

CDT Induction Event

Gert Aarts (Swansea University)

Steve Fairhurst (Cardiff University)



Welcome to

STFC CDT on Data-Intensive Science (DI-CDT)

UKRI CDT in Artificial Intelligence, Machine Learning & Advanced Computing (AIMLAC)

Outline

- Introduction
- Some background info
- Plan of the coming days

Informal: please feel free to interrupt, ask questions

STFC CDT on Data-Intensive Science (DI-CDT)

- First Cohort of students started in 2017
- Third cohort starting now!
- Cardiff, Bristol, Swansea
- Particle physics, astrophysics, gravitational physics, cosmology
- Stephen Fairhurst (CDT Director, Cardiff University)
- Rosemary Granger (CDT Manager)
- data-intensive-cdt.ac.uk

Data Intensive CDT
Centre for Doctoral Training



UKRI CDT in Artificial Intelligence, Machine Learning & Advanced Computing



- First cohort starting now!
- 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)
- Rhian Morris (CDT manager), Roz Toft (CDT support officer)
- cdt-aimlac.org

Introduction



Research councils in the UK

UK Research and Innovation

Investing in the highest quality research and innovation across the UK and fostering a collaborative environment for universities, researchers and businesses



Innovate UK



VIA (very important acronyms...)

- **UKRI**: UK Research and Innovation – sits above the Research Councils, created in 2018, enables (interdisciplinary) research and training
- **EPSRC**: Engineering and Physical Sciences Research Council
- **STFC**: Science and Technology Facilities Council
- **CDT**: Centre for Doctoral Training – train PhD students via cohort delivery, in collaboration with industrial partners

UKRI call for CDTs in AI

- Call: February 5, 2018
- Outline proposal submission: March 28, 2018
84 proposals submitted, 37 invited for full proposal
- Full proposal deadline: July 31, 2018
- Interview: November 7, 2018
- Announcement: February 6, 2019
16 CDTs awarded
- First cohorts to start in October 2019

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 Overview



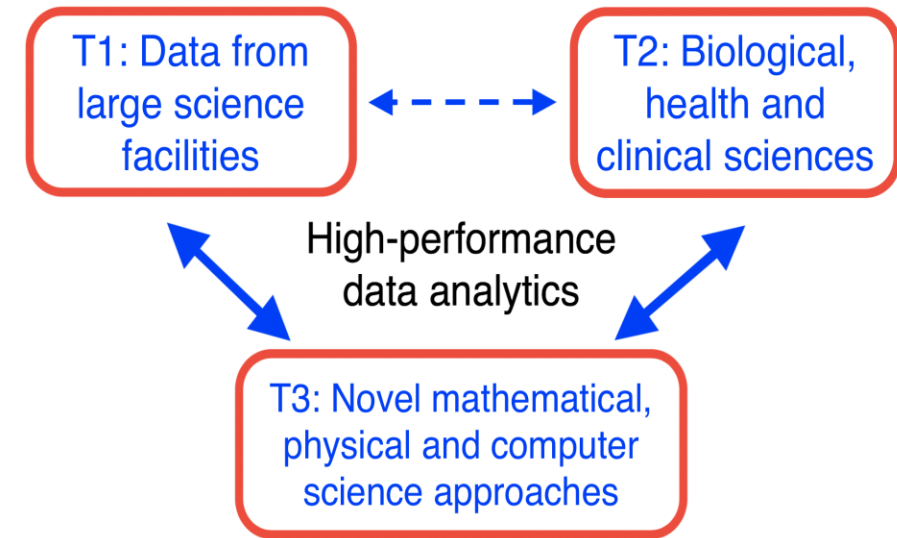
25



AIMLAC CDT research

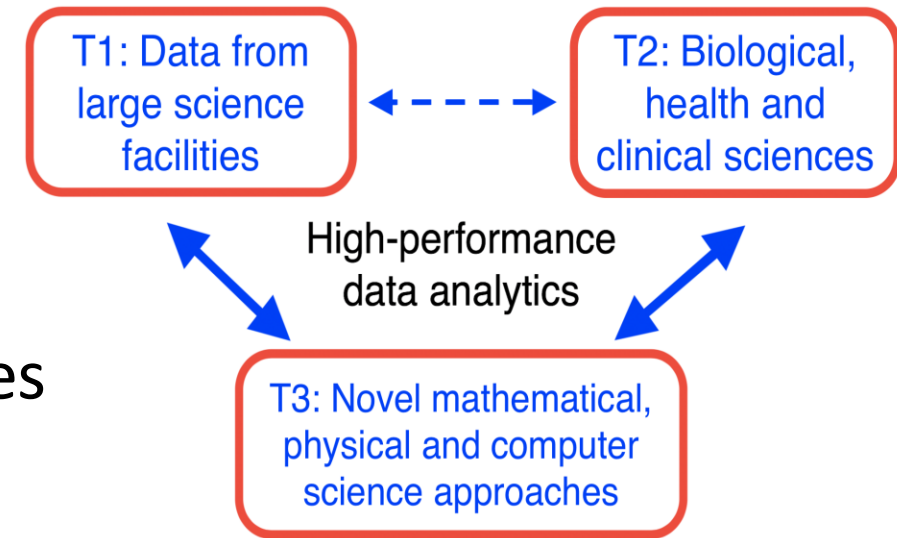
Research is organised in three **research themes**:

- **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



T1: data from large science facilities

Particle physics, astronomy, cosmology: example projects

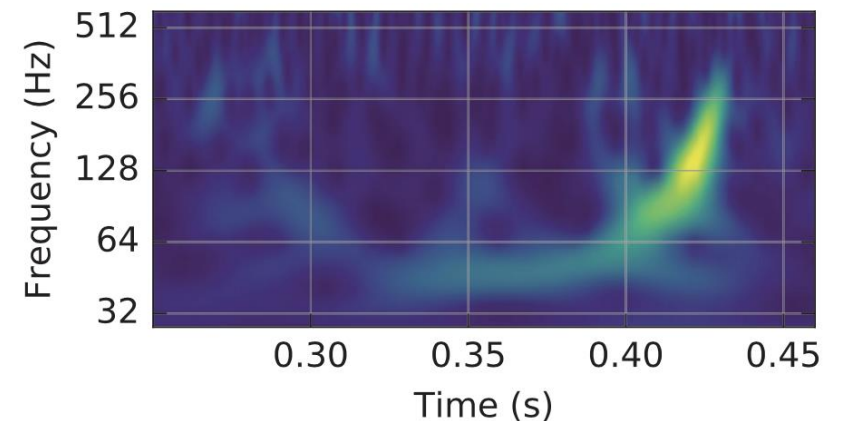
- Dark matter detection at LZ experiment in South Dakota

- Event reconstruction from raw waveform data

B.P. Abbott et al,
PRL **116** (2016)
061102

- Real-time gravitational wave detection

- Identify signals from merging neutron stars and black holes



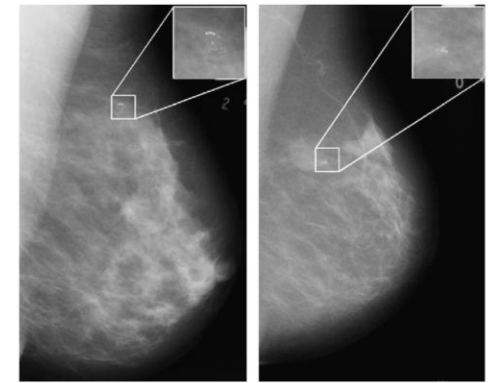
- Phases of matter from Monte Carlo data

- Classify phases and symmetry breaking in theories without order parameters

T2: biological, health and clinical sciences

Medical imaging, health data, bioinformatics: example projects

- Development of breast cancer abnormalities
 - Mammographic morphology, manifold modelling
- Early prediction of cancer from electronic health records
 - Processing of primary and secondary EHRs for Wales
- Tracking and treatment of sexually transmitted diseases
 - Linking of genomic sequence data to population-level health data



A. Oliver et al,
Knowledge-Based System
28 (2012) 68

T3: novel mathematical, physical, and computer science approaches

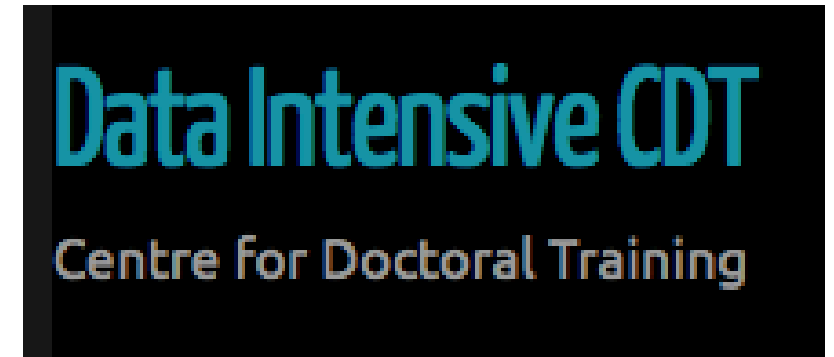
Data, hardware, software, algorithms: example projects

- Spiking neural networks for enhanced situational awareness
 - In collaboration with Crime and Security Institute
- Precision and convergence of ML, using HPC and GPUs
 - Linking HPC experience with progress in AI
- Artificial intelligence for immersive analytics
 - Context-aware, context-adaptive and predictive interfaces for AI



STFC CDT on Data-Intensive Science (DI-CDT)

- 16 students in first cohort (2017)
- 6 students in second cohort (2018)
- 3 students in third cohort (2019)
funded through industry contributions
- Steve Fairhurst, Paul Clark (Cardiff),
Henning Flaecher, Malcolm Bremer (Bristol),
Simon Hands, Biagio Lucini (Swansea)
- Rosemary Granger (CDT manager)
- <https://data-intensive-cdt.ac.uk>



DI-CDT Update

- STFC funded 8 Data-Intensive CDTs
 - STFC review of CDTs completed, January 2019, with positive feedback ... but, no future funding at present
 - 2017 cohort undertaking placements, now or soon
 - 2018 cohort will be matched to external partners at this meeting
 - 2019 cohort starting now
-
- 2017 and 2018 cohorts applying DI skills and knowledge to their research
 - Starting to see research publications (please let us know about them)
 - James Dawson (3rd year student) won the ARIEL AI challenge

DI-CDT Residential Courses and Workshops

- Induction event, September 2018, Vale Resort, Cardiff
 - Introduction to CDT for new students, version control, hackathon
 - Matching of students with placement projects
 - Cohort building
 - Attended by industry partners
- National CDT event, November 2018, Edinburgh
 - Training events
 - Careers
 - Chance to meet students from other CDTs
- Software Carpentry, November 2018, Swansea
 - Version control
 - UNIX, Python programming
- Data Carpentry, March 2019, Bristol
 - Introduction to SQL, R
 - External stakeholder board meeting
- CDT Summer school, Sussex, July 2019
 - AI and Machine learning training
 - Attended by industry partners from various CDTs

Student cohorts and mentoring

- STFC DI 2017: 15 students
- STFC DI 2018: 6 students
- AIMLAC + DI 2019: 13 + 3 students
- AIMLAC + DI 2020-2023: at least 11 students/cohort
- Collaboration across research themes, transfer of knowledge
- More experienced students can mentor and guide the newer students
- Student Advisory Board, Staff-Student Panel: participate!

Annual meetings

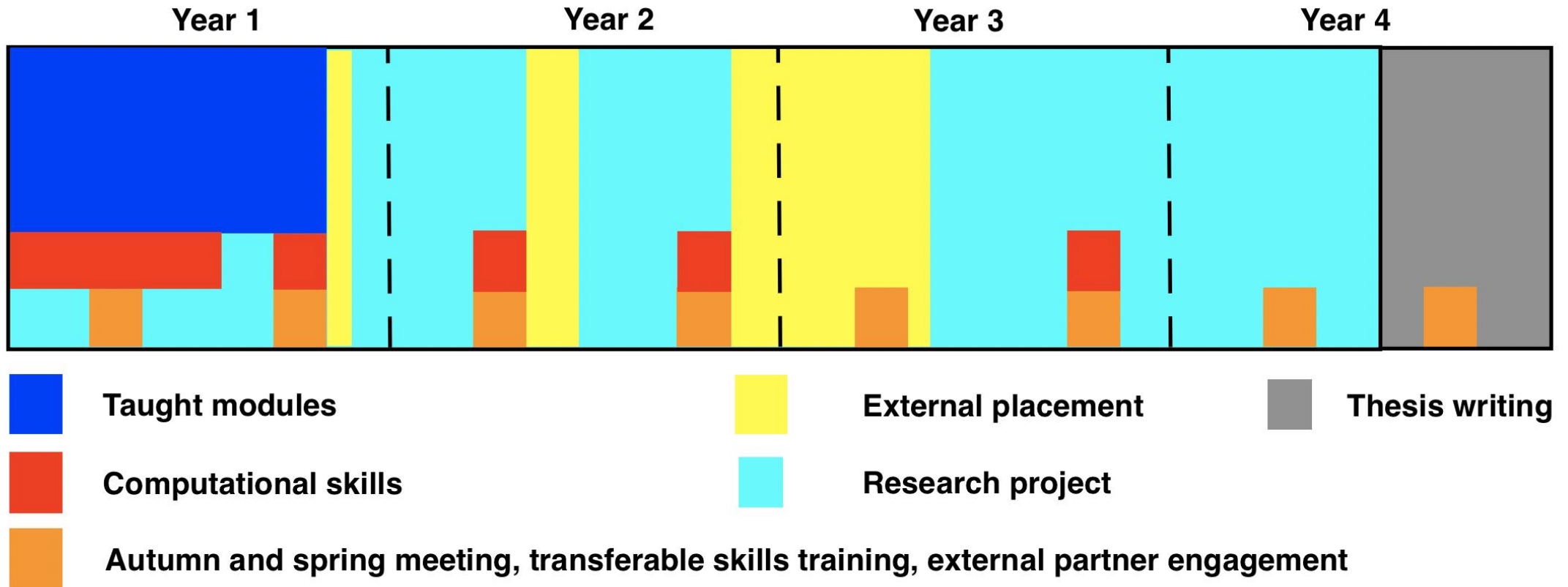
Two meetings per year for all students, supervisors, external partners

- All cohorts and external partners invited, networking and feedback
- Autumn meeting:
 - welcome to new cohort
 - Pairing of students with external partners
 - Mentoring, feedback, transferable skills, ...
- Spring meeting: annual conference
 - Science! oral and poster presentations, best poster prize
 - EDI, RI, industrial research presentations by external speakers
 - Mentoring, feedback, transferable skills, ...
- Stakeholder board and External Board meetings

PhD timeline

- 4-year PhD
- Fully funded
- Taught components
- Responsible innovation
- Transferable skills
- External partner placements
- Intra- and inter-cohort meetings

Student trajectory



More details in afternoon session for 2019 cohort

Details of the placements

- The students will remain enrolled at the University during the placement
 - They will be paid by the host university
 - They will not be members of staff at the partner organisations
 - CDT and academic supervisors will retain project oversight
- An agreement will be signed between the student and partner detailing IP expectations
- If possible, and appropriate, students will be encouraged to publish the results obtained during the placement.

DI/AIMLAC Partners



AMPLIFYFI

Atos

DiRAC Distributed Research utilizing Advanced Computing



[dstl]

e Evolved Intelligence



IBM



Microsoft

n-eos characterising coatings and membranes

nVIDIA

ORACLE



QUANTUM ADVISORY WORKING IN PARTNERSHIP WITH YOU

Q.F. The Quant Foundry

QinetiQ

Stanley Black & Decker

Science & Technology Facilities Council



we predict

Dŵr Cymru Welsh Water

PRIFYSGOL ABERYSTWYTH UNIVERSITY

PRIFYSGOL BANGOR UNIVERSITY

University of BRISTOL

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

Swansea University Prifysgol Abertawe

SUPERCOMPUTING WALES UWCHGYFRIFIADURA CYMRU

DI-CDT Placements

- All CDT students are expected to undertake a six-month placement with an external partner
- All students from the 2017 cohort have been matched with a placement partner and undertaken two-week initial placement
- Many students are on 6-month placement now. Others will start in the coming weeks/months.
- Talk from Rhys Green about placement in the afternoon session
- Session this morning for students to give placement updates, flag any issues, give advice to 2018 cohort

DI-CDT Placements

- All 2nd year DI students have prepared a 1-page pro-forma summarizing their expertise and interests
- We have received project descriptions from partners
- This afternoon, we have a session for the partners to introduce projects, students to introduce themselves and their research and time for one-on-one discussions
- Following this, we will finalize placement details and set up 2-week initial engagements



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

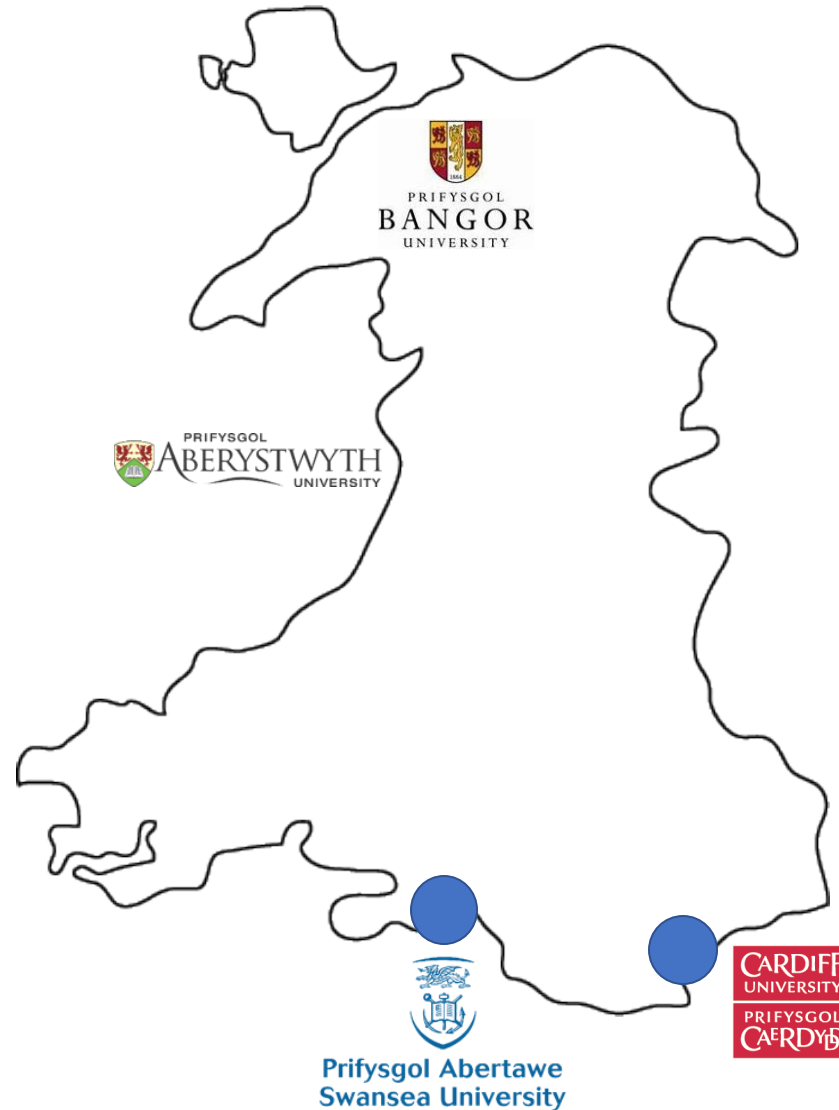
providing technical and programme support to the consortium



The supercomputing landscape in Wales



The **world's first joint Atos Dell EMC Supercomputing Centre of Excellence**, with development opportunities for the Supercomputing Wales Research Software Engineers



Two hubs – 8,040 core hub in Cardiff and a 5,040 core hub in Swansea

High Performance Data Analytics platform for data science and AI applications

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



Agenda: Wednesday

Wednesday 25 th September	
From 7am	Breakfast in the Grey restaurant
8.45am	Arrival tea and coffee in the Sophia suite
9.00am	Student welcome and overview of AIMLAC and DI CDT's in the Sophia suite
10.15am	Tea and coffee break

	1 st year students	2 nd and 3 rd year students	
10.30am	Introduction to Coding Challenge in the Sophia suite	Placement feedback and discussion in the Roath	
12.30pm	Lunch and Industry arrival in the Atrium		
1.30pm	Industry and student session in the Castle suite		
2.45pm	Tea and coffee break in the Atrium		
	1 st year students	2 nd years & 2019 Industry placement DI CDT partners	3 rd Year students & Industry partners
3pm-5pm	Taught components overview in the Sophia suite	Industry placement matching and discussion in the Castle suite	Careers session in the Roath suite
5.00pm close			
6.30pm	Drinks reception in the Atrium		
7.30pm	Dinner in the Castle Suite		

Agenda: Wednesday Afternoon

1.30 pm	Welcome & CDT Overview	Gert Aarts
1.40 pm	Introductions	All
1.55 pm	Equality, diversity and inclusivity	Biagio Lucini
2.00	AI and big data in academia	Cosimo Inserra (Cardiff)
2.15	AI and big data at EDF Energy	Hannah LoebI (EDF)
2.30	AI and big data at Amplyfi	Mark Woods (Amplyfi)
2.45	Placement experience at Oracle	Rhys Green (Cardiff)

Agenda: Wednesday Afternoon -- Careers

Careers talks from: Hannah Loeb, Cosimo Inserra, Lorenzo Bongiovanni (Amplifyfi)

Panel discussion

Agenda: Wednesday Afternoon -- Placements

3.00-3.30: student posters

3.30-4.00: Project descriptions

4.00-5.00: one on one discussions of placements

Agenda: Wednesday Afternoon -- Teaching

Overview of teaching and other activities for 2019 cohort

Agenda: Thursday

Thursday 26 th September			
From 7am	Breakfast in the Grey restaurant		
	DI CDT Industry partners	AIMLAC CDT Industry partners	All students
9.00am	External Advisory Board in the Roath suite Industry and CDT Management meeting in the Bute room		Sophia Suite: 9.00am -Dr Emma Tonkin - Data Ethics 9.45 – Prof Reyer Zwiggerlaar- AI Ethics

Note: Cardiff Physics/Astronomy has student induction on Thursday.
 Please attend CDT event – we have agreed for you to catch up on local induction later

	DI CDT Industry partners	AIMLAC CDT Industry partners	All students
10.30am	Industry departure		Team Building off-site
Lunch in the meeting lounge 1.30-2.00pm			
	All students		
2:00pm	Equality, Diversity and Inclusion in the Sophia suite		
	1st year students	2nd and 3rd year students	
2.30pm	Coding Challenge in the Sophia suite	Feedback session in the Roath suite and departure	
3.30pm	Tea and coffee break		
3.45pm	Coding Challenge continued		
4.30pm	Feedback and future CDT events		
4:45pm close and departure			

Wrap-up

- Any other business
- cdt-aimlac.org
- cdt-aimlac@swansea.c.uk
- @AimlaCcommunity
- data-intensive-cdt.ac.uk
- data-intensive-cdt@cardiff.ac.uk
- @CardiffUniDII

CDT management teams:

Rhian Melita Morris
AIMLAC CDT manager

Roz Toft
AIMLAC Research support officer

Rosemary Granger
DI CDT manager