NEWCASTLE DATA 2020

2020 WAS A YEAR LIKE NO OTHER...

Professor Paul Watson
Director, Newcastle Data
Newcastle University

Looking back at notes I made in meetings earlier in the year, what strikes me now is how quickly our lives changed - even in late February we were planning face to face events for March and April without any apparent concern that they might have to be postponed or moved online. For anyone wanting a historical perspective on the pandemic, I can recommend Laura Spinney's excellent book on the "Spanish" Flu of 1918; it turns out that arguments over masks, controversies over limits on gatherings, travel bans, and even football matches held behind closed doors are not new to COVID-19.

In response to the impact of the virus, Newcastle Data has not only moved its planned activities on-line, but has also added new ones to support the many in our community who pivoted their research onto mitigating the impact of COVID on society and the economy. The pandemic has highlighted how critical data now is to our health and wellbeing, and it was impressive to see the Newcastle Data community stepping up to apply their rare skills to assisting organisations drowning in COVID-related data, but needing our help to extract value from it - those we worked with included the NHS, charities, Newcastle City Council, and the Government.

There have also been examples of turning the fact that everyone is working from home into a virtue. We hosted over 40 events in 2020, attracting over 3000 attendees We signed-up eminent speakers from around the world and increased our audience reach. Pre-pandemic we would have been limited to local speakers or those who could travel to Newcastle.

Over the year we have continued to move forward on Newcastle Data's three goals:

- using data to transform research across the University
- training the next generation of leaders in data science
- exchanging data expertise with those outside the university, for the benefit of society, the economy and our own research and teaching

We believe that the key to success is building communities that combine diverse disciplines and perspectives. In the rest of this report you'll find many examples of this, showing that whether we are tackling a global pandemic or making a step-change in our research, no one discipline has all the answers.

DATA NUCORE

At the heart of our research is the Newcastle University Centre of Research Excellence in Data (Data NUCoRE).

There are seven themes driving forward research in key areas:

Biomedical informatics	Developing new computational and statistical approaches to the converging fields of bioinformatics and health informatics to consolidate and grow research strengths.
Data for manufacturing	Collaborating across disciplines and with industrial partners to develop new AI to support the manufacturing sector. Led by Professor Nick Wright (Engineering)
Data visualization	Setting the national agenda for data visualization and building the latest advanced visualization facilities to link digital data to human thinking. Led by Professor Nick Holliman (Computing)
Spatial analytics	Working across disciplines to integrate and analyse spatial data to address complex societal challenges in areas such as health, mobility, education and economic development. Led by Professor Rachel Franklin (Geography, Politics & Sociology)
Streaming data modelling	Working together to bring about major changes to computational infrastructure, statistical methodology, and algorithms in order to extract value from big data. Led by Professor Darren Wilkinson (Maths, Stats & Physics)
Text in a digital age	Transforming the way we work together and developing new tools to answer questions about how we understand text in the digital future. Led by Professor Jenny Richards (English Literature, Language and Linguistics)
Regulation of data driven technologies	Growing research in the regulation of the digital society and data driven innovation (including data, algorithms, machine learning and AI, ethics, synthetic biology and genetic engineering) through law, code and ethics. Led by Professor Lilian Edwards (Law)

RESEARCH

Our expertise is transforming research across the University by driving forward the development and application of new methods for extracting value from data.

We build multi-disciplinary communities that foster the exchange of knowledge and ideas, and combine our research strengths to address real world problems through externally funded research projects. Throughout 2020 the theme leaders have been focused on building communities and exchanging knowledge. We hosted 28 events across the themes, drawing over 1,500 attendees. These have included:

Spatial Analytics Data
Collaborating with Bristol University and the Turing, Prof Rachel Franklin has curated an extensive virtual seminar series centred around Spatial Analytics Data. This series has attracted hundreds of attendees from across the world. Initially scheduled as a winter series, Rachel is now planning for a spring and summer series due its popularity.

Text and Image: Exploring Cognition
A short series of provocations and discussions on the influence of different types of media on our understanding of ideas and the decisions we make, seeking to address the limits of human perceptions and cognition, how different modes of representation informs perceptions and how we might create new ways of thinking through digital design.This series combined our themes of Text in a Digital Age and Data Visualisation, working closely with colleagues in the Newcastle University Humanities Research Institute.

Animating Texts at Newcastle University (ATNU)
ATNU is a digital collaboration between scholarly editors based in humanities disciplines and Newcastle Data. The virtual Speaker Series welcomed international speakers to discuss how data science can impact the ways in which readers and users can interact with texts.

IN THE YEAR AHEAD, WE'RE LOOKING TO BUILD ON OUR COLLABORATIVE CULTURE AND DIVERSE NETWORKS TO COMPETE FOR EXTERNAL FUNDING AND TACKLE COMPLEX INDUSTRIAL AND SOCIETAL CHALLENGES.

RESEARCH SOFTWARE ENGINEERING

The Research Software Engineering (RSE) team supports the transformation of research at Newcastle through the application of software engineering best practices.

We bring specialist programming expertise, modern development practices and engineering rigour to academic software. We provide expert software engineering consulting services to world-leading research teams and collaborate with scientists and scholars to build software to meet the next generation of research challenges.

2020 has been an exceptionally challenging year for everyone in the RSE team. COVID-19 has affected project priorities, UK research funding, access to the tools we need for our work and most importantly our ability to see each other and grow as a team.

Despite this, 2020 saw continued growth both in staff and in project partnerships. We recruited three more people to the team and started work on several exciting new research projects. Importantly, the team is now known to a larger number of Newcastle academics, and crucially project support staff, enabling them to use their knowledge and skills to help to shape the academic's ideas earlier in the process. In September, the team became qualified Software Carpentry instructors, allowing us to move into delivering training courses for students and staff across the University. We currently cover introductory courses on the UNIX shell, version control and programming with Python.

32
RESEARCH PROJECTS SUPPORTED BY THE RSE TEAM IN 2020

3
NEW TEAM MEMBERS

£1.1M
OF RESEARCH PROJECTS SUPPORTED BY THE RSE TEAM

RESEARCH

MERLON
Principal Investigator: Dr David Greenwood
Science, Agriculture and Engineering Faculty

As part of the MERLON Horizon 2020 project, Newcastle University researchers were tasked with producing forecasting and scheduling modules for island local energy systems. The project aims to make better use of renewable energy systems and improve reliability to customers by coordinating local resources and enabling the local network to operate as an energy island. This required a scheduling approach which could keep the local energy resources in balance, and forecasting methods to predict the short-term future behaviour of the system. While the researcher had the skills and knowledge to develop these algorithms, the RSE team was able to produce robust implementations or to integrate the algorithms with software modules developed by other partners.

"The project would have been extremely challenging for us without RSE support. the team was very easy to work with, gave clear communication, explaining what they were going to do as well as why it was the best approach.

Next year, we look forward to working with the RSE team again as we deliver an updated version of the modules informed by trial results."

Dr David Greenwood
**Principal Investigator, MERLON**

Human Cell Atlas
Principal Investigator: Prof Muzlifah Haniffa
Faculty of Medical Sciences

Cells are the fundamental units of life, and the human body contains around 37 trillion of them. Most cells contain the same genome, but gene activity varies from cell to cell. To truly understand the genome, we need to understand how it instructs cells to carry out their unique functions in the body. A new global initiative called the Human Cell Atlas is setting out to tackle this challenge, using powerful genomics approaches to define the cell types in the human body and reveal how they behave in health and disease.

ROMETRANS
Principal Investigator: Prof Ian Haynes
Humanities and Social Sciences Faculty

Rome Transformed (ROMETRANS) aims to advance our understanding of Rome and its place in cultural change across the Mediterranean World by mapping political, military, and religious changes to the eastern Caelian from the first to eighth centuries CE. The programme offers multiple gains for archaeologists, historians, topographers, and geographers by documenting both the mundane and monumental elements of the city fabric in chronological, geographical, and ideological relationship to one another.
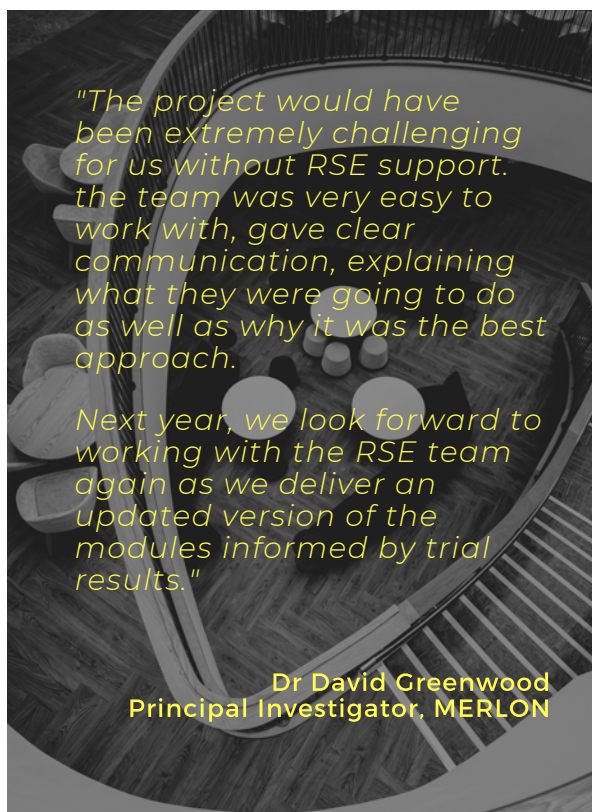
THE TURING

Our fellows have continued to work closely with the Turing's unique network of universities and external partners to develop ambitious programmes and advance research and policy across data science and AI.

Our partnership in The Alan Turing Institute continued to bring benefits despite COVID preventing face to face networking with other partners through most of the year.

Several of our Newcastle Turing Fellows have taken on leadership roles within the Turing. Lilian Edwards (Professor of Law, Innovation and Society) is a Group Leader on the AI Programme, providing expertise on legal aspects of AI, algorithms and the data driven economy, and Darren Wilkinson (Professor of Stochastic Modelling) is a strategic theme leader in the Health and Medical Sciences Programme – a major programme of research using innovations in AI and statistical science

to accelerate the scientific understanding of human disease Meanwhile, the Turing Fellows continue to make highly valuable contributions to the Turing network and their fields of research, developing new methods and applying them in innovative ways. Prof Chris Oates was author of one of the Most Highly Cited Paper in the journal SIAM Review, Bayesian Probabilistic Numerical Methods - this paper provides foundations for numerical methods that are used as part of the data-centric engineering programme at The Turing, where Chris leads a group of researchers.

Before lockdown, we were able to run a two-day Data Science Perspectives workshop for PhD students who travelled to Newcastle from across the North of England. This brought together students from across computing, mathematics, engineering and social sciences to share their research and form networks beyond their institutions.

For many, this was their first presentation on their research, and we were pleased to have created a friendly and constructive environment to build their confidence. As part of the workshop, we welcomed Dr Malvika Sharan from the Turing Institute to introduce the students to The Turing Way - an open-source community-driven guide to reproducible, ethical, inclusive and collaborative data science.

NATIONAL INNOVATION CENTRE FOR DATA

The National Innovation Centre for Data (NICD) addresses the gap in the availability of data science skills by providing practical expertise to organisations struggling with data. We enable them to explore their business questions in creative and entrepreneurial ways.

The Catalyst is home to the National Innovation Centre for Data. We opened it to the public in January 2020, with a full programme of events and a fantastic group of data-focussed tenants, running their businesses out of the building. The Catalyst met all of our ambitions for a collaborative space that can facilitate NICD's events, activities and engagement style.

In early March we had exhibitions in the Atrium on Building Innovation; meetings in the Boardroom held by external organisations including Newcastle City Council, Newcastle Building Society and NHS Business Services Authority. Our workshop space was used to run events by the North East LEP, the Digital Union and Blocknorth Block Party; and we had a number of NICD projects running out of the collaborative space. We also held a PhD and MSc module on data science, enabling students to interact with industry, using the workshops to problem solve and the TED theatre to feedback and communicate their ideas.

We're really excited about the innovative projects we're running with many forward thinking companies. We have private sector projects across Energy, Housing, Banking and Pharma and beyond and public sector projects in health and in the emergency services.

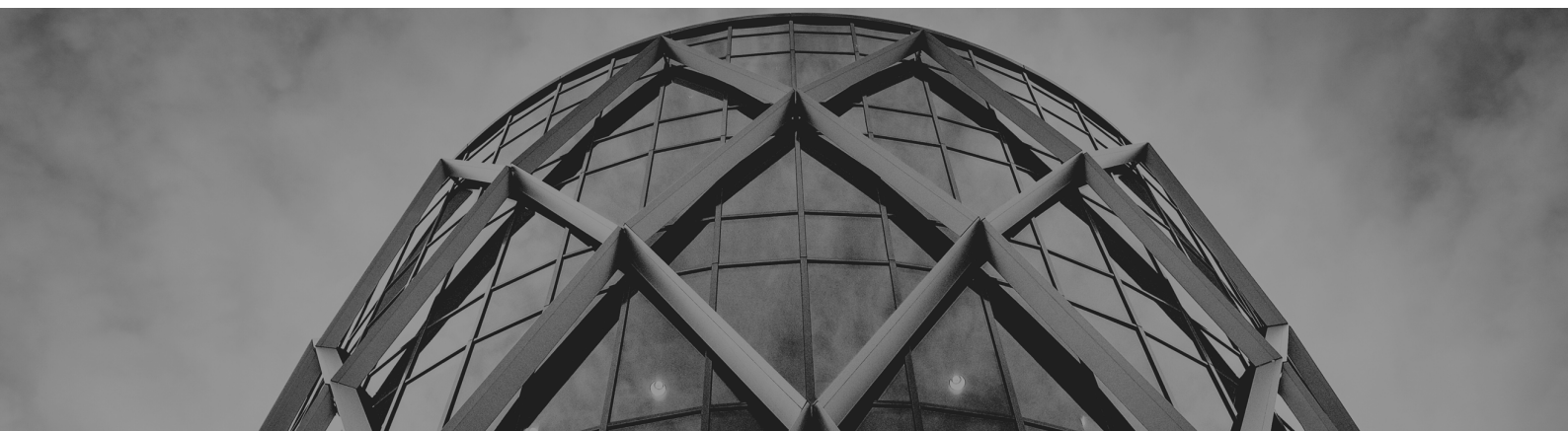| **500**	**120**	**£2M**	**7**
ENGAGED ORGANISATIONS AND 5000 PARTICIPANTS	DATA DRIVEN COLLABORATIVE PROJECTS AND DEEP DIVE DISCOVERY WORKSHOPS	NEW EXTERNAL GRANT FUNDING	CATALYST CONSTRUCTION AWARDS INC. NE BUILDING OF THE YEAR

ENGAGEMENT

All these projects are breaking new ground, using data to change decision making and service provision. Working with the Arrow project we are directly addressing SME needs. Recent successes include working with Nebula Labs to introduce machine learning techniques to better understand travel patterns – enabling a new skill set and a richer solution to a client problem.

As well as running client facing skills projects we are now engaged on applied research projects exploring some ground breaking areas, particularly around machine learning and machine vision. This represents an interesting challenge and a great learning curve for our data scientists to explore the state of the art in these areas.

We continue to market the Data Science MSc and CDT PhD programs with our clients both as upskilling opportunities, and for placements and internships. We continue to host the CDT and MSc Professional Skills module for Data Science students with over 50 students working on industrial problems set by AkzoNobel to better understand business models.

Like everyone else, when COVID-19 hit, we had to transform our delivery method and we volunteered our skills and expertise for projects requiring urgent delivery.

We partnered with the Urban Observatory and Newcastle Data's Research Software Engineer team on a project for Newcastle City Council and NE1 to inform citizens and make them feel safer when they come into the city centre; partnered with Newcastle upon Tyne Hospital Trust and Newcastle University to prioritise care within emergency and intensive care units for coronavirus patients; worked with the West End Food Bank to understand how their clients interact with their food parcel service; and brought together the region's data community to gather useful health, economic and social data.

Despite the effects of COVID-19, we hosted around twenty events throughout 2020 reaching an audience of approximately 1,300 people from across the region and beyond. Most of these events were delivered remotely and successfully contributed towards building our data ecosystem.

These included: the North East Response to COVID-19 which highlighted key projects addressing COVID-19 issues and identified opportunities for collaboration within the region; DataJam North East Event Series which spanned five separate events focusing on the key issues surrounding Health, Environment, Mobility, Inclusivity and Smart Living; the Regional Coordination Group Data Sharing Workshop which brought together key figures to identify data that could aid recovery following the effects of COVID-19; and the National Data Strategy Roundtable hosted with DCMS and the North of Tyne Combined Authority.

ENGAGEMENT

We have worked hard to offer a real contribution to regional strategies, plans and policies. We are becoming an important asset to support the North East economy.

NICD has been cited in:

- *Digital Narrative* - Newcastle City Council
- *Newcastle's System of Systems* - Newcastle City Futures and Newcastle University
- *Data for Growth* - North East Innovation Board Digital Sub-Group,
- *Growing the Digital Technology Sector* - Newcastle City Council/Gateshead Council
- *Strategic Economic Plan* - North East LEP
- *Dynamics of Data Science Skills* - Royal Society
- *Digital Strategy* - North of Tyne Combined Authority
- *Local Industrial Strategy* - North East LEP
- *UK Tech for a Changing World* - Tech Nation Report 2020

We are delivering value back to the nation, the UK Government and to our local region through the North of Tyne Combined Authority.

"The National Innovation Centre for Data continues to play an important role in delivering on the Governments ambitions to unlock the value of data."

The Rt Hon John Whittingdale MP, Minister of State for Media and Data, DCMS

"We know Data is going to be one of the things that transforms our world. That's why the North of Tyne Combined Authority is investing £1.6 million, to make sure that we can use NICD to shape a new world and how we use data to drive innovation, productivity, jobs and prosperity."

Cllr Nick Forbes
Leader
Newcastle City Council

THE CATALYST

The National Innovation Centre for Data and the Research Software Engineering team are located in The Catalyst on the new £350 million Helix development, a ground-breaking 24-acre hybrid city quartet in the centre of Newcastle that brings together businesses, academia, the local community, technology and science..

The Catalyst is designed to encourage collaborative innovation. It includes a visualisation suite, flexible workspaces, an event space and a TED-style lecture theatre. Our facilities are designed to create a supportive ecosystem for businesses to grow and develop new ideas, products and services, alongside world-leading thinkers.

Here, Newcastle Data staff and students will come together with industry and the public sector to solve societal and industrial challenges. Some of the companies we work,with, like Red Hat, have become permanent residents in The Catalyst, while others occupy the flexible spaces when they are working on collaborative projects with us.

BUILDING PROJECT OF THE YEAR

Constructing Excellence North East Awards 2020

EXCELLENCE IN ARCHITECTURAL TECHNOLOGY

AT AWARDS 2020

CHAIR'S AWARD

RTPI North East Awards for Planning Excellence 2020

DESIGN EXCELLENCE AWARD
North East Property Awards 2020

OUTSTANDING BREEAM RATING

First commercial office building in North East to be awarded this certification.

WINNER SUSTAINABILITY

Constructing Excellence North East Awards 2020

WINNER INTEGRATION AND COLLABORATIVE WORKING

CENE Awards 2020

CENTRE FOR DOCTORAL TRAINING
in Cloud Computing for Big Data

The Centre for Doctoral Training (CDT) in Cloud Computing for Big Data has been at the forefront in fostering new types of interdisciplinary. Originally funded by the EPSRC, the CDT continues to be supported through investment from our industrial partners and Newcastle University.

We wanted to help tackle the enormous skills shortage in data analytics in the UK and designed a programme that would produce future leaders who are able to solve real, data driven problems by combining their skills and deep technical knowledge in both Computing and Statistics.

The Centre for Doctoral Training in Cloud Computing for Big Data was first established in 2014, and since then it has recruited seven excellent cohorts of students. Our graduates have gone on to exciting careers in industry (including technical roles at Red Hat, Ubisoft, Huawei, and DWP), academia, and the National Innovation Centre for Data.

In September 2019 we recruited our sixth cohort of students into the CDT. All students have performed exceptionally well during the taught component of the programme (October 2019 to May 2020), and quickly adapted to remote working when COVID-19 forced us to move to home working. During this time, the students completed two very successful group projects working entirely remotely:

One group, working with the NHS Business Service Authority as a client, conducted a study to determine the potential effects of air pollution on life expectancy and identified a relationship that links an increase in mean, ambient benzene levels with a decrease in life expectancy. The students were invited to present their findings to senior members of the NHS Business Service Authority's Data Science team.

The other group, working with ADLink, developed a machine-learning model to predict the inference time of various edge devices for different object detection models. Following the initial project, ADlink are now co-supervising a PhD project in this area.

Meanwhile, the CDT students in the research phase of the programme have continued to excel, despite the challenging circumstances. Ten vivas took place in 2020 (eight of which were conducted remotely), We're pleased that nearly all our students have chosen to stay in the North East and are helping to drive forward the local economy.
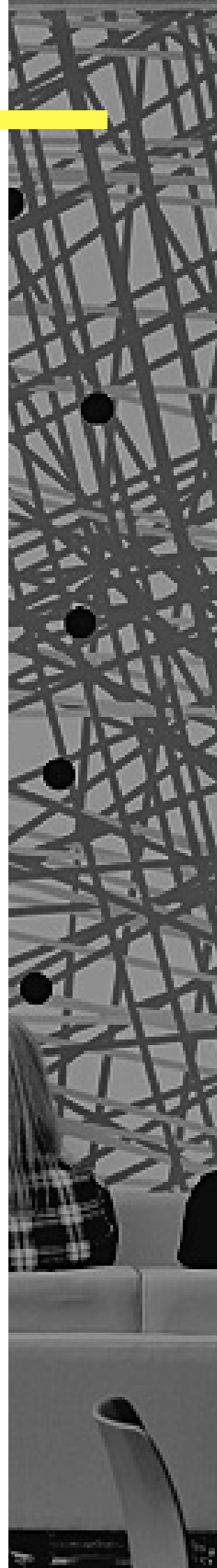
TRAINING

64
FUTURE LEADERS
IN DATA SCIENCE

23
INDUSTRIAL
PLACEMENTS

20
GRADUATES IN
DATA SCIENCE ROLES

50
SUPERVISORS ACROSS
THREE FACULTIES

Most of our students elect to undertake a placement during their PhD. Like all our activities, these had to be conducted remotely. Fortunately, this did not deter the students and we saw a number of successful placements including:
.

- **Coconut Lizard** – Dale Whinham worked at as programmer, porting an in-house memory diagnostics tool used for tracking down performance issues in Unreal Engine to another console platform.

- **Jumping Rivers** – Isaac Matthews built a system to collect data, generate reports on contentsecurity policies and developed an app to explore and visualise the data gathered.

- **Ansys** – Kostas Georgopoulos is working to Ansys identify technical areas of interest in the still emerging field of Quantum Computing.

In addition to this, the students continue to share their research findings at high profile international conferences including CVPR2020, Neural Information Processing Systems (NeurIPS), International Conference on Machine Learning and Quantum Computing Theory in Practice Conference.

We anticipate the Data Science MSc will provide an excellent talent pipeline of students with a strong foundational knowledge in both Computing and Statistics. To retain this talent at Newcastle, we'll be creating a cohort of students who are ready to start their PhD research in September 2021.

"There's a strong connection between the programme and industry, and it definitely helps you to meet key people in these areas. For me, I can't imagine a better experience, from my supervisors, the programme itself, and AkzoNobel."

- Antonia Kontaratou
CDT Cohort 2
NICD Software Specialist

DATA SCIENCE MSC

Newcastle University's Data Science MSc embody the University's aims of 'Education for Life', providing a programme which inspires, challenges and supports a diverse set of students. In 2020/21 we are equipping 50 students with the technical skills, commercial awareness, and entrepreneurial outlook necessary to be successful in the labour market. We are positioning ourselves as a vital source of data science training, with applications increasing by a factor of 20 since the programme launched in 2018.

At the heart of the programme is a successful collaboration between Newcastle University, NICD and AkzoNobel, on a data driven innovation curriculum. We equip students with commercial awareness around the use of data and AI, through a 10-day incubator where students are immersed in an industry setting, working with stakeholders on real business problems.

In addition to developing the Data Science programmes, Matt Forshaw was appointed as National Skills Lead at The Alan Turing Institute.

Matt is passionate about inspiring, educating and upskilling data talent across education and industry. He has made it a priority to widen participation and democratise access to data science training nationally. To strive towards this, Newcastle received funding from DCMS and the Office for AI (via Office for Students) for a project to widen participation in Data Science and AI. This includes 45 MSc scholarships for historically underrepresented groups in the field, with focus on female, black and registered disabled students, students from POLAR Q1 and Q2, care leavers, estranged students, Gypsy/Roma/traveller students, refugees, children from military families, veterans and partners of military personnel.

We are also running pre-sessional summer schools to attract far-and non-STEM applicants and have established an industry mentor network for scholarship holders to support them to build confidence in their abilities, and to grow their professional networks.

20
INDUSTRY-LED PROJECTS

45
SCHOLARSHIPS FOR WIDENING PARTICIPATION

99
STUDENTS SINCE 2018

"Combining technical knowledge with business acumen and the ability to communicate effectively will create tomorrow's leaders in data science.

That's what this programme sets out to deliver."

- Richie Ramsden,
Head of Data Science, Innovation Incubator,
AkzoNobel

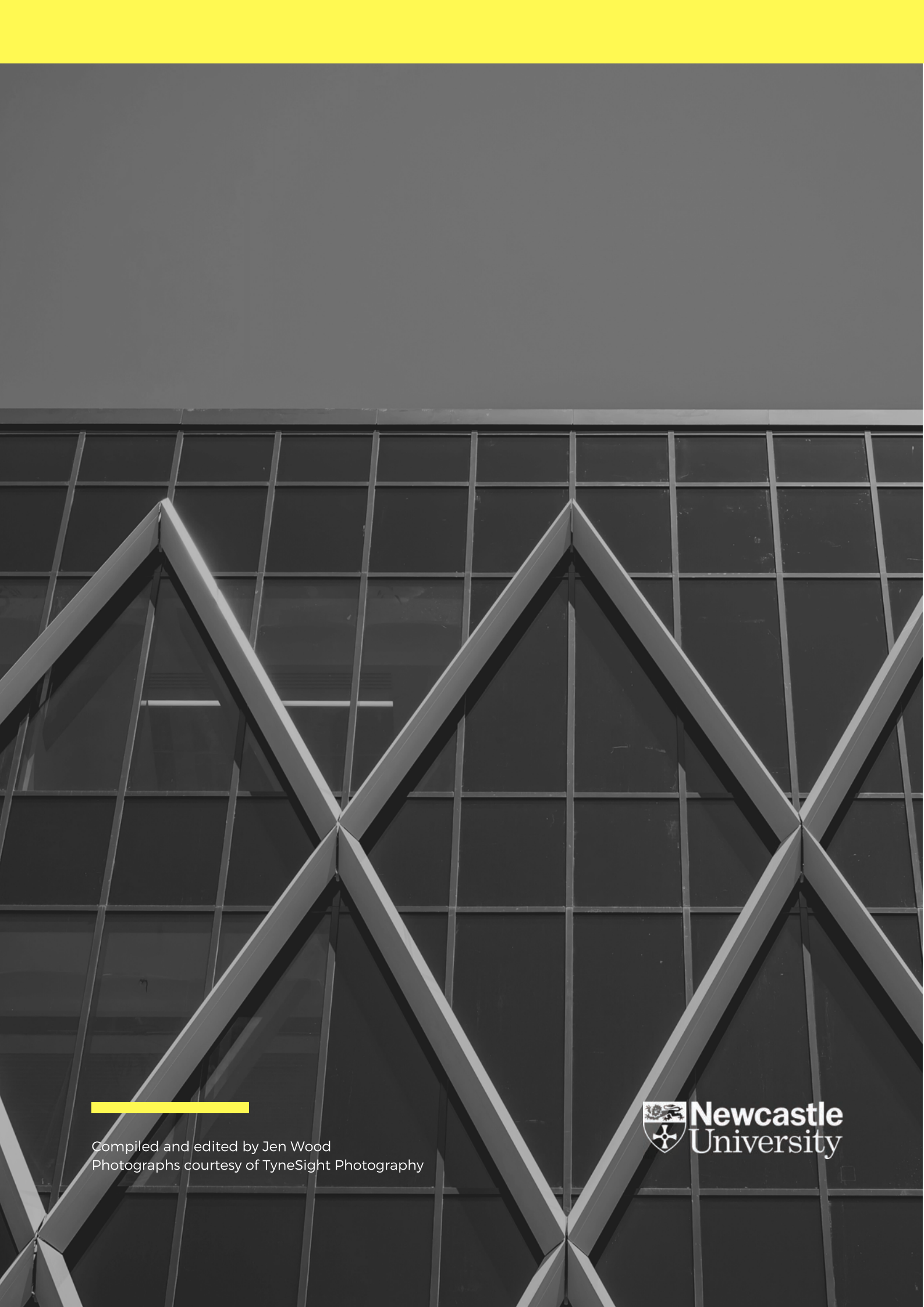"The MSc Data Science course was a hugely challenging and rewarding experience. The course covered the broad range of underlying theoretical subject areas required in order to gain a solid understanding of key data science concepts as well as providing experience of the practical, real-world delivery of data science solutions.

Alongside the taught elements of the course, there were several opportunities to hear from and interact with businesses operating at the forefront of data science commercially, giving a real sense of potential career pathways. Even though I came to the course with a non-technical background, I was supported in developing both the mathematical and computer science skills required to bring data science to life in an organisational context"

- Graham Cole
Data Science MSc graduate and NICD Data Scientist

Newcastle University