficiency for provenance capturing and sharing. This project looks into how we can keep track of both data and programs by automatically associating version controlled programs to version controlled datasets, independent of systems used. The published provenance should be easily usable for a reproduction of the current dataset. This provenance information can also be used for many other purposes such as undo/redo of steps and reproducing steps from environment/system A to B (e.g. R on laptop to Spark in cluster; small data to big data; exploration to production).**Project Duties/Tasks*** Task 1: Build a data provenance tool to track version controlled programs and version controlled datasets
* Task 2: Evaluate the proposed solution through comparing with state-of- practise

**Relevant Fields of Study**Software Engineering, Computer science**Location:** Eveleigh, Sydney**Contact:** Xiwei Xu via email Xiwei.Xu@data61.csiro.au or Liming Zhu via email Liming.Zhu@data61.csiro.au  |
| Data61 45 | **Project Title**Dependable Auditing on Operations of in-Cloud Applications.**Project Description**Despite the tremendous potential of cloud computing, organisations that deploy their applications in the cloud may have concerns related to loss of control and dependability issues (e.g. security, privacy and availability).Operations of in-cloud applications are usually a mix of (semi-) automated tasks by both in-house administrators and cloud infrastructure providers reacting to rapidly changing environment. It is often difficult to establish assurance cases and track accountabilities of these operations in three cases1. for compliance testing with respect to legal requirements such as Basel III
2. forensic operations when there has been a security breach, and
3. when the applications encounter problems and it is necessary to get compensation from the cloud vendor.

This project will investigate the both the technical mechanisms and legal issues in auditing operations of in-cloud applications. Specifically, the student will examine the current logging facilities related to typical operations of in-cloud applications and cloud infrastructure vendor’s comliance process and evidence requirements. The student will then propose improvements in logging and auditing so better assurances and accountability could be established.**Project Duties/Tasks*** Task 1: Examine the current logging facilities related to typical operations of in-cloud applications and cloud infrastructure vendor’s compliance process and evidence requirement
* Task 2: Propose improvements in logging and auditing to achieve better assurances and accountability

**Relevant Fields of Study**Software Engineering, Computer science**Location:** Eveleigh, Sydney**Contact:** Liming Zhu via email Liming.Zhu@data61.csiro.au  |
| Data61 46 | **Project Title**Continuous Deployment for Big Data Analytics Applications.**Project Description**Data scientists are increasingly moving from small-scale data analytics on a laptop to big data analytics in clusters (e.g. for genome analysis and financial data analysis). However, data scientists still need to perform explorative analytics development on their laptop or a small-scale environment and then deploy the data analytics application to clusters, often with extensive support from data product teams and engineering teams. If any issue arises during on the large-scale deployment or operation, data scientists need to revise their models back at their laptop and repeat the process again. Continuous Deployment/Delivery (CD) is a practice that copes with high frequency and automated deployment of applications. CD practices have been used in many types of applications but its use for data analytics applications and model development is still limited due to distinct model development cycle, data sampling and cluster deployment challenges. This project will expose you to data science and big data workflow. You will work within a team to develop new solutions to automate and simplify the workflow.**Project Duties/Tasks*** Task 1: Explore the current practise of data science and big data workflow
* Task 2: Develop new solutions to automate the simplify the workflow
* Task 3: Evaluate the proposed solution

**Relevant Fields of Study**Software Engineering, Computer science**Location:** Eveleigh, Sydney**Contact:** Liming Zhu via email Liming.Zhu@data61.csiro.au  |
| Data61 47 | **Project Title**Dependable Blockchain Crowd-Funding Application.**Project Description**Advances in the context of Bitcoin-like distributed platforms allow peers to exchange digital asset ownerships through more elaborate transactions or smart contracts. Although traditional blockchain systems (e.g., Ethereum) may experience anomalies [1], few consistent alternatives were proposed recently. This project consists of designing a dependable crowd-funding application on a new consistent blockchain system deployed as a Data61 private chain. The application will allow peers to open a new account for a project and place project funding orders. Smart contacts will have to be implemented to allow a project to be funded once a threshold is reached and to provide return on investment. This application could serve as a building block for a Decentralized Autonomous Organisation (DAO) [3].[1] The Blockchain Anomaly. Christopher Natoli and Vincent Gramoli. arXiv:1605.05438v1, 2016.[2] The Blockchain as a Software Connector. Xiwei Xu, Cesare Pautasso, Liming Zhu, Vincent Gramoli, Alexander Ponomarev, An Binh Tran and Shiping Chen. Proceedings of the 13th Working IEEE/IFIP Conference on Software Architecture (WICSA), 2016.[3]<http://techcrunch.com/2016/05/16/the-tao-of-the-dao-or-how-the-autonomous-corporation-is-already-here/> **Project Duties/Tasks*** Task 1: Demo
* Task 2: performance evaluation under real workloads and results under stress-tests of pathological scenarios

**Relevant Fields of Study**Software Engineering, Computer science**Location:** Eveleigh, Sydney**Contact:** Ingo Weber via email Ingo.Weber@data61.csiro.au  |
| Data61 48 | **Project Title**Exploring the risks of software monoculture.**Project Description**Many software products today have external dependencies, which are often retrieved from online repositories such as Github, Bitbucket, or npn.Languages like Go and JavaScript even incorporate the automatic inclusion of libraries from these services.However, this growing dominance has given rise to a problem in software engineering that has implications for security: what happens if the external package is critically changed or removed? How many systems would be affected, and how?In this project, we are going to carry out an empirical analysis to give a first answer. We will analyse dependencies of existing projects and analyse their properties.**Project Duties/Tasks*** Task 1: A tool set to carry out automatic analyses of software dependencies in the networked world, focusing on one language (most likely Go).
* Task 2: A short report describing the chosen methodology and results

**Relevant Fields of Study**Software Engineering, Computer science**Location:** Eveleigh, Sydney**Contact:** **Contact:** Ralph Holz via email ralph.holz@data61.csiro.au or Ingo Weber via email Ingo.Weber@data61.csiro.au  |
| Data61 49 | **Project Title**Block chain on Cloud.**Project Description**This project is going to build a private block chain system on a real cloud platform, on which we will test a spectrum of distributed applications for various attacks.**Project Duties/Tasks*** Task 1: Using Ethereum to build block chain system in virtual machine images
* Task 2: Using Puppet and the images to spawn arbitrary BC system in Nectar cloud
* Task 3: Testing and reporting

**Relevant Fields of Study*** Software engineering
* Computer systems

**Location:** Canberra city, ACT **Contact:** Daniel Sun via phone on (02) 6218 3810 or email daniel.sun@data61.csiro.au |
| Data61 50 | **Project Title**Self-adaptive IoT.**Project Description**The project aims to conduct research and development on IoT security. Resource constrained IoT devices are vulnerable to attacks. It is desired that an IoT network can be self-adaptive or self-healing against various problems. Due to the restrictions of lightweight devices, existing solutions are unaffordable. This project is to develop a lightweight self-adaptive mechanism for recovering an IoT network from attacks and unexpected status.**Project Duties/Tasks*** Conduct a survey on attacks against IoT networks.
* Understand lightweight mutual authentication protocols and key update techniques.
* Design and implement a self-adaptive solution to different problems.

**Relevant Fields of Study*** Computer Science

**Location:** Marsfield, Sydney**Contact:** Nan Li via phone on (02) 9372 4193 or email nan.li@csiro.au |
| Data61 51 | **Project Title**Home IoT and Security.**Project Description**This project is going to build software infrastructure for Home IoT systems and protection for data and accessibility. A student should better have good knowledge of operating system and programming on it, especially hardware drivers, and also good at network programming.**Project Duties/Tasks*** Task 1: On a set of IoT hardware and controllers, a student will program data collection, information exchange, and control logic on and between IoT devices and its host, a cloud.
* Task 2: Attack and defence practice on the system built in Task 1
* Task 3: Testing and reporting

**Relevant Fields of Study*** Software engineering
* Computer systems

**Location:** Canberra city, ACT **Contact:** Daniel Sun via phone on (02) 6218 3810 or email daniel.sun@data61.csiro.au |
| Data61 52 | **Project Title**Rule-Based Reporting System (RuleRS).**Project Description**Data61, CSIRO designed a Rule-based Reporting Systems (RuleRS), which integrates databases, with logic reasoner and rule engine [1].Students can have an idea by reading [1] Islam MB, Governatori G (2015) Ruleoms: A rule-based online management system. In: Proceedings of the ICAIL 2015, ACM, New York, NY, USA, pp 187–191**Project Duties/Tasks*** Task 1: to develop user friendly Graphical User Interface (GUI) for RuleRS
* Task 2: to develop user-friendly IOS/Android app for RuleRS.
* Task 3: to test database compatibility with MySQL/ SQLite database (currently the project is implemented using PostgreSQL database)

**Relevant Fields of Study*** Science
* Engineering
* Information Technology

**Location:** Spring Hill, QLD**Contact:** Dr Badiul Islam via Phone on (07) 3214 2762 or email badiul.islam@data61.csiro.au |
| Data61 53 | **Project Title**Pig+.**Project Description**This project is aiming at revising Pig source code to provide full and fine grain provenance functionality. Pig is a open source software in Hadoop eco-system and a critical workflow management tool for big data analytics.**Project Duties/Tasks*** Task 1: Reading and decomposing Pig source code
* Task 2: Revising the code with respect to existing provenance system
* Task 3: Testing and reporting

**Relevant Fields of Study*** Software engineering
* Computer systems

**Location:** Canberra city, ACT **Contact:** Daniel Sun via phone on (02) 6218 3810 or email daniel.sun@data61.csiro.au |
| Data61 54 | **Project Title**Cryptographically protected Cloud Data.**Project Description**Cryptographic protocols are always considered to be a solution where the confidentiality and integrity are required such as a cloud network. In this project, a program which encrypts (decrypts) files when those files are uploaded to (downloaded from, respectively) a third party cloud storage such as Google Drive and Drop-box will be implemented. A knowledge of cryptographic algorithms is not necessary but will be helpful.**Project Duties/Tasks*** Understanding cryptographic algorithms such as attribute based encryption and signature schemes
* Implementing cryptographic algorithms using open source libraries
* Applying a cryptographic protection to the data in a cloud storage

**Relevant Fields of Study*** Computer Science
* Network/Data Security

**Location:** Marsfield, Sydney**Contact:** Jongkil Kim via phone on (02) 93724223 or email Jongkil.kim@data61.csiro.au  |
| Data61 55 | **Project Title**Fighting Ransomware on Mobile Devices with Document Randomization and Encryption.**Project Description**Mobile devices (e.g., smart phones) are being integrated into our daily life and work. We can store various valuable documents on mobile devices, such as personal data and photos. However, mobile devices are increasingly becoming attractive attack targets in recent years. In particular, ransomware starts to attack mobile devices. The victims of ransomware may lose their valuable documents if they do not want to pay attackers for decryption keys. In this project, we will randomize the storage of documents on smart devices and encrypt them, such that even if a smart phone is compromised, the ransomware will not find recognizable or really valuable documents to encrypt as in ordinary smartphone storage system.**Project Duties/Tasks*** Document encryption on the Android environment
* Efficient protocol to access encrypted documents
* Implementation of a demonstrable system on smart phones

**Relevant Fields of Study*** Information security
* Computer science

**Location:** Marsfield, Sydney**Contact:** Dongxi Liu via phone on (02) 93724152 or email dongxi.liu@csiro.au |
| Data61 56 | **Project Title**Real forecasting with consideration of data uncertainty.**Project Description**Forecasting is becoming more and more important in many research fields. While more and more data are available due to technical innovation, data uncertainty is also an increasing issues for both model calibration/estimation and forecasting. This project aims to excise some techniques for real forecasting when the data are subject to uncertainty, with possibility to produce a add-on software module/package for general use.**Project Duties/Tasks*** Task 1: Literature review of different techniques and their underlying assumptions.
* Task 2: Programming in R to develop statistical model for model calibration/estimation.
* Task 3: Writing up with a case study (data and model with be ready to use).

**Relevant Fields of Study*** Statistics
* Analytics
* Mathematics
* Computing
* Other disciplines with strong analytic skills

**Location:** Floreat, WA**Contact:** Quanxi Shao via phone on (08) 9333 6613 or email Quanxi.Shao@csiro.au |
| Data61 57 | **Project Title**Workload Analysis Toolkit.**Project Description**The need to understand operator workload is a key requirement across numerous sectors, including maritime shipping, nuclear power operations, air traffic control, driving, and many other contexts that impose a high demand on the human attentional system. While the physical elements of workload are generally well understood, the concept of cognitive workload, or mental workload, is less self-evident. Human attention is by nature a limited resource and under certain conditions our comprehension of a situation can break down, with the result being that accidents happen, causing damage to property, human life, and to the environment.This project aims at developing a toolkit to analyse and measure mental workload unobtrusively in real-time from. For this we are assessing and validating of a number of workload measurement techniques including psychophysiological measures (measuring changes in people’s physical reactions). The vacation scholarship will focus on the use of Galvanic Skin Response (GSR) or skin conductance levels for workload assessment.**Project Duties/Tasks*** Define research scope with supervisors and background research
* Data collection and preparation
* Develop and/or adapt existing signal processing tools
* Develop and/or adapt machine learning algorithms
* Design and conduct small scale validation study with human participants

**Relevant Fields of Study**Electrical Engineering, etc* Psychology,
* Computer science or related discipline with interest and skills in one or more of the following areas: Human-Computer Interaction, Signal processing, Machine learning, Designing, conducting and analysing studies with human participants.

**Location:** Hobart, Tasmania**Contact:** Andreas Duenser via phone on (03) 62375678 or email andreas.duenser@csiro.au  |
| Data61 58 | **Project Title**Skeletal motion capture using Microsoft Kinect™ for sports and rehabilitation simulations.**Project Description**CSIRO has well developed human body simulation capabilities that are used for developing detailed models of Australia’s elite sports people. We need to be able to intuitively modify the technique used by the model athletes and the Microsoft Kinect™ hardware is a suitable, cheap and widely available hardware for capturing body motions. By developing an intuitive interface between the user and the models, coaches and athletes will be able to interact with the models more efficiently and also extend the use of the capability to the wider community such as for rehabilitation and workplace training.**Project Duties/Tasks*** Familiarisation with the Microsoft Kinect API (C++, C# bindings) and CSIRO’s Workspace platform (C++ development)
* Build C++ operations to capture, visualise and analyse data from the Microsoft Kinect™ and provide skeletal motion capture data in a suitable form for existing biomechanical models
* Capture body motions using the Microsoft Kinect™ and input these motions into existing sports models
* Optimise the use of the new capability for non-technical users

**Relevant Fields of Study*** Computer programming
* Engineering

**Location:** Clayton, VIC**Contact:** Dr. Simon Harrison via phone on (03) 9545 8450 or email Simon.Harrison@csiro.au |
| Data61 59 | **Project Title**Hedging FX risk LSV-style!**Project Description**In the foreign exchange market, accurate pricing models are essential to the creation and operation of hedging portfolios. This project will provide the first analysis of both the gamma and vega hedging performances of a new pricing model termed the lognormal local-stochastic volatility model.**Project Duties/Tasks*** Task 1: Setup gamma and vega hedging portfolios using existing testing interface
* Task 2: Compute hedging performance over a range of market inputs and compare with previously computed delta-hedged LSV results
* Task 3: Document hedging performance for possible publication

**Relevant Fields of Study*** applied mathematics
* probability theory
* computer engineering

**Location:** Clayton, VIC**Contact:** Geoff Lee via phone on (03) 9545 8040 or email Geoffrey.Lee@data61.csiro.au |
| Data61 60 | **Project Title**Virtual and Augmented Reality Visual Analytics for Computational Modelling and Simulation.**Project Description**This project seeks to further develop a Virtual and Augmented Reality visualisation platform in [Workspace](https://research.csiro.au/workspace/) with a focus on applications in computational modelling and simulation.The successful applicant will work with world leading research and computational models from CSIRO scientists in areas including bushfire behaviour and aquatic athlete performance. Working as a member of the software engineering team, you will develop immersive methods for exploration and communication of this and other research to stakeholders and the public. This capability is being developed as part of Data61’s workflow framework [Workspace](https://research.csiro.au/workspace/) utilising the [Unity](https://unity3d.com/) game engine. The framework is used in projects across CSIRO and externally meaning the capabilities developed by the project are likely to be used in a diverse range of international research activities.**Project Duties/Tasks*** Develop an asset and data source pipeline from CSIRO’s [Workspace](http://research.csiro.au/workspace/) framework to the [Unity](https://unity3d.com/) game engine for AR and VR visualisation and visual analytics use cases. Development aspects will include
* C# development of [Unity](https://unity3d.com/) modules
* C++ development of [Workspace](http://research.csiro.au/workspace/) plugins for data acquisition and pre-processing
* Develop Cg/HLSL shaders for visualisation of simulation data

**Relevant Fields of Study*** Computer Science
* Software Engineering
* Information Technology (Games and Graphics Programming)

**Location:** Clayton, VIC**Contact:** Matt Bolger via phone on (03) 9545 7892 or email matt.bolger@data61.csiro.au Lachlan Hetherton via phone on (03) 9545 8041 or email lachlan.hetherton@data61.csiro.au or email workspace@csiro.au |
| Data61 61 | **Project Title**UX Design for Graphical Workflow Software.**Project Description**This project seeks to design and develop a new graphical user interface for CSIRO’s workflow software: [Workspace](https://research.csiro.au/workspace/). The successful applicant will work with the Workspace team to design new graphical enhancements, streamline common functionality and automate repetitive processes. The applicant will use JavaScript frameworks and agile development methodology to prototype their work. The end goal is to bring Workspace’s user interface to professional level of quality on par with other tech industry software products.The Workspace framework is used in projects across CSIRO and externally. The student’s work will have significant impact by enhancing the product for existing users and by increasing uptake.**Project Duties/Tasks*** Develop skills with JavaScript frameworks
* Iteratively design and develop new features for UX testing by the Workspace team

**Relevant Fields of Study*** Computer Science
* Software Engineering
* Information Technology (Games and Graphics Programming)

**Location:** Clayton, VIC**Contact:** Chris Rucinski via phone on (03) 9545 7895 or email chris.rucinski@data61.csiro.au, Lachlan Hetherton via phone on (03) 9545 8041 or email lachlan.hetherton@data61.csiro.au or email workspace@csiro.au |
| Data61 62 | **Project Title**OurClimate.**Project Description**OurClimate is an urban sustainability related project that aims at using new image processing algorithms and calculating a “sustainability score” for a dwelling. The sustainability score is an indication of how conducive the dwelling is for installation of solar PV and/or rainwater tanks.The image processing algorithm needs to be fine-tuned and tested. Ground truth has to be calculated against existing highly accurate analysis output for LiDAR imagery.A rooftop edge detection algorithm has to be identified and implemented.A database of street addresses needs to be created using the output of an image analysis algorithm.A working prototype of OurClimate called MyClimate can be found at: <http://thermalweb.it.csiro.au/arcgis/myclimate/index.html>Note that this currently works only for the City of Port Phillip in Victoria.**Project Duties/Tasks*** Use Workspace - <https://research.csiro.au/workspace/> - to create a workflow that integrates the image analysis algorithm with edge detection and creation of a database of street addresses. A
* Assign a sustainability score to each dwelling.
* Carry out test tasks.

Note that the student will not be expected to complete all tasks above but instead will be asked to focus on one or two tasks based on their level of expertise and interest.**Relevant Fields of Study*** Computer Science
* Software Engineering
* Information Technology
* GIS

**Location:** Clayton, VIC**Contact:** Dr. Mahesh Prakash via email Mahesh.Prakash@data61.csiro.au  |
| Data61 63 | **Project Title**Evaluation of a Microstructure Model of Titanium for Additive Manufacturing.**Project Description**The titanium alloy Ti-6Al-4V is widely used because of its high strength, toughness and corrosion resistance. Its mechanical properties are very sensitive to the microstructure so the ability to control the microstructure evolution during a titanium fabrication process is vital in order to obtain a high quality final product. CSIRO has recently implemented a microstructure evolution model within the software package MSC Marc for Ti-6Al-4V which requires testing, validation and senstivity studies. Specifically we are interested in evaluating the model for multi-pass welding and metal deposition processes. These processes involve severe temperature variations and therefore changes to the microstructure. Results from the project will be valuable in assessing whether a more detailed microstructure model is required for predicting microstructure evolution in multi-pass metal deposition processes. **Project Duties/Tasks*** Validation and sensitivity study of the material parameters in the model for a single cooling cycle (beta to alpha transition) at various cooling rates.
* Validation and sensitivity study of the model for a laser metal deposition experiment published in the literature
* Application of the model to simulate a multi-pass metal deposition process recently performed experimentally in Manufacturing. Evaluation of the model against the experimental results

**Relevant Fields of Study*** Materials Engineering
* Mechanical Engineering
* Applied Mathematics
* Experience with Finite Element Method is desirable

**Location:** Clayton, VIC**Contact:** Dr. Sharen Cummins via phone on (03) 9454 8061 or email sharen.cummins@csiro.au |
| Data61 64 | **Project Title**Fair Allocation of Chores.**Project Description**In multi-agent allocation and fair division, the goal is to allocate items to agents in a fair and efficient manner based on the preferences of the agents. Immediate applications include task scheduling, online barter markets, and allocation of CPU and memory resources in cloud computing. In most of the research in this area, the items to be allocated are "goods" for which agents have positive utility. There has been relatively less research on settings when the items to be allocated are chores or tasks for which agents have negative utility. The goal of this project is to devise new algorithms and protocols for fair allocations of chores among agents. The research project draws on principles and ideas from algorithm design as well as game theory/social choice theory.**Project Duties/Tasks*** Literature survey of existing approaches
* Algorithm design and analysis
* Possible implementation of the algorithm and experiments

**Relevant Fields of Study*** Algorithm design
* Game theory
* Artificial intelligence
* Software Engineering (in case the student is more interested in implementing algorithms)

**Location:** Kensington, NSW**Contact:** Haris Aziz via phone on (02) 8306 5909 or email haris.aziz@data61.csiro.au  |
| Data61 65 | **Project Title**Text Mining to Assist Physicians in Patient Diagnosis and Treatment.**Project Description**Project Outline:Text mining, also known as text analytics, can be used to find and extract information used for decision making. One application is clinical decision making, where doctors may need to search for information relevant to a specific patient, in order to make a decision for diagnosis or treatment. Technologies that can help this process are search, information retrieval and information extraction. In this project, we take advantage of an international competition for clinical decision support. We investigate techniques to search over a large body of medical literature, and find information that can best help doctors with specific queries. More information about this competition can be found here: <http://www.trec-cds.org/>This project will include activities such as: (1) processing big data on high performance computing systems; (2) using search and information retrieval techniques; (3) setting up a search engine; (4) pre-processing text; (5) using machine learning techniques.Requirements:- Interest in learning natural language processing techniques- Familiarity with Linux - Knowing a scripting language such as Python in a plus**Project Duties/Tasks*** Software engineering; Natural Language Processing and Machine Learning Research tasks

**Relevant Fields of Study*** Computer science
* Natural Language Processing
* Machine Learning

**Location:** Marsfield, Sydney**Contact:** Sarvnaz Karimi via phone on (02) 9372 4353 or email Sarvnaz.Karimi@csiro.au |
| Data61 66 | **Project Title**Natural Language Queries to Structured Data.**Project Description**Project Outline:Natural language queries provide an intuitive way to search within structured data, such as a spreadsheets or databases, to find the portion of the data set that corresponds to a user’s interests. However, mapping from the natural language query to data representations can pose a challenge for most current systems. In this project, you will perform research and develop technology for mapping natural language queries to structured data. Activities may include: preparing data sets (including using crowdsourcing methods), using natural language processing methods (including machine learning methods), and producing web demonstrators.Requirements:* Excellent communication skills
* Good programming skills
* Experience with Java and Python
* Familiarity with natural language processing techniques

Familiarity with machine learning concepts**Project Duties/Tasks*** Software engineering; Natural Language Processing and Machine Learning Research tasks

**Relevant Fields of Study*** Computer science
* Natural Language Processing
* Machine Learning

**Location:** Marsfield, Sydney**Contact:** Stephen Wan via phone on (02) 9372 4703 or email Stephen.Wan@csiro.au  |
| Data61 67 | **Project Title**Augmented Human-Bee Interaction.**Project Description**In the Global Initiative for Honey bee Health collaboration, an abundant amount of data from sensor ‘backpacks’ carried by honey bees is being collected to study their behaviour and monitor their health. Novel interfaces are needed to further optimise and accelerate the path from data to insight for field workers that are directly interacting with the honey bees, such as bee keepers, entomologists, and scientists. In this project, the student will develop and evaluate a visual analytics interface for an Augmented Reality headset that will allow to gain faster and deeper insight into bee behaviour in the field.**Project Duties/Tasks*** Develop a visual analytics interface for an Epson BT-200 augmented reality headset.
* Perform a formal evaluation of the interface in relation to an existing tablet-based interface with users in the field.
* Disseminate the outcomes of the development and study in a conference paper, journal article, and/or CSIRO report.

**Relevant Fields of Study*** Computer Science
* Human Computer Interaction
* Immersive Systems (AR/VR)

**Location:** Sandy Bay, TAS**Contact:** Huyen Nguyen via phone on 03 6237 5694 or email huyen.nguyen@csiro.au  |
| Data61 68 | **Project Title**Mobile IDE for OpenIoT platform.**Project Description**The existing IDE for OpenIoT middleware platform request definition and presentation is web-based. This limits deployment of user-driven workflow development interfaces. Porting of OpenIoT IDE to Android-based mobile platform will widen the scope of OpenIoT applications and make the platform more appealing. The student will study zero-programming principles of OpenIoT IDE, design adaptive IDE for Android-based platforms, ingest the data from a smartphone into OpenIoT, fetch relevant data streams and visualise the results of analytics on the smartphone.**Project Duties/Tasks*** **Task 1**. Study zero-programming drag&drop principles of OpenIoT IDE
* **Task 2**. Design and implement adaptive OpenIoT IDE for Android-based platforms
* **Task 3.** Develop and implement algorithms for ingesting the data from a smartphone into OpenIoT, fetch relevant data streams and visualise the results of analytics on the smartphone.

**Relevant Fields of Study*** Computer Science & Engineering
* Internet of Things platforms and middleware
* Android programming, sensors, data streams

**Location:** Clayton, VIC**Contact:** Arkady Zaslavsky via phone on (03) 9545 8016 or email arkady.zaslavsky@csiro.au |
| Data61 69 | **Project Title**Applying Machine Learning to the Design of New High Performance Granular Materials.**Project Description**Machine Learning allows a system to learn from inputted sets of data to predict outcomes and dynamically adjust when presented with new scenarios or inputs. This capability can offer unprecedented potential to accelerate our discovery of new materials and assessment of the suitability of large numbers of known materials for use in various industrial applications such as Additive Manufacturing (3D printing). In this project we will develop a machine learning model capable of modifying the internal structure of a simulated material composed of granular particles of various shapes with the aim of reducing the density of the system below that normally observed while maintaining mechanical stability. The behaviour of the model will be analysed to gain insights into the key structural features of the particle contact network that generates mechanically stability across a range of material densities. The model will be used to predict combinations of particle shapes that are capable of generating the lowest density mechanically stable systems.**Project Duties/Tasks*** Train a Machine Learning model using existing data from simulated granular materials composed of particles with different shapes.
* Use the trained model to iteratively modify the internal structure of a set of granular materials to attempt to reduce the density while maintaining mechanical stability.
* Analyse the internal structure evolution of the materials to understand the key features that contribute to mechanical stability as the density is varied and to predict optimal particle shape combinations for mechanical stability.

**Relevant Fields of Study*** Computer Science / Computational Science / Physical Science
* Software Engineering

**Location:** Clayton, VIC**Contact:** Gary Delaney via email gary.delaney@data61.csiro.au or Ritaban Dutta via email Ritaban.Dutta@data61.csiro.au  |
| Data61 70 | **Project Title**Exploring high dimensional data sets in Virtual Reality Environments.**Project Description**Virtual Reality Environments offer the potential to more deeply investigate scientific data and visualise simulation output in ways that were previously impossible. This project will build on existing data analytics and visualisation capabilities to create an integrated tool for visualisation and analysis of simulation data in an interactive Virtual Reality Environment. The project will focus on high dimensional data sets, and seek to take advantage of the additional spatial, auditory and haptic interaction that is possible in a VR environment to allow for deeper more intuitive exploration of high dimensional scientific data. Specific data sets that will be considered will include output from high throughput materials synthesis, simulations of 3D printing, and particle based models of industrial devices.**Project Duties/Tasks*** Develop enhancements to existing data processing pipelines in C++ and C# from CSIRO’s [Workspace](http://research.csiro.au/workspace/) platform to the Unity engine for analysis and display of high dimensional scientific data
* Develop within the Unity engine an example of a spatial, auditory and haptic interaction capability for interrogation and display of high dimensional scientific data
* Report on the effectiveness of the developed capabilities in enhancing the user’s speed and depth of understanding of complex scientific data

**Relevant Fields of Study*** Computer Science / Computational Science / Physical Science
* Software Engineering
* Information Technology (Games and Graphics Programming)

**Location:** Clayton, VIC**Contact:** Gary Delaney via email gary.delaney@data61.csiro.au or David Alexander via email david.alexander@data61.csiro.au  |
| Data61 71 | **Project Title**3D Data Management (VoxelNet) for Intelligent Mining.**Project Description**CSIRO is developing a multiuser, distributed virtual world system (the VoxelNet) for gathering, sharing and using 3D data supporting remote control of Unmanned Aerial Vehicles (UAVs) in underground mines. This project will develop methods for combining different types of 3D UAV data within the platform, including visual images, lidar, sonar range and inertial navigation data.**Project Duties/Tasks*** Implement a representation of sonar range data in a 3D virtual model
* Implement an additional representation of inertial navigation data in the 3D virtual model
* Compare these two paths together and also with a photogrammetrically derived path and 3D model in the same environment (the photogrammetry software already exists)
* Assign shape probabilities to mine void surfaces in voxels based upon comparing these different data sources

(Note that this project can be suitable for 1 to 3 vacation students.)**Relevant Fields of Study**Electrical Engineering, etc* 3D programming and computing science
* Mechatronics/robotics engineering

**Location:** Sandy Bay, Tasmania**Contact:** Charlotte Sennersten via phone on 0499240053 or email  charlotte.sennersten@csiro.au |
| Data61 72 | **Project Title**Radically Transparent Data Logging**Project Description**Working on the Confidential Computing's N1 product, design and implement a radically transparent data log.**Project Duties/Tasks*** Research distributed ledger systems such as blockchain.
* Create a standalone proof of concept
* Design and implement proposed solution in the N1 system

**Relevant Fields of Study**Electrical Engineering, etc* Computer Science

**Location:** Eveleigh, Sydney**Contact:** Stephen Hardy via email stephen.hardy@csiro.au  |
| Data61 73 | **Project Title**Privacy Preserving Voting.**Project Description**Working with the confidential computing group, create a secure, private electronic voting system **Project Duties/Tasks*** Get up to speed on state of the art cryptography for digital voting such as

the Damgard-Jurik partially homomorphic cryptosystem.* Implement a digital voting server application
* Design and implement a front end application

**Relevant Fields of Study*** Computer Science
* User Experience

**Location:** Eveleigh, Sydney**Contact:** Brian Thorne via phone on (02) 9490 5666 or email brian.thorne@csiro.au  |
| Data61 74 | **Project Title**ASPIRE to engage with industry.**Project Description**This project, [ASPIRE](http://www.csiro.au/en/Research/MF/Areas/Innovation/Agile/Reducing-waste-through-ASPIRE), is building a digital tool to support the match making of companies with complementary resources, one company with a by-product or waste that can be used as an input for another company. The student project will contribute to information, data gathering, system testing and industry engagement for ASPIRE in Melbourne – combining your ability to understand technical details with polished written skills to deliver this practical project for Victorian industry.**Project Duties/Tasks*** ASPIRE system testing & development
* Support the ASPIRE team’s engagement activities
* Research recyclable resource network flows

**Relevant Fields of Study*** Business & Marketing
* Engineering
* Computer Science

**Location:** Melbourne (CBD or Clayton) or Geelong**Contact:** Melanie Ayre via phone on (03) 9545 8473 or email melanie.ayre@csiro.au  |
| Data61 75 | **Project Title**Capability extraction.**Project Description**Data published by organisations and people (websites, social media posts, publications, job adverts, projects) contain clues to find expertise or capabilities of these organisations and people. For example, CSIRO on the website state “At the Commonwealth Scientific and Industrial Research Organisation (CSIRO), we shape the future. We do this by using science to solve real issues. Our research makes a difference to people, industry and the planet.”, which indicates a primary capability of the organization “scientific research”. Now if you further drill down you may find clues to figure out other capabilities of the organization, such as data analytics, and mining, manufacturing etc. Similar, clues exist in other published data. Recognising people and organisation capabilities are useful in many scenarios. Recruiting experts for projects, and selecting organisations to award tender projects are two such scenarios where such intelligence can be useful.In this project, we investigate the following two research questions?How capabilities can be extracted from free text published by people and organisations?How credible the extracted capabilities are?How can we rank organisations / people for capabilities?In the first phase of the project, we investigate how data published on websites can be used to build models to answer the above two questions. The project involves crawling the websites, and analyzing structural and textual data in websites. In the subsequent phases we investigate how information derived from other sources can be used to derive more accurate combined intelligence. Skills involved: programming, natural language processing, machine learning (text mining), and information retrieval.**Project Duties/Tasks*** Develop a software framework to gather and analyse data.
* Build analytical models.
* Author publication(s) based on findings

**Relevant Fields of Study*** Computer science
* Information Retrieval
* Machine Learning

**Location:** Marsfield, NSW or Clayton, Vic**Contact:** Gaya Jayasinghe via phone on (03) 95457965 or email Gaya.Jayasinghe@csiro.au  |
| Data61 76 | **Project Title**Automatic RDB2RDF schema mapping.**Project Description**Working on the Data Integration platform, design and implement an algorithm to map relational data sources to a RDF-based graph database**Project Duties/Tasks*** Research state-of-the-art RDB2RDF systems and semantic modelling for
* relational data sources
* Implement parts of a (semi-)automatic algorithm for schema mapping
* Apply to a real world scenario -e.g., commercial data, linked open data, semantic web

**Relevant Fields of Study*** Computer Science

**Location:** Eveleigh, Sydney**Contact:** Natalia Ruemmele via email Natalia.Ruemmele@data61.csiro.au  |