CDT Induction Event

Gert Aarts

(Swansea University)
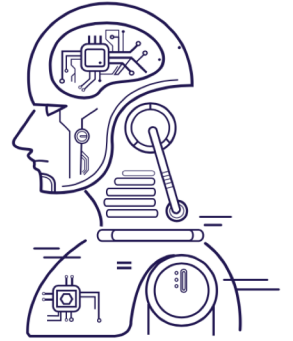
Outline

- Introduction
- Plan of the coming days
- Overview of CDT activities

Please ask questions/indicate you have a question through the **chat box**

If there are technical issues (lost connections, etc), please email Roz Toft r.toft@swansea.ac.uk or connect via Slack

Welcome to everyone!

- 2nd, 3rd and 4th year CDT students
- Supervisors
- External partners
- Supercomputing Wales Research Software Engineers
- CDT management and support teams

- **1st year cohort**

2020 cohort

Name	University	Project Title	Theme	Supervisor(s)
Luke Ian Lunn	Aberystwyth	Approximating the colour of Mars	T1, T3	Helen Miles
Bishnu Paduel	Aberystwyth	Automatic stroke recovery prediction using artificial intelligence	T2	Otar Akanyeti, Reyer Zwiggelaar
Will Robinson	Aberystwyth	Detecting when deep learning goes wrong in medical image analysis	T2	Bernie Tiddeman, Reyer Zwiggelaar
Franciszek Krzyzowski	Bangor	Learning from badly behaving data	T3	Lucy Kuncheva, Franck Vidal
Iwan Mitchell	Bangor	Automated optimisation of industrial X-ray computed tomography	T3	Franck Vidal, Simon Middleburgh
Jake Amey	Bristol	New Physics searches in B and D meson decays with machine learning	T1	Jonas Rademacker, Konstantinos Petridis
Matthew Selwood	Bristol	Using machine learning to explore the evolution of active galaxies with Euclid	T1	Sotiria Fotopoulou, Malcolm Bremer
Drew Barratt	Cardiff	Examination of SARS-CoV-2 severity, transmissibility and spread within Wales through the analysis of linked patient health records and genomic sequence data	T3	Tom Connor
Matthew Walker	Cardiff	Inferring brain tissue microstructure from standard structural imaging	T2	Leandro Beltrachini, Kevin Murphy
Samuel Wincott	Cardiff	AI and neuro-evolution: Exploiting network motifs to enhance prediction of contagion in complex networks	T3	Roger Whitaker, Alun Preece
Natalia Sikora	Swansea	Enhancing the diagnostic performance of a bowel cancer blood test using advanced machine learning algorithms and the incorporation of information from the patient's medical record	T2	Peter Dunstan, Dean Harris
Lukas Golino	Swansea	Machine learning with anti-hydrogen	T1	Niels Madsen, Gert Aarts
Maciej Glowacki*	Bristol	Searches for Beyond-Standard-Model signatures with jets + missing energy	T1	Henning Flaecher
Jacob Elford*	Cardiff	Monsters in the dark: gas, dust and star formation around supermassive black holes	T1	Timothy A. Davis, Mattia Negrello
David Mason*	Swansea	Non-perturbative dynamics and compositeness	T1	Biagio Lucini, Maurizio Piai
Jack Furby**	Cardiff	Human-machine collaboration with deep learning agents	T3	Alun Preece
Paul Murphy**	Cardiff	Adaptive neural networks through epigenetic processes	T3	Roger Whitaker
Ben Page**	Swansea	Studies of thermal QCD using lattice gauge theory	T1	Chris Allton

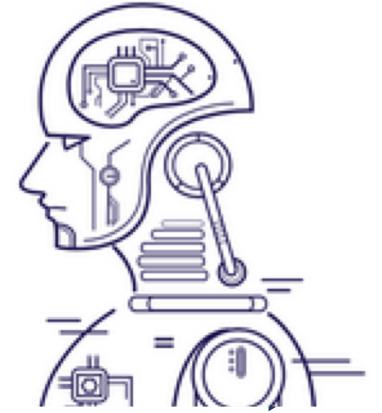*STFC CDT on Data-Intensive Science
**Associate member

First-Year Students

External partners

Many thanks for the placement positions

AIMLAC Slack channel

Slack is used for communication, for AIMLAC, DI and associated students

How to join our Slack channel: google how to join slack for generic instructions

slack.com/intl/en-gb/help/articles/212675257-Join-a-Slack-workspace

Our workspace is aimlac.slack.com

The 2020 cohort have received an invitation email, please join!

STFC Data Intensive Centre for Doctoral Training

University of BRISTOL — CARDIFF UNIVERSITY / PRIFYSGOL CAERDYDD Data Innovation Research Institute Sefydliad Ymchwil Arloesedd Data — Swansea University Prifysgol Abertawe

AIMLAC

// UKRI CENTRE FOR DOCTORAL TRAINING IN ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND ADVANCED COMPUTING

2020 CDT Induction Event, 22nd-24th September 2020

https://swanseauniversity.zoom.us/j/98098780985?pwd=TldLWlFNb2VkOVdMekV5NzgyQitqZz09

Tuesday 22nd September			
All Welcome (Students, Supervisors, Industry partners and Management Board members)			
9.15 a.m.	Welcome and CDT Overview - Ice breaker in small groups Prof Gert Aarts (AIMLAC) Prof Steve Fairhurst (DI)		
11.00 a.m.	Overview of taught components – Prof. Gert Aarts (All welcome, mostly relevant to 1st year students and supervisors)		
11:30 a.m.	**Break**		
11.45 a.m.	Introduction to the coding challenge – Mark Dawson and Ed Bennett (All welcome, mostly relevant to 1st year students and supervisors)		
12.30 p.m.	Equality, Diversity and Inclusion - Mr Andrew Lane, Cardiff University		
1.20 p.m.	**Break for lunch**		
	1st year students and supervisors	**3rd and 4th year students**	**2nd year students and supervisors, Industry partners and Management Board**
2.05 p.m.	Data Ethics Dr. Emma Tonkin, Bristol University	Future Thinking: Career Development, Collaborations and Networks Steve Hutchinson, Hutchinson Training Zoom link	Placement project overview talks from industry Zoom Link
3:00 p.m.	**Break**		
	1st year students and supervisors	**3rd and 4th year students**	**2nd year students and supervisors, Industry partners and Management Board**
3.15 p.m.	AI Ethics Prof. Reyer Zwiggelaar (50 mins)	Future Thinking Continued	Placement project overview talks from industry (continued). Zoom Link
5 p.m. close			
7.00 p.m.	Drop into the Virtual Pub – bring a drink, informal chat – breakout rooms available!		

https://swanseauniversity.zoom.us/j/98098780985?pwd=TldLWlFNb2VkOVdMekV5NzgyQitqZz09

Wednesday 23rd September **All welcome – (Students, Supervisors, Industry partners and Management Board members)**			
9:15 a.m.	Algorithmic bias Dr Bertie Mueller, Swansea University		
10.00 a.m.	Data Intensive CDT Research presentations		
10.50 a.m.	**Break**		
11.00 a.m.	4th year Data Intensive Science CDT students present placement experience (5-minute talks)		
	Industry Partners , Management Board members and student reps		
11:30 a.m.	External Stakeholder Board meeting (formerly External Advisory Board)		
12.20 p.m.	**Break for lunch**		
	Industry Partners and 2nd year students and supervisors	**3rd and 4th year students**	**1st year CDT students**
1.00 p.m.	Introduction to Placement session and student research poster presentations and breakout rooms	Break	Coding challenge (breaks allocated during session) Zoom link
2.00 p.m.	**Break**	2pm-4pm Hugh Kearns – Thinkwell "Turbocharge your writing" ZOOM LINK	
2.15 p.m.	Industry Placement matching (AKA speed dating) and further discussions		
3.45 p.m.	Students depart		
	Industry Partners and Management Board members		
3.45 p.m.	Feedback session		
4.15 p.m.	**Close**		
7.00 p.m.	Drop into the Virtual Pub – bring a drink, informal chat – breakout rooms available!		

https://swanseauniversity.zoom.us/j/98098780985?pwd=TldLWlFNb2VkOVdMekV5NzgyQitqZz09

Thursday 24th September Students only	
	All students
10.00 a.m.	Pastoral mentoring for 1st year students – informal discussions, Q&A Student rep to host
11.00 a.m.	Student Board Meeting and DataAid presentation – Ben Henderson
12 noon	**All students and Management Board members**
	Event feedback session
12.30 p.m.	**Close**
	All students (Management Board members optional)
1.30 p.m.	Team Building exercise – all students and Management Board members (Links and instructions will be provided on the day)

- There is one Zoom link at the top of page one, for the three-day event
- Some individual sessions have their own Zoom link – please check before joining

Comments on Agenda

One zoom link for all three days, including the Virtual Pub, see top of day 1
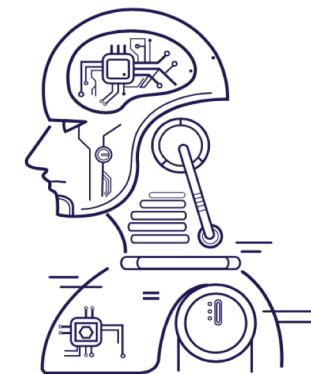
Zoom links are clickable in

- Word document, attached to invite
- PDF on Slack, after downloading

- Team building exercise (virtual escape room) on Thursday, please confirm attendance by end of today, by contacting Roz (email or Slack)

- For any connectivity issues, etc, please contact Roz r.toft@swansea.ac.uk

About the CDTs

![AIMLAC logo — UKRI Centre for Doctoral Training in Artificial Intelligence, Machine Learning and Advanced Computing]

Data Intensive CDT
Centre for Doctoral Training

PRIFYSGOL ABERYSTWYTH UNIVERSITY · PRIFYSGOL BANGOR UNIVERSITY · University of BRISTOL · CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD · Swansea University Prifysgol Abertawe · SUPERCOMPUTING WALES UWCHGYFRIFIADURA CYMRU

STFC CDT on Data-Intensive Science (DI)

- Since 2017

- Cardiff, Bristol, Swansea

- Particle physics, astrophysics, cosmology

- Stephen Fairhurst (CDT Director, Cardiff University)

- Fourth cohort starting today!

UKRI CDT in Artificial Intelligence, Machine Learning & Advanced Computing

- Since 2019

- Swansea, Aberystwyth, Bangor, Bristol, Cardiff, Supercomputing Wales

- Particle physics, astronomy, cosmology, biological and health sciences, mathematical and computer sciences

- Gert Aarts (CDT Director, Swansea University)

- Second cohort starting today!

Student cohorts

- 2017 DI CDT: 15 students
- 2018 DI CDT: 6 students
- 2019 AIMLAC + DI: 12 + 3 students
- 2020 AIMLAC + DI + associated: 12+3+3 students
- 2021-2023:   >11 students/cohort

54 students currently enrolled

- Collaboration across research themes, transfer of knowledge

- More experienced students can mentor and guide the newer students

Questions?

Break-out rooms: brief introduction

20 minute break for introduction and chat:

- You will be randomly assigned to a break-out room
- About 10 people per room

In two minutes:

➤ Introduce yourself

➤ Why you are here

➤ What you hope to achieve (in the context of the CDT)

About the CDTs

CDT proposition, compare with standard PhDs

- Research project in internationally leading groups, at cutting edge of science, resulting in PhD thesis

- Train PhD students to become fluent in AI, machine learning, data science, computing

- **Emphasis on computational and transferable skills training, delivered to the cohorts, to enhance wider skills set**

- **Ongoing engagement, including placements, with external partners to develop and fine-tune expectations from industry**

AIMLAC CDT research

Research is organised in three **research themes**:

- **T1**: **data from large science facilities**

  particle physics, astronomy, cosmology – Bristol, Cardiff, Swansea

- **T2**: **biological, health and clinical sciences**

  medical imaging, health data, bioinformatics – Aberystwyth, Cardiff, Swansea

- **T3**: **novel mathematical, physical, and computer science approaches**

  data, hardware, software, algorithms – Aberystwyth, Bangor, Cardiff, Swansea
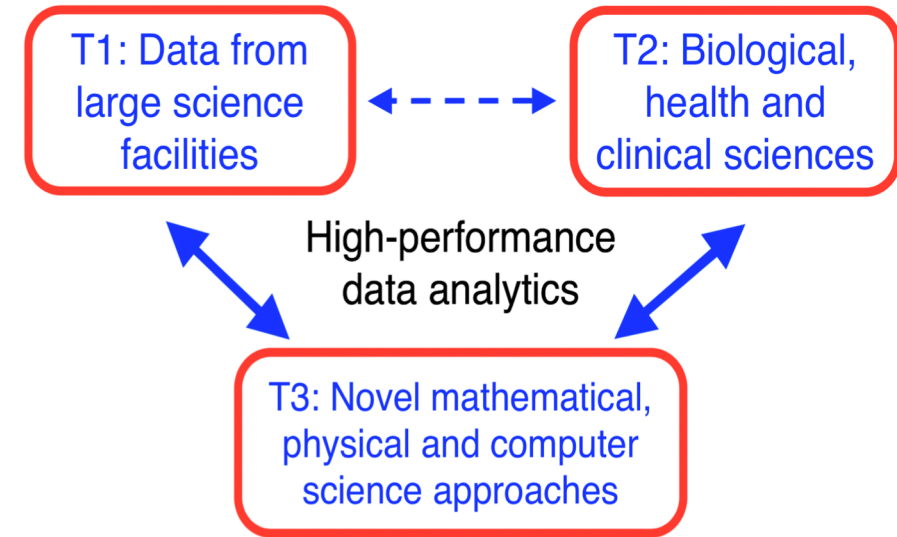
CDT research

- PhD project sits in one theme or across themes

- Interaction expected in and across cohorts: exchange of methods, experiences, best practice

- Students from diverse academic backgrounds – computer science, physics, mathematics, data science, engineering

- Training and cohort building essential to establish common base
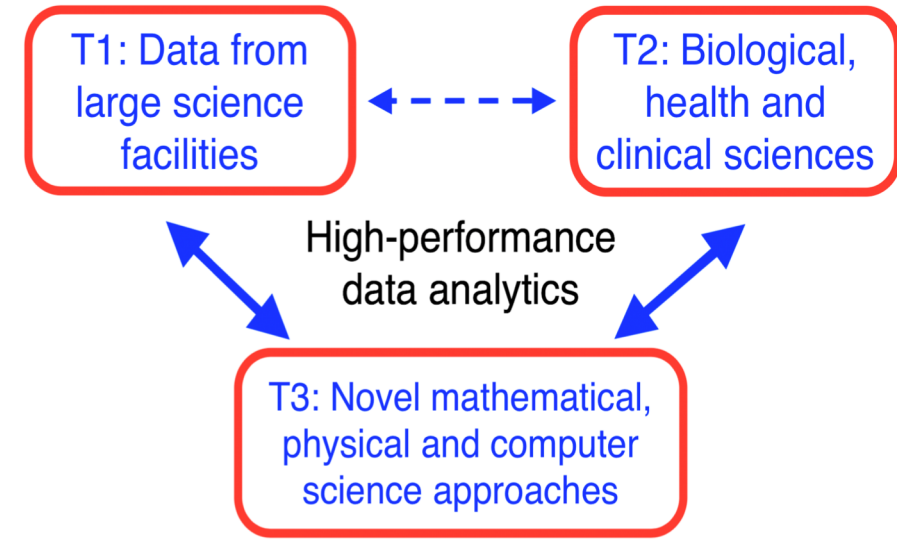
**CDT Student Handbook
2020 - 21**

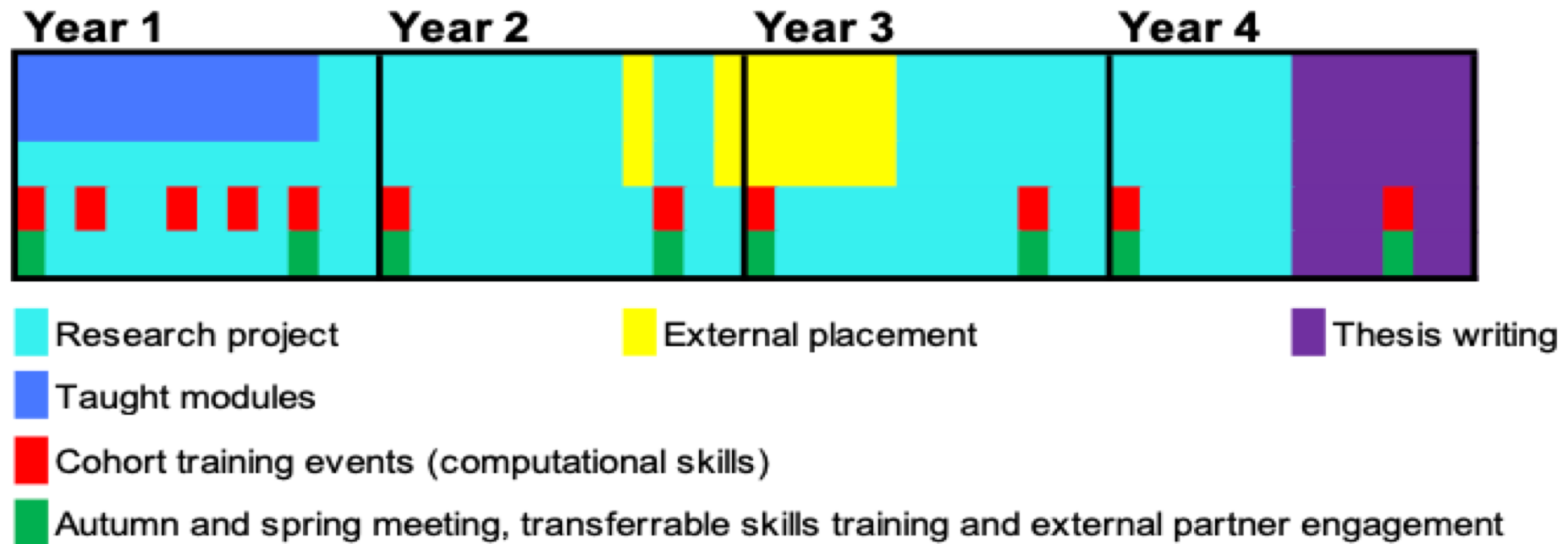
Handbook:

Provides comprehensive information on CDT delivery

Sent out to all students and supervisors

Also applies to DI students!

Timeline

	Research project		External placement		Thesis writing

Research project

Taught modules

Cohort training events (computational skills)

Autumn and spring meeting, transferrable skills training and external partner engagement

6 month placement can take place anytime during years 2 and or 3.

Timeline

- Activities in Year 1 discussed today at 11 and 11:45
    - aimed at 1st year students, but all are welcome

- Placements discussed extensively today/tomorrow
    - aimed at 2nd year students
    - presentations from 3rd, 4th year students on Wednesday, all are welcome

- Transferable skills training, responsible innovation, and equality, diversity and inclusivity are embedded throughout the 4 years

- Residential or virtual meetings

Events: 2019-2020

2019 Cardiff University

- September 24-26, 2019
- Induction Event
- Programme
- Main presentation
- Taught components
- Equality, diversity and inclusion

2019 Aberystwyth University

- November 18-20, 2019
- Software Carpentry

2020 Bristol University

- February 11-13, 2020
- Module Feedback
- Cloud Containers

Events: 2019-2020

2020 Bangor University (virtual)

- April 28-30, 2020
- Responsible Research and Innovation
- Presenting your Research
- Automated Testing and Continuous Integration

2020 Swansea University (virtual)

- June 9-11, 2020
- Jointly with the STFC CDT on Data-Intensive Science
- Object-Oriented Programming
- Responsible Research and Innovation
- Partner talks
- Poster presentations
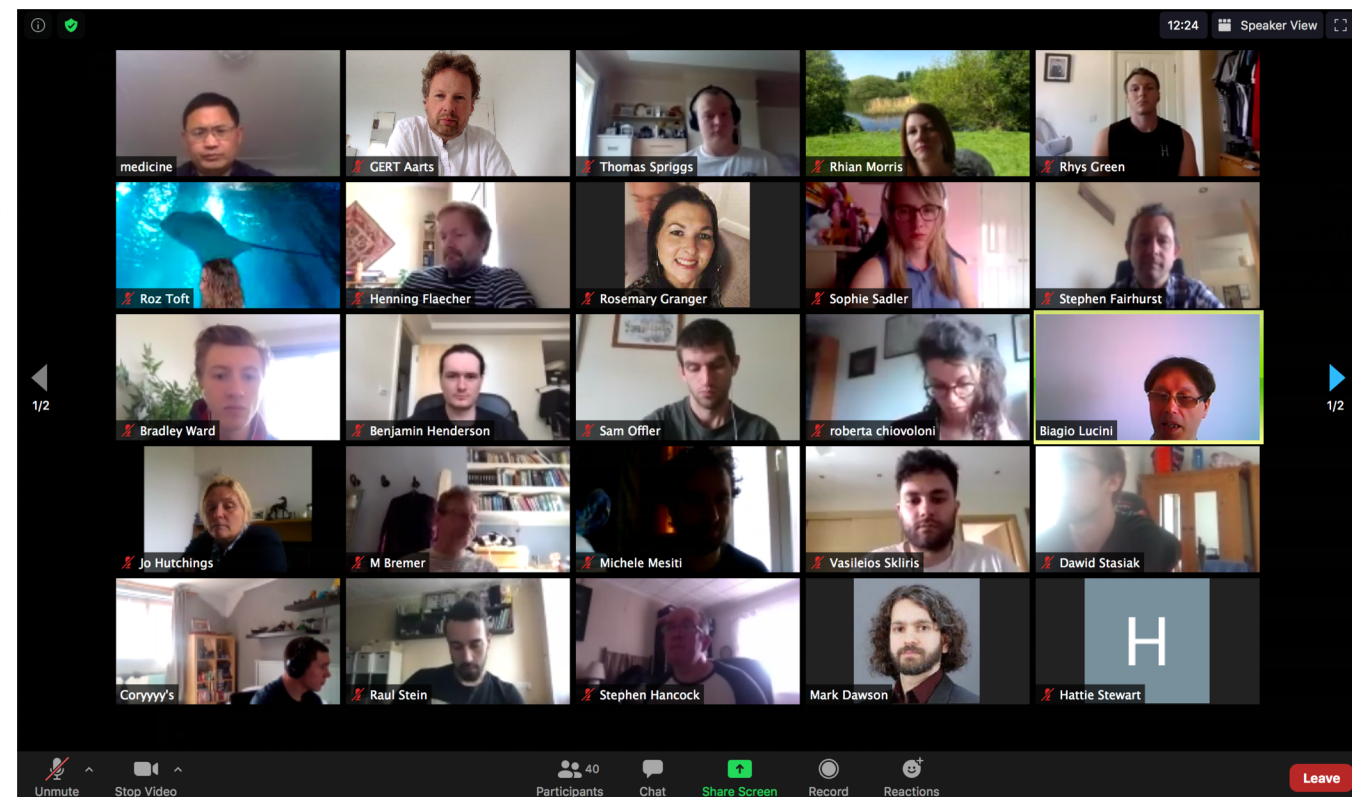
Resources: cdt-aimlac.org

Presenting your research

These sessions on *Presenting your Research* were delivered by Pamela Styles.

- Presenting Your Research Session 1, password 9o?8G47V
- Presenting Your Research Session 2, password 2j%=Y.4$
- Slides of the presentations

Training Resources

Our training is developed and delivered in collaboration with Supercomputing Wales and their outstanding team of Research Software Engineers. All the material that is used is available online under open-source licenses.

Software Carpentry

- Introduction to the Unix Shell: swcarpentry.github.io/shell-novice
- Introduction to programming with Python: swcarpentry.github.io/python-novice-inflammation
- Introduction to Version Control with Git: swcarpentry.github.io/git-novice

Resources: cdt-aimlac.org

Advanced introductory topics

- Performant Numpy: edbennett.github.io/performant-numpy
- Git: Beyond the Basics: markgdawson.github.io/git-beyond-the-basics

High Performance Computing

- Introduction to High Performance Computing with Supercomputing Wales: supercomputingwales.github.io/SCW-tutorial

Reproducible environments and containers

- Introduction to reproducible environments with Binder: zenodo.org/record/2598530
- Reproducible computational environments using containers: carpentries-incubator.github.io/docker-introduction

Automated testing and continuous integration in Python

- Introduction to automated testing: milliams.com/courses/software_engineering_best_practices/Testing.html milliams.com/courses/software_engineering_best_practices/Fixtures.html
- Advanced testing and continuous integration: chryswoods.com/python_and_data/testing

Object-Oriented Programming

- Introduction to Object-Oriented Programming in Python: edbennett.github.io/python-oop-novice

SUPERCOMPUTING WALES

UWCHGYFRIFIADURA CYMRU

The national
supercomputing
research facility
for Wales

Owain Huw
Programme Manager

**Mark Dawson, Ed Bennett,
Michele Mesiti, Colin Sauzé**
CDT (Senior) Research Software Engineers

Supercomputing Wales overview

£15m programme
until September 2021 with £9m ERDF grant through Welsh Government

13,080 cores
in two supercomputing hubs at Cardiff and Swansea deployed during 2018

15 RSEs
Research Software Engineers to leverage the power of the facilities

4 universities
across Wales delivering services as a consortium

1 petaflop
of compute power with access to >2 petabytes storage

Pan Wales team
providing technical and programme support to the consortium

RSEs – leveraging the power of HPC

- 15 Research Software Engineers working as a cohort across Wales

- Provide expertise required to leverage power of HPC facilities – not just about raw 'core count', but about efficient, optimised use of the resources

- Supporting delivery of major projects and scientific breakthroughs

- Increase capacity to attract competitive research funding

- Support world-leading outputs and build an international reputation

Two important Boards

Student Boards:

Current reps: Tom Spriggs, Tonicha Crook (Yr2)
               Andrew Hannington (Yr3) Dawid Stasiak (Yr3)

Equality, Diversity and Inclusivity (EDI) Board:

Chair: Roberta Chiovoloni (Yr4), EDI Lead: Biagio Lucini

CDT Management teams

Rhian Melita Morris – AIMLAC CDT manager

Roz Toft – AIMLAC Research support officer

Rosemary Granger – DI CDT manager

cdt-aimlac.org

data-intensive-cdt.ac.uk

STFC Data Intensive Centre for Doctoral Training

AIMLAC

// UKRI CENTRE FOR DOCTORAL TRAINING
IN ARTIFICIAL INTELLIGENCE, MACHINE
LEARNING AND ADVANCED COMPUTING

2020 CDT Induction Event, 22nd-24th September 2020
https://swanseauniversity.zoom.us/j/98098780985?pwd=TldLWIFNb2VkOVdMekV5NzgyQitqZz09

Tuesday 22nd September **All Welcome (Students, Supervisors, Industry partners and Management Board members)**			
9.15 a.m.	Welcome and CDT Overview - Ice breaker in small groups Prof Gert Aarts (AIMLAC) Prof Steve Fairhurst (DI)		
11.00 a.m.	Overview of taught components – Prof. Gert Aarts (All welcome, mostly relevant to 1st year students and supervisors)		
11:30 a.m.	**Break**		
11.45 a.m.	Introduction to the coding challenge – Mark Dawson and Ed Bennett (All welcome, mostly relevant to 1st year students and supervisors)		
12.30 p.m.	Equality, Diversity and Inclusion - Mr Andrew Lane, Cardiff University		
1.20 p.m.	**Break for lunch**		
	1st year students and supervisors	**3rd and 4th year students**	**2nd year students and supervisors, Industry partners and Management Board**
2.05 p.m.	Data Ethics Dr. Emma Tonkin, Bristol University	Future Thinking: Career Development, Collaborations and Networks Steve Hutchinson, Hutchinson Training Zoom link	Placement project overview talks from industry Zoom Link
3:00 p.m.	**Break**		
	1st year students and supervisors	**3rd and 4th year students**	**2nd year students and supervisors, Industry partners and Management Board**
3.15 p.m.	AI Ethics Prof. Reyer Zwiggelaar (50 mins)	Future Thinking Continued	Placement project overview talks from industry (continued). Zoom Link
5 p.m. close			
7.00 p.m.	Drop into the Virtual Pub – bring a drink, informal chat – breakout rooms available!		