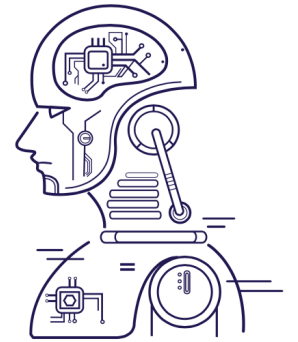# CDT Induction Event

Gert Aarts

(Swansea University)

Outline

- Introduction
- Who is who
- Overview of CDT activities

Please ask questions if things are unclear!

Agenda

Tuesday 20th September – Kidwelly Suite	
1st year students only	
10.00am	Arrival tea and coffee *in Second floor breakout area*
11.00am	Student welcome and overview of AIMLAC and DI CDT's by Prof Gert Aarts in TBC
12 - 1pm Lunch - *Urban Restaurant*	
1.00pm	Introduction to Collaborative Software Development
2.30pm	Tea and coffee break *in Second floor breakout area*
3.30pm	Introduction to Collaborative Software Development Cont'd
4.30pm	Taught components overview and training events in the TBC
5.00pm close	
7.00pm	Dinner - TBC

Agenda

Wednesday 21st September – various rooms see below

	CDT Industry Partners		
10.00am	External Partner Meeting - *Tintern Suite*		
10.30am	**Coffee break** – poster board discussions by 2nd years – 3rd floor Breakout area		
	1st year students	**2nd Year Students**	**3rd / 4th year students**
11.00am	Dr Emma Tonkin- Data Ethics *in Portmeiron Suite*	Industry and student session - industry placement presentations *in Caernarfon Suite*	Louisa Clark - Challenge Confident *in Tintern Room*
12.00pm		Speed dating (*in Caernarfon Suite*	

1.00pm - 2.00pm	Lunch - *Urban Restaurant*	
	All Students - *Caernarfon Suite*	**Partner and Academics - *KIdwelly Suite***
2:00pm	Prof Reyer Zwiggelaar- AI Ethics	External Partner Consultation
		Academic Consultation
3.30pm	**Coffee break with networking**	
4.00pm	Presentation from AIMLAC's Second Year Student - Collaborative Software Development – *Caernarfon Suite*	
4.45pm	Prof Gert Aarts - Consultation with Students	
	5.30pm close	
6.00pm	Pre-dinner network	
6.30pm	Industry and student Dinner - *Brecon Suite*	
8.00pm	Phil Hall from Elzware – I am Echoborg - *Brecon Suite*	

Thursday 22nd September

All students - *Caerphilly Suite*	
09.30am	Tea and Coffee – *Third floor breakout area*
10.00am	Student Board Meeting – Chair Robbie Webbe
11.00am	Closing the Feedback from students and Chair of the Student board
11.30am	Cohort Building - Escape Rooms Cardiff.co.uk Churchill Way Exodus / Boat trip round the bay
11.30am - Close of Induction, Training and Cohort Building	

Who are we? Who are you?

Gert Aarts – CDT Director AIMLAC

Steve Fairhurst – CDT Director Data-Intensive Science (DIS)

Rhian Morris – CDT Manager

Roz Toft – CDT Research Manager

Rosemary Granger – CDT Manager DIS

cdt-aimlac.org

data-intensive-cdt.ac.uk

AIMLAC Code of Conduct

The Code of Conduct applies to AIMLAC CDT students, associated and aligned students, management board members, supervisors and professional services staff involved with CDT activities. In addition, this Code of Conduct has to be abided by all external stakeholders in their interactions with the CDT, including industrial partners and external participants to CDT events.

AIMLAC is committed to making its activities productive and enjoyable for everyone. Creating a supportive professional environment where open and frank discussion of ideas can take place, where everyone is treated with courtesy and respect, and in which diversity and inclusion are valued, is the responsibility of all those involved. We will not tolerate harassment of AIMLAC members or others involved in AIMLAC in any form. For the entire duration of an AIMLAC activity and in other professional interactions with colleagues you agree to follow these guidelines:

- Behave professionally in personal interactions as well as in any other form of communication including social media. Harassment and sexist, racist, or exclusionary comments or jokes are not appropriate. Harassment includes sustained disruption of talks or other events, inappropriate physical contact, sexual attention or innuendo, deliberate intimidation, stalking, and photography or recording of an individual without consent. It also includes, but is not limited to, offensive comments related to gender identity, sexual orientation, disability, physical appearance, body size, race, nationality, or the religion or non-religion of participants.
- Be kind to others. Do not insult or put down AIMLAC members or other individuals associated with AIMLAC. Scientific discussion and criticism are vital and should be conducted in this spirit.
- All communication should be appropriate for a professional audience including people of many different backgrounds.
- If participants wish to share photos of a speaker on social media, we strongly recommend that they first obtain the speaker's permission. Participants may share the contents of talks/slides via social media unless speakers have asked that specific details/slides should not be shared.

Should an AIMLAC member or another individual be asked to stop any inappropriate behaviour, they are expected to comply immediately. In serious cases, they may be asked to leave the activity at the sole discretion of the AIMLAC Management Board. They may also be banned from participation at future activities.

Should an individual witness events of bullying, harassment or aggression, we recommend that they approach the affected person and show support. The witness may also suggest that the inappropriate behaviour be reported and offer to facilitate that reporting if requested. Participants can report any violation of these guidelines to the activity organizers or members of the AIMLAC Management Board. Such reports will be treated confidentially.

AIMLAC members, non-affiliated students, and other participants to CDT events are subject to the policies and regulations at their Home Institution. Violations of the Code of Conduct will be reported to those.

Thank you for helping to make AIMLAC welcoming to all.

The AIMLAC Management Board

cdt-aimlac.org/cdt-coc.html

2022 cohort

Name	University	Project Title	Theme	Supervisor(s)
Michael Casaletto	Aberystwyth	Prediction of facial growth for children with cleft lip and palate using 3D data mining and machine learning	T2, T3	Richard Jensen
Luke Williams	Aberystwyth	Collaborative mapping of large scale outdoor environments	T3	Myra Wilson
Jamie Pointon	Bangor	X-ray simulation and deep learning: Application to Automatic segmentation of defects in CT images corrupted by artefacts	T3	Franck Vidal
Preben Vangberg	Bangor	Automatically Analysing Big Language Data	T3	William Teahan
Rhys Shaw	Bristol	Machine learning and radio source multiplicity	T1	Mark Birkinshaw
Tanya Kushwahaa	Cardiff	Exploiting GAIA data and understanding the galaxies' past histories with machine learning	T1	Mikako Matsuura
Sama Al-Shammari	Cardiff	Simulation-based Inference of gravitational waves signals from black holes and neutron stars	T1	Vivien Raymond
Chanju Park	Swansea	Learning (from) lattice field theory	T1	Gert Aarts, Biagio Lucini
Vasiles Balabanis	Swansea	Multimodal analysis of Anatomical and Functional features to enhance the understanding of Brain Processing Phenomena: A Machine Learning Approach.	T2, T3	Scott Yang

First-Year
Students

2021 cohort

Name	University	Project Title	Theme	Supervisor(s)
Myles Clayton	Aberystwyth	A deep learning framework for agricultural plant breeding that predicts genotype-phenotype associations	T2	Martin Swain, Chuan Lu
Ong Ding Sheng	Aberystwyth	Few-shot Learning for Environment Adaptive Multi-modal Vision System	T3	Jungong Han
Leena Sarah Farhat	Bangor	Bringing big-data to social science	T3	Simon Willcock, William Teahan
Dan Farmer	Bangor	Edge-based object recognition for immersive analytics in Web-based XR	T3	Panagiotis (Panos) Ritsos
Sam Hennessey	Bangor	Ensembles of Deep Neural Networks for Semi-supervised Learning	T3	Lucy Kuncheva
Fergus Baker	Bristol	Machine learning to study accretion flows around black holes	T1	Andy Young
Laura Ballisat	Bristol	Advanced computational methods for dosimetry, planning and verification in emergent radiotherapy treatments	T1, T2	Jaap Velthuis, Richard Hugtenburg (Swansea)
Matthew Powell	Cardiff	Real-time Situational Understanding using Deep Neural Networks and Knowledge Graphs	T3	Alun Preece
Zara Siddique	Cardiff	Evolving Ethical Deep Neural Networks	T3	Roger Whitaker
Luke Golby	Swansea	AI based approaches multi-dimensional functional genomics	T2	Steve Conlan
Tabitha Lewis	Swansea	ML-guided dynamical systems modelling of sepsis	T2, T3	Noemi Picco
Shobhna Singh*	Cardiff	Dimer models on quasicrystals	T1	Felix Flicker

*Associate member

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

2019 cohort

| Name	University	Project Title	Theme	Supervisor(s)
Lily Major	Aberystwyth	Big Data algorithmics for efficient search and analysis of large collections of genomes	T2	Amanda Clare, Jacqueline Daykin, Benjamin Mora, Christine Zarges
Cory Thomas	Aberystwyth	Modelling the development of breast cancer abnormalities	T2, T3	Reyer Zwiggelaar, Tom Tornsey-Weir, Jason Xie
Benjamin Winter	Bangor	The research of neuroevolution algorithms	T3	William Teahan, Franck Vidal
Hattie Stewart	Bristol	AI techniques for extracting source information from Square Kilometre Array (SKA) datasets	T1	Mark Birkingshaw
Robbie Webbe	Bristol	X-Ray Astronomy, concerning the identification and classification of highly variable AGN	T1	Andy Young
Christopher Wright	Bristol	Multi-channel waveform reconstruction for dark matter searches with LUX-ZEPLIN	T1	Henning Flaecher, Stephen Fairhurst
Michael Norman	Cardiff	Deep learning for real-time gravitational wave detection	T1	Patrick Sutton
Bradley Ward	Cardiff	Investigating the epoch of galaxy formation using artificial intelligence	T1	Steve Eales
Tonicha Crook	Swansea	Game theory	T3	Arno Pauly, Edwin Beggs
Jamie Duell	Swansea	Machine learning in medical science	T2	Xiuyi Fan, Shangming Zhou, Gert Aarts
Sophie Sadler	Swansea	Visual analytics for explainable graph-based machine learning	T3	Daniel Archambault, Mike Edwards
Raul Stein*	Bristol	FPGA implementation of machine learning for low latency data processing in particle detectors	T1	Jim Brooke
Eleonora Parrag*	Cardiff	Rewinding supernovae with machine learning	T1	Cosimo Inserra
Thomas Spriggs*	Swansea	Spectral features of hadronic states in thermal QCD	T1	Chris Allton, Tim Burns

*STFC CDT on Data-Intensive Science

4

Boards/activities: later in the meeting

o Student board

o EDI: Equality, diversity and inclusion board

o DataAid

o RSE: Research software engineer
  practical training, coding challenge

AIMLAC Slack channel
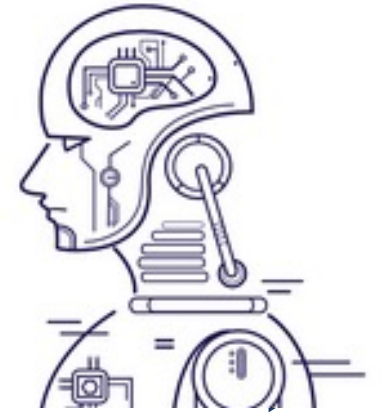
Slack is used for communication

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

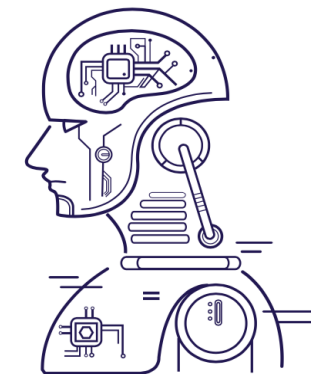The 2022 cohort will receive an invitation email, please join!

Questions?

About the CDTs

STFC CDT on Data-Intensive Science (DI)

- Cohorts started in 2017-2020
- Cardiff, Bristol, Swansea
- Particle physics, astrophysics, cosmology
- Stephen Fairhurst (CDT Director, Cardiff University)

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)
- Fourth cohort starting now (five cohorts in total)

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

PRIFYSGOL ABERYSTWYTH UNIVERSITY

PRIFYSGOL BANGOR UNIVERSITY

University of BRISTOL

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

Swansea University Prifysgol Abertawe

SUPERCOMPUTING WALES UWCHGYFRIFIADURA CYMRU

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 AIMLAC + associated: 11 + 1 students
- 2022 AIMLAC: 10 students

currently 61 students enrolled

- Collaboration across research themes, transfer of knowledge

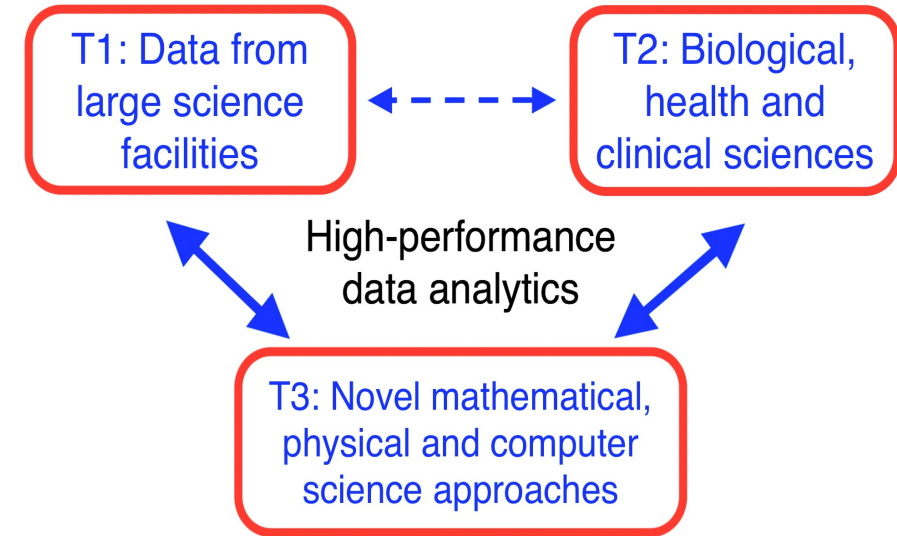- More experienced students can mentor and guide the newer students

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

T1: Data from large science facilities ← - - - - → T2: Biological, health and clinical sciences

High-performance data analytics

T3: Novel mathematical, physical and computer science approaches

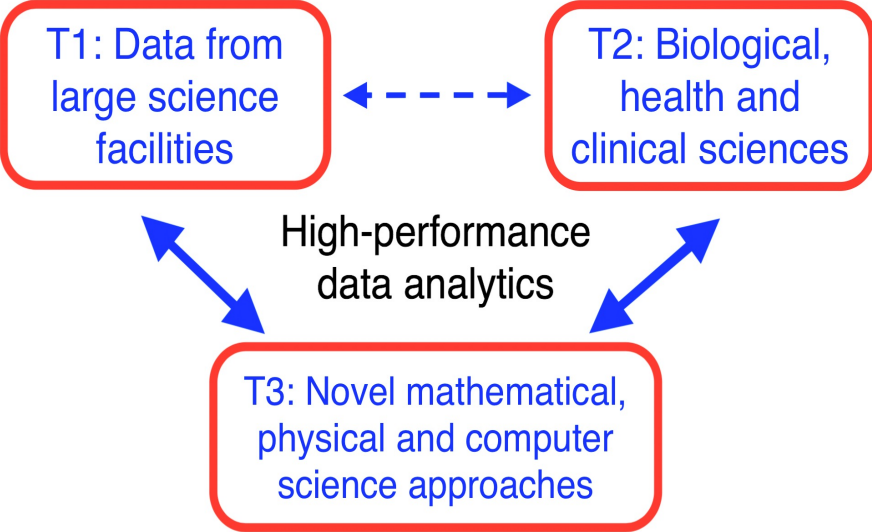Research is organised in three **research themes**:

- **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

T1: Data from large science facilities

T2: Biological, health and clinical sciences

High-performance data analytics

T3: Novel mathematical, physical and computer science approaches

- 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

Handbook

Provides comprehensive information on CDT delivery

Sent out to all students and supervisors

CDT Student Handbook
2022 -2023

Information: `cdt-aimlac.org`

- past events

- resources, technical and transferable skills training

- cohorts/people/partners

- code of conduct

- DataAid

- ….

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

Questions?