



UNIVERSITY OF LEEDS



LEEDS *Institute for
Data Analytics*

**Strategic
Plan**
2020

Leeds Institute for Data Analytics

A vision for data science and data analytics
at the University of Leeds: A strategic proposal

Index

1	Vision	3
2	Mission	4
3	Background	6
3.1	Context	8
3.2	Inception	9
3.3	Multidisciplinary and interdisciplinary	9
3.4	Achievements	9
3.5	Data management infrastructure	11
4	A Vision for Data Science and Analytics at the University of Leeds – The Strategy	12
4.1	Aim and objectives	14
4.2	Data analytics	15
4.3	Training the data scientists of tomorrow	16
4.4	Stakeholders and beneficiaries	17
4.5	How we will work	19
5	Our Enabling Capabilities	22
5.1	Research IT platform	24
5.2	Research Software Engineering Team	24
5.3	Information Governance	25
5.4	Research Management Process	25
5.5	Communications	28
6	Financial Model	29
7	Summary	29
Appendix 1	Leeds Data Platform	31
Appendix 2	Summary of Portfolio Funding	32
Appendix 3	Stakeholder Engagement and Contributorship	34

The quest for the
Leeds Institute
for Data Analytics
(LIDA) is to be
at the forefront
of advanced data
driven research for
societal benefit.

Vision

To advance knowledge
and improve well-being
and prosperity through
internationally outstanding
research and education in
interdisciplinary data science
and artificial intelligence.

Mission

To generate societal impact through the excellence of our research and education in data science with real world applications, underpinned by state-of-the-art developments in data analytics and artificial intelligence (AI).

The quest for the Leeds Institute for Data Analytics (LIDA) is to be at the forefront of advanced data driven research for societal benefit. This will be achieved within an environment where data science and analytics innovation can flourish, and where data scientists are equipped and motivated to undertake data science and analytics research for public good.

The purpose of LIDA is to leverage and develop local, national, and international talent to enable high impact data science research.

LIDA advocates data analytics as the development and application of a bundle of scientific technologies which create applied value and impact from data including statistics, informatics, machine learning (ML), and AI.

LIDA will serve as the ‘go to’ Institute of expertise and infrastructure for data science and data analytics.

LIDA will have a state-of-the-art agile information technology (IT) infrastructure with a wrap-around Research Management Process that upholds the highest levels of information governance. This will be accessible and responsive to LIDA researchers and external stakeholders.

LIDA will continue to grow a multidisciplinary community which promotes synergy through interdisciplinary collaboration with world-class researchers from every Faculty.

LIDA will have a wide range of beneficiaries including, but not limited to, the University of Leeds, the public, academia, data custodians, funders, public services, Government, the National Health Service (NHS), researchers, policy makers and industry.

LIDA will continue to grow a multidisciplinary community which promotes synergy through interdisciplinary collaboration with

World-class researchers

The purpose of LIDA is to leverage and develop local, national, and international talent to enable high impact data science research.

20,000

square feet on Level 11 of the Worsley Building occupied by LIDA

April 2014

LIDA was established with Faculty Management Group approval

£12m

from the union of the Medical Research Council (MRC) Medical Bioinformatics Centre and the ESRC Consumer Data Research Centre (CDRC)

3.1 Context	8
3.2 Inception	9
3.3 Multidisciplinary and interdisciplinary	9
3.4 Achievements	9
3.5 Data management infrastructure	11

There is increasingly compelling evidence that data analytics, including AI and data science, will be foremost amongst the major drivers of research advances, and social and economic growth in the coming decades.

3.1 Context

There is increasingly compelling evidence that data analytics, and in particular AI, will be foremost amongst the major drivers of research advances, and social and economic growth in the coming decades. Data volumes are continuing to explode towards exabyte computing, while new scientific developments are transforming the landscape, for example in precision medicine, transportation (connected autonomous vehicles, mobility as a service), social simulation and immersive technology. Moreover, AI and data science is specifically recognised as one of the four pillars in the UK Industrial Strategy. The Alan Turing Institute will be a key delivery mechanism for the national strategy in AI and data, with the University of Leeds as one of its core university partners.

The significance of data and advanced data analytics for addressing major societal challenges is well recognised. This is of escalating importance across multiple commercial, public and healthcare sectors, exemplified by substantial financial investments from UK Research and Innovation (UKRI), research charities and industry in data science, as well as being a prominent component of their strategic delivery plans. That is, data-driven research and innovation is a central component of the current funding landscape for Higher Education Institutions (HEIs), and a significant proportion (£343 million, 20%) of the investment via Waves 1 and 2 of the UK Industry Strategy Challenge Fund has been allocated to support data-driven research and innovation, with further investment planned through Wave 3.

It is anticipated that innovations in new therapies, improved understanding of the determinants of health and wellbeing (and targeted treatment of diseases), increased commercial productivity, economic growth, and the transformation of the way in which public and private organisations do business will be driven by data and how it is analysed. This ‘technology-led change’ has been termed the fourth industrial revolution and is an opportunity to help the broader society harness converging technologies in order to create an inclusive, human-centred future.

Interdisciplinary teams that combine computer science, mathematics, statistics, health services research and domain expertise such as that in geography, engineering, environment and medicine, among others, are essential to understand, curate and exploit data for research. Indeed, several national and international initiatives such as the Alan Turing Institute, the UK Economic and Social Research Council (ESRC) Big Data Network, Health Data Research UK (HDIR), the Wellcome Trust Programme for Data for Science and Health, the British Heart Foundation (BHF) Data Science Centre, and NHSX have prioritised data science and digital transformation to expedite the use of data for societal benefit.

Undoubtedly, the UK has some of the richest health and societal data assets in the world. Whilst harnessing these data is a priority area, a clear unmet need is the necessary bespoke infrastructure, data science research capability, the skilled work force to study and innovate with these data, as well as the necessary supporting eco-systems for data governance, ethics and data ownership. The University of Leeds has clearly articulated this opportunity, and LIDA offers the potential to deliver this as a flagship for the University, HEIs, and the UK.

3.2 Inception

LIDA was established with Faculty Management Group approval on 1st April 2014, initially from the union of the Medical Research Council (MRC) Medical Bioinformatics Centre and the ESRC Consumer Data Research Centre (CDRC) to the amount of £12M. Its key aims were, within a multidisciplinary environment, to develop data analytics as a transformative technology and play a crucial role at the intersection between research, education and effective partnership working with a wide range of industry partners.

3.3 Multidisciplinary and interdisciplinary

LIDA brings together methodologists and applied researchers in data science and data analytics from a broad range of disciplinary backgrounds, all of whom are working with data – developing and exploiting the latest technologies to generate advances in human and societal health and wellbeing. LIDA also serves as a physical and organisational environment that allows researchers to draw on knowledge and expertise from the many different disciplines across the University of Leeds that are concerned with data, data science and data analytics. Exemplars include the UKRI Centre for Doctoral Training in AI for Medical Diagnosis and Care, and the University’s Economic and Social Research Council (ESRC) Timescapes Archive for ‘big qualitative’ analytics and the Inter-Disciplinary Ethics Applied (IDEA) Centre.

To facilitate this, LIDA occupies about 20,000 square feet on Level 11 of the Worsley Building, University of Leeds. This space includes a range of purposefully designed office and meeting rooms, a visualisation suite and a number of safe rooms for access to controlled datasets.

3.4 Achievements

LIDA currently supports a large and productive research and innovation portfolio. This includes 45 individual centres, programmes and projects, with a combined value to the University of Leeds of £60.5m. This value has increased by £23m in the last 12 months, with the following high value awards secured:

£60.5m *

combined value to the University of Leeds

*as referenced on pp 32-33

- i) UKRI Centre for Doctoral Training in AI for Medical Diagnosis and Care (£5.7m, PI: Prof. David Hogg);
- ii) Future Fashion Factory (£5.5m, PI: Prof. Steve Russell);
- iii) Alan Turing Institute strategic priority funding (£1.5m), seed corn projects (£1.3m) and intern Fellowships (£0.2m);
- iv) Northern Pathology Imaging Collaborative (£1.4m, PI: Dr Darren Treanor);
- v) Consumer Data Research Centre (CDRC) phase 2 (£1.3m, PI: Prof. Mark Birkin);
- vi) Timescapes Archive for the ESRC Research Methods Training Centre (PI Dr Kahryn Hughes);
- vii) HDR UK National Cancer Digital Innovation Hub, DATA-CAN (£4.5m, PI: Prof. Geoff Hall);
- viii) HDR UK North (£3.9m, PI: Prof. Andrew Clegg).

Over the past three years, there has been considerable diversification in terms of funding sources. From an initial reliance on the founding grants from ERSC and Medical Research Council, funds have now been secured from an additional 19 sources, including:

- Arts and Humanities Research Council (£5.5m);
- British Heart Foundation (£1.5m);
- Cancer Research UK (£3.2m);
- EPSRC (£2.1m);
- European Research Council (£1.1m);
- Innovate UK (£1.7m);
- National Institute for Health Research (£5.8m);
- Yorkshire Cancer Research (£6.9m);
- Horizon 2020 (£0.9m).

Figure 1 and 2

A summary of the portfolio funding is presented within Figures 1 and 2 and in detail in Appendix 2.

Figure 1: Grant value to the University of Leeds (£60.5m)

Research Councils: 62%
UK Charities: 23%
UK Government: 12%
EU: 3%

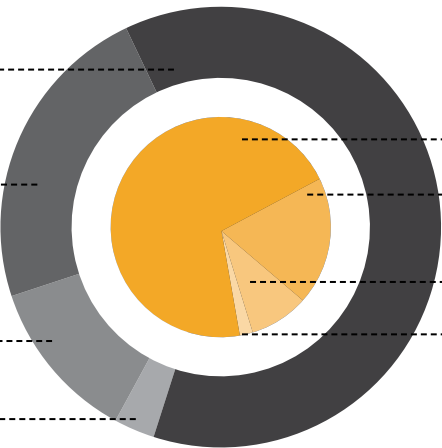


Figure 2: Grant value to LIDA (by funder type) (£38.5m)

Research Councils: 70%
UK Charities: 19%
UK Government: 9%
EU: 2%

100%

Growth target had been achieved by November 2016

Moreover, the Institute’s originally stated strategic ambition was to double its initial grant income over a five year planning horizon by leveraging multidisciplinary projects drawn from across campus and involving a wide range of academic and non-academic external partners.

By November 2016 that 100% growth target had been achieved through projects which have a specific dependence on LIDA providing a combined income of £4.8 million per annum (from an initial baseline of £2.4 million per annum). The grant portfolio for which LIDA is a named collaborator currently has a combined value of approximately £60.3 million (£8.4 million per annum).

Consequently, LIDA has had a substantial positive impact across many of the University’s Faculties. Its success is based on a model of collaboration within the University and with a large network of local, national and international partners including linked data providers, other HEIs, government and commercial organisations.

It is host to 158 LIDA personnel, comprising full time academic staff, PhD students and graduate interns as well as project staff and IT professionals, and its activities include funded research programmes, hosting and controlling access to data, and delivering training, seminars and workshops.

Furthermore, LIDA is a trusted research partner to over 40 government, public and commercial bodies and has formal collaborations and data sharing agreements. Access to LIDA’s secure data services, resources and expertise enables our partners to derive additional value from their data and add value to their internal data analytics capability.

3.5 Data management infrastructure

Historically, LIDA has been supported by two data management infrastructures (IT platforms), the Integrated Research Campus (IRC) and the Secure Electronic Environment for Data (SEED). The IRC is an advanced and secure computational platform, designed, implemented (at LIDA inception) and maintained by University of Leeds IT to provide specialist technical services for data handling, analysis and application within LIDA. The IRC has accredited certification to the international standard for information security management, ISO/IEC 27001:2013 and both the IRC and SEED are NHS Data Security and Protection Toolkit compliant. This allows researchers secure access to sensitive data, and has enabled and encouraged external data owners to work with the University of Leeds. LIDA personnel also have access to the University of Leeds comprehensive Advanced Research Computing (ARC) service which consists of a large High Performance Computing (HPC or Supercomputing) resource. LIDA, has recently successfully completed its ISO 27001 recertification audit for its Information Security Management System (ISMS). The IRC has been approved by external auditors and is compliant with the requirements of ISO 27001: 2013.

A Vision for Data Science and Analytics at the University of Leeds

The Strategy

LIDA will support new groups with ambitious ideas who have a compelling strategy by which to realise their vision

4.1 Aim and objectives	14
4.2 Data analytics	15
4.3 Training the data scientists of tomorrow	16
4.4 Stakeholders and beneficiaries	17
4.5 How we will work	19

Strong and productive relationships with key stakeholders (Section 4.4) are essential to realising our vision. To date, LIDA has a solid foundation with many academic and industry parties, from which we will build and expand.

LIDA will make fundamental data driven discoveries for societal benefit by working with data applications arising from each of the seven University of Leeds' Faculties. This will require harmonised working within the University and beyond, be founded on LIDA's core values and principles (Section 4.5), and delivered by means of a collaborative, interdisciplinary approach.

Research domains that will be of strategic importance to LIDA will include, but not be limited to urban analytics, health data, and AI.

Notwithstanding these strategic research domains, LIDA will realise and promote areas of academic excellence evidenced by: high impact internationally competitive academic outputs; major societal impact; competitive personal awards; successive project, and/or programmatic, and/or infrastructural awards; research and teaching excellence demonstrating impact in data science.

Equally, LIDA will support new groups with ambitious ideas who have a compelling strategy by which to realise their vision, and who ultimately align with LIDA's vision.

LIDA will provide the environment to allow researchers to exchange ideas on data science and data analytics, in particular AI methods and its real world application for societal benefit. LIDA will help develop people who wish to excel in data science and data analytics. LIDA will have and maintain an outstanding infrastructure and estate to offer opportunities for high quality student and researcher, and external collaborator experience.

4.1 Aim and objectives

Academic excellence

in teaching and research in data science and data analytics

4.1.1 Aim

To be a world leading Institute of expertise and infrastructure applying data science and artificial intelligence to yield valuable real world insight from health and societal data.

4.1.2 Objectives

- To maintain strong alignment between the Institute and the University of Leeds’ Strategic Plan for the period 2020–2025.
- To work in partnership with key stakeholders.
- To align with national and international priorities in data science and analytics.
- To provide unique opportunities for students with respect to data science and data analytics.
- To provide academic excellence in teaching and research in data science and data analytics.
- To provide first-class facilities for researchers and external collaborators which meets the highest of standards for governance.
- To invest in people and academic activities.
- To be strategy-led, but ensuring that activities are academically and financially sustainable.

LIDA will bring together the best minds in data science and data analytics and co-work with its stakeholders to ensure:

- Early, yet sustainable ‘return of investment’ with regards to new knowledge and societal benefit from the study of data.
- New insights from data analytics are made widely available at pace and, therefore, scaled for public good
- An environment that will accelerate the pace of growth of the next generation of data scientists.
- That it is at the cutting edge of the development and use of the data science and data analytics methods.
- It abides by the highest standard of research integrity and quality.

4.2 Data analytics

LIDA recognises and values advanced analytics for handling complex and high dimensional data from multidisciplinary research domains, including quantitative, qualitative, structured and unstructured data. Moreover, the application and development of methodology at the interface of statistics, machine learning and AI is an essential part of data analytics.

Notwithstanding the importance of these techniques, LIDA will have a strong focus on AI. There is an unprecedented unmet need in this area, scope for considerable grant funding, a strategic drive from the UK Government and UKRI, and a major opportunity for the University of Leeds to have a foremost institute which majors on AI.

Furthermore, there is substantial interest and huge potential across the University in applying AI to key challenges in science and engineering, humanities, and health. As such, there is a critical need to build a University-wide community and academic pipeline of researchers and educators, both to develop the necessary knowledge and skills in AI, and to exploit new synergies between disciplines that are enabled by AI. Notably, the University of Leeds, School of Computing has been awarded an UKRI Centre for Doctoral Training in AI for Medical Diagnosis and Care, which is hosted in LIDA.

Thus, LIDA will actively seek opportunities to strengthen the University’s foundation in this area through the development and appointment of associated postdoctoral, mid-career grade and senior researchers in this discipline.

Accelerate

the pace of growth of the next generation of data scientists

Professor Christopher Gale, LIDA Co-Director and Chair of Cardiovascular Medicine at the Leeds Institute of Cardiovascular and Metabolic Medicine.



4.3 Training the data scientists of tomorrow

40

PhD students are hosted by LIDA

There is currently a national imperative to create an innovative data environment in the UK that will sustain the excellence of the research base, increase productivity and transform the way in which public and private sector organisations do business. LIDA will facilitate, in partnership with and through the Schools and Faculties at the University of Leeds, the training of the data scientists of tomorrow. LIDA will contribute to the training of professionals from a diverse range of backgrounds and abilities to include those from the healthcare, industry and public sectors in data science and data analytics. In the health data domain, the University of Leeds has contributed to the Topol Review (<https://topol.hee.nhs.uk>), which highlights the need for greater digital education in the undergraduate health curricula as well as upskilling the existing workforce, and training tomorrow’s clinical academics and allied healthcare professionals in this domain.

LIDA already delivers a range of educational activities including the Seminar Series, Training Workshops, Summer School, affiliated MSc programmes, and hosts more than 40 PhD students. Notably, the LIDA-centred Leeds Data Science Society is one of the most popular of student-run societies at the University and a focus for much student activity concerning data science. Presently, LIDA’s teaching and educational activities are integrated within many of the University’s initiatives including the Alan Turing Institute at Leeds, the Consumer Data Research Centre, the Wolfson Centre and University Academic Fellowships. Notably, LIDA has received funding from HDR UK for the delivery of a Masters in Data Science (£600k over 3 years).

LIDA will work in partnership with the Schools and Faculties to manage and deliver a portfolio of affiliated MSc programmes, which already embraces ten individual courses with more than 600 students. Opportunities will be actively sought to build new programmes which align to LIDA’s domain priorities in Urban Analytics, Digital Health and AI. Harmonisation of content across all programmes will be pursued to maximise the quality, effectiveness, and impact of our education, while being led by both research excellence and applied relevance. Acting as a champion for data analytics across the institution, LIDA will actively promote the expansion of capacity to meet these objectives. LIDA will look to exploit the potential for adding value through programme management and the creation and delivery of its own content, offering more flexible learning and to a wider market (undergraduate and continual professional development). This will include the advancement of plans for a Postgraduate Credit Route and for cross-faculty undergraduate content which is accessible to students from any discipline and closely aligned to the University’s status as a QStep Centre for Quantitative Methods. LIDA is exploring opportunities to work with other organisations to train the workforce in data science (for example a trailblazer group in collaboration with the Royal Geographical Society and Geospatial Commission to establish a LIDA Degree Apprenticeship in Geographical Information Science).

4.4 Stakeholders and beneficiaries

£46m

Alan Turing Fellowships under the AI Sector Deal

An essential component of LIDA will be its partnership with key stakeholders. LIDA will build on its existing relationships, and seek new partners. Partnerships will be forged which promote and align with LIDA’s vision.

Academic partnership will not be limited to internal cross-faculty relations; interdisciplinary working will be core business within the University. LIDA will reach out to other HEI’s in the UK and abroad to develop ground-breaking and effective partnerships between universities working in data science and data analytics – creating a culture of ‘team data science’. Having a recognised national footprint which is built upon excellence in our research domains of urban analytics, health data and AI will be fundamental to the development of LIDA’s partnerships and, therefore, its success.

LIDA is strongly placed to capitalise upon four key trends across the current funding landscape, these are:

- An increasing share of funds allocated to data-driven research and innovation encompassing AI, data analytics and data science.
- A greater emphasis on mobilising inter- and multidisciplinary research teams to address real-world ‘Grand Challenges’.
- Innovation through collaboration based on a wide range of academic and non-academic stakeholders working together to achieve mutual benefits.
- An increasing requirement to demonstrate a place-based approach to research and innovation that harnesses regional assets, including: intellectual capital; physical infrastructure; and investment from the public and private sectors.

These trends are evident in the strategies and delivery plans of several major funders, with current or planned funding opportunities including:

- UKRI Industrial Strategy Challenge Fund (£486.8 million);
- UKRI Strategic Priorities Fund (£55.9 million);
- Research England Strength in Places Fund (£31.9 million);
- Alan Turing Fellowships under the AI Sector Deal (£46 million);
- HDR UK Digital Innovation Hubs and Better Care Programme (£39.5 million);
- British Heart Foundation Centre for Data Science (£10 million);

(Values quoted represent stated funding commitments).

Close working with industry will also be important for income generation, knowledge transfer, routes to market and societal benefit. This is an area in which LIDA has yet to capitalise fully, but for example, has good potential for pharmaceutical industry funded research using healthcare data. LIDA will capitalise on the newly awarded HDR UK National Cancer Digital Innovation Hub (DATA-CAN), engage with Innovate UK, directly with industry partners, and with industry through the University of Leeds’ dedicated NEXUS Hub for research and innovation collaborations.

Table 1:
Beneficiaries

Ultimately, the beneficiaries of LIDA’s work will be the public and society at large. Details of anticipated beneficiaries are given in the table below.

Stakeholders	Benefits
University of Leeds	Maintaining a leading reputation in data science and analytics research and teaching Providing a stimulating and enhanced environment for students
Academia	Developing the data science researchers and teachers of the future
Data custodians	Abiding by the highest levels of governance to exploit data for public benefit
Funders	Returning high impact research applicable to society on time and within budget
NHS	Better informed healthcare delivery
Researchers	Access to a purpose built environment for the investigation of and learning about data and data analytics Part of a community of data scientists from a range of disciplines
Policy makers	Providing reliable and robust information needed to drive changes to Government policy, healthcare practice and society
Industry	Rapidly returning new insights from the study of data for commercial productivity and economic growth Providing partners with access to experts in data science
The public	Enhanced knowledge about health, disease, treatments, and society at large transforming the way in which public and private organisations do business

4.5 How we will work

Table 2:
Shared values

LIDA will enjoy a collaborative environment which brings together researchers of international quality from every Faculty of the University, national centres in consumer data research, digital imaging, as well as external partnerships spanning commerce to Government to the third sector. It will host an enviable portfolio of multi-disciplinary post-graduate training courses, and a network of partnerships with other research leaders, including that through the Alan Turing Institute. This environment will allow and encourage data-driven research and innovation to flourish, and ensure data scientists are equipped and motivated to work with data for public good.

LIDA's vision will be centred on shared values (Table 2) and underlying principles (Table 3), and be compatible with the University of Leeds’ values.

Science-based	Driving discovery through data
Partnership-forming	Building strong national and international academic and industrial collaborations that benefit each party
Society-centred	Co-working with and for the public to determine strategic direction and delivery
Team-orientated	Removing people pigeon-holing and discipline specific silos and enabling effective professional development opportunities utilising matrix management approaches which involve staff from both LIDA and Faculties/University Specialist Services

Table 3:
Underlying principles

Be driven by data science and data analytics that impacts on the lives of the public to ensure that we have our roots in the use of data for societal benefit
Be agile in and at the forefront of research IT architecture, therefore being competitive in a fast moving IT environment
Be underpinned by the highest standards of information governance co-created with data scientists, researchers, IT experts, and data managers
Be innovative in ways of extracting and sharing new knowledge from data
Be proactive in developing close partnerships with beneficiaries, so that LIDA is recognised nationally and internationally for excellence in data science and analytics

LIDA will have a ‘Code of Conduct’ that builds a community capable of responding to the needs of its beneficiaries – the University of Leeds, academia, data custodians, funders, public services, the NHS, researchers, students, policy makers, industry and, above all the public (Table 1: Beneficiaries). Whilst LIDA will have an ‘open door’ policy, and welcome activities relating to its vision, it will abide by the highest standards of information governance and require our LIDA personnel and collaborators to do the same.

LIDA’s vision will be upheld by means of synergistic working relationships with several key University of Leeds stakeholders, including:

- The University of Leeds Research IT Professional Service will be integral to the provision and maintenance of our state of the art IT platform, the Research Management Process, operational governance, information governance and professional development of our staff and associates.
- The University of Leeds Research and Innovation Service will host Research and Innovation Development personnel within LIDA to provide dedicated support to researchers in the development of successful applications to national, international, commercial and non-commercial funders.
- The University of Leeds Data Protection Officer, Research Ethics Service, ResearchData@Leeds Service and UseMyData.org will work with LIDA in providing specialist advice to researchers about ethical and legal issues.
- The LIDA Team will work closely with Faculty Research Office staff and Institute Business Managers to support researchers in managing their research grants.

Whilst there are unique benefits to functioning as an Institute which spans the University, inherent to this are challenges. As such, LIDA will modify its present operating model within the University of Leeds in the following ways:

- Designing the operating model, work environment and IT platform to facilitate stronger collaboration that will accelerate the pace of translation of research outputs into demonstrable benefits to society and the professional development of the next generation of data science researchers.
- Investing in senior operational leadership to provide a driver for effective strategic management of all of our activities (Operations Director).
- Re-focussing our staff, associates and collaborators on our shared values and principles.
- Working synergistically with key University of Leeds specialist facilities.
- Ensuring meaningful public involvement in our research design, delivery, strategy development, governance and operating models.
- Actively and regularly communicating with our associates and stakeholders through a variety of media.

Investing

in senior operational leadership to provide a driver for effective strategic management

Professor Mark Birkin, LIDA Co-Director, ESRC Consumer Data Research Centre and Professor of Spatial Analysis & Policy, School of Geography and Fellow of The Alan Turing Institute.



Our Enabling Capabilities

Quality assurance

LIDA will continue to embed quality assurance into all of our activities

Highest standards

of security for data analytics

5.1	Research IT platform	24
5.2	Research Software Engineering Team	24
5.3	Information Governance	25
5.4	Research Management Process	25
5.5	Communications	28

Members of the team will be strategically aligned to the key domains of LIDA research activity thereby enabling them to provide specialist insight for grant applications and scholarly publications.

5.1 Research IT platform

As part of this new phase of the Institute’s development significant investment has been made in a new LIDA Research IT Platform (hereinafter called LASER (Leeds Analytic Secure Environment for Research)) to provide a powerful and resilient IT infrastructure designed to empower researchers and drive excellence in LIDA.

The range of capabilities of this new infrastructure are outlined in Appendix 1. Its distinctive offering is the combination of meeting the highest standards of security for data analytics, of course ensuring ISO27001 and NHS Data Security and Protection Toolkit compliance with the flexibility to enable constant agility in design and function; alongside scalability depending on the researcher need. This platform is supported through a commitment backed up by both LIDA and the University of Leeds IT Service to ensure high quality service wrap around to maximise ease and efficiency of time for researchers.

5.2 Data Analytics Team

LIDA recognises the critical importance of providing and growing its specialist Data Analytics Team, who also have research engineering capabilities. The Data Analytics Team will work alongside and in collaboration with researchers.

This team, located in LIDA, provide expertise in data handling, data wrangling, data curation and quality standards, data linkage, database set up, software development and data visualisation. Members of the team will be strategically aligned to the key domains of LIDA research activity thereby enabling them to provide specialist insight for grant applications and scholarly publications. The Team will also provide support and advice for use of the Leeds Data Platform and will have key responsibilities in operationalising the LIDA Information Governance Framework.

5.3 Information Governance

The provision of first class Information Governance robustly applied to all of the projects, partnerships and collaborations will be a key characteristic of the way in which LIDA operates.

This will be delivered through a combination of our highly trained staff, robust secure systems and process and good research practice, data management and protection. Our work is at all times underpinned by an ethical understanding of the subject, supported by the University of Leeds Inter-Disciplinary Ethics Applied (IDEA) Centre – a national Centre of Excellence in Applied Ethics. LIDA will continue to embed quality assurance into all of our activities, achieving quality management standards and putting in place a robust governance framework to ensure our work continues to be conducted to the highest legal, ethical and quality standards.

5.4 Research Management Process

The refreshed LIDA Research Management Process will span the lifecycle of research projects from pre-award (design, costing and funding application) through to post-award (research conduct, delivery and reporting).

The process has been designed with quality and professionalism in mind, to provide support and added value at all stages for our researchers and collaborators, be efficient and, to this end, will utilise a new automated workflow system to support the process from start to finish. It will be operationalised by the LIDA Team through a controlled suite of documents including a quality manual and standard operating procedures with supporting documentation to guide researchers through the process.

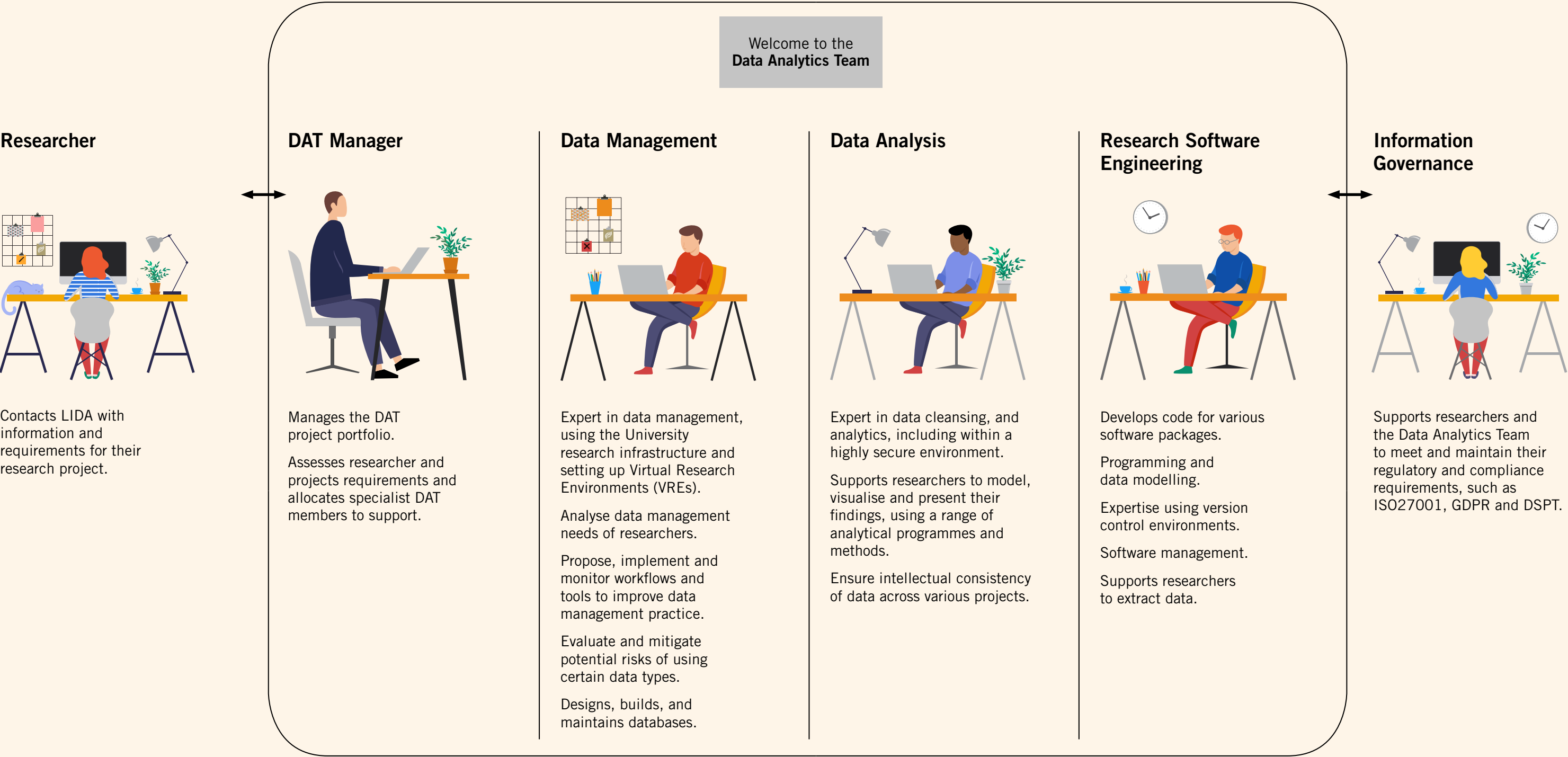
Understand the team

LIDA's Data Analytics Team (DAT) have expertise and experience across different areas of data management, analytics, research software engineering and information governance.

These skills are complemented by their expertise and experience of working within specific domains, including health and urban analytics.

DAT members can provide support to researchers at the grant opportunity stage, right through to project delivery.

If successful the DAT team are assigned to a project based on its individual requirements, as seen in the project example to the left.



5.5 Communications

LIDA recognises that an important part of maintaining reach and engagement is effective communications. LIDA will continue to focus on ensuring it raises the profile of academic research through LIDA, extending our reach to target key leaders, influencers and audiences, drive engagement with our education and training activities and build strong relationships with industry, research councils and partners. We will build on the extensive tools already being used via digital media, a refreshed web presence, multi-channel social media engagement and partner channels such as Turing/ESRC as well as extensive event activity hosted by LIDA and participation in partner events. Our communications activity will be delivered collaboratively with Faculty and University marketing and communication teams; ensuring we are maximising our impact and are efficient with our messaging and use of available media channels.



6

Financial Model

We have created a financial model that incorporates cost recovery from a Faculty contribution model, research grant income, industry partnerships, and education and professional development activities, as well as public, private and philanthropic support. We aim to provide maximum return on investment.

7

Summary

‘Leeds Institute for Data Analytics – A vision for data science and data analytics at the University of Leeds: A strategic proposal’, is a bold and ambitious, but necessary proposal for the University of Leeds which will ensure that the University maintains its position in the rapidly moving field of data science and data analytics. Critically, this is a unique opportunity for the University of Leeds to show HEIs and the world that it is fully committed to its aims to be a leading player in health and societal data and data analytics. LIDA will require strong leadership, a new approach to ‘team science’, a dedicated senior team, infrastructure, committed investment, and the unfaltering support of the University of Leeds. We have set out a clear Vision and associated strategy, and we are confident that following this will propel the University of Leeds into a position of leadership in the field of data science and analytics for societal benefit.

The importance of data and advanced data analytics for addressing major societal challenges is well recognised.



Appendix 1 LASER Platform

A compelling argument for LIDA's success is a state-of-the-art IT platform. The LASER Platform is a powerful and resilient IT infrastructure designed to empower researchers and drive excellence in LIDA. The design allows agile development to be undertaken into the future; a commitment backed up by both LIDA and the University of Leeds IT Service.

The LASER Platform has been designed with and for researchers and includes the following capabilities:

- Fully flexible and scalable to enable researchers to align spend to research requirements.
- Agile and quick to provision, to support a range of research user cases.
- Access to the latest tools and capabilities such as machine learning to support researchers.

The system has been specified and is maintained to ensure the highest standards of data security:

- Resilience levels and backups provided by one of the most cyber-aware and heavily invested companies in the world.
- Regularly patched with the latest software and operating system updates.
- Encryption at rest – all data at rest is encrypted to AES256 Bit encryption.
- Encryption in transit – all data is encrypted using TLS 1.2.
- Residency – all data stored in the UK.
- Access Controls – role based access controls and principles of least privileged access are implemented to ensure only authorised users are able to view data.
- Data is classified according to sensitivity and criticality with robust checks on data input and output to and from the Cloud including a fully maintained asset register.
- Multifactor authentication implemented with the ability to restrict access to specified safe rooms on Campus for data requiring the highest level of security.
- Monitoring and threat detection to provide unified security management with supporting network controls to isolate information in our networks and its supporting information processing facilities.
- ISO27001 compliant.
- NHS Data Security and Protection Toolkit compliant.

Appendix 2
Summary of Portfolio Funding

Project	Location	Total Value to Leeds (Price in KRISTAL) / £	Total LIDA Share %	Funder
Medical Bioinformatics Centre (MBC)	LIDA / LIBACS	£5m+	75%–100%	MRC
UKRI CDT in AI for Medical Diagnosis and Care	Computing / LIDA	£5m+	75%–100%	UKRI
Consumer Data Research Centre (CDRC)	LIDA / Geography	£5m+	75%–100%	ESRC
Yorkshire Lung Screening Trial	LIHS	£5m+	0%–25%	YCR
Future Fashion Factory	Design	£5m+	0%–25%	AHRC
National Colorectal Cancer Intelligence Hub	LIDA / LICAMM	£1m–£5m	75%–100%	Cancer Research UK
DATA-CAN (HDR-UK Cancer Data Hub)	LIMR	£1m–£5m	75%–100%	Health Data Research UK
CDRC Transition Funding	LIDA / Geography	£1m–£5m	75%–100%	ESRC
Data Assimilation for Agent-Based Models: Applications to Civil Emergencies	LIDA / Geography	£1m–£5m	75%–100%	ERC
Towards UK poSt Arthroplasty Follow-up rEcommendation: UK SAFE	Clinical & Translational Rheumatology	£1m–£5m	75%–100%	NIHR
Centre for Doctoral Training in Data Analytics and Society	LIDA / Geography	£1m–£5m	25%–50%	ESRC
PigSustain (Modelling the resilience of the UK pig industry)	Biology	£1m–£5m	0%–25%	BBSRC
The Northern Pathology Imaging Collaborative	Pathology & Tumour Biology	£1m–£5m	0%–25%	Innovate UK
ALABAMA: Allergy, Antibiotics and microbial resistance	Medicine	£1m–£5m	TBC	NIHR
PICANET 5 (Paediatric Intensive Care Audit Network)	Div. Epidemiology and Biostatistics	£0.5m–£1m	75%–100%	Healthcare Quality Improvement Partnership
QuantiCode (Intelligent infrastructure for quantitative, coded longitudinal data)	Computing / LIDA	£0.5m–£1m	75%–100%	EPSRC
Using tumour, peripheral blood and sentinel nodal transcriptomics to understand the interaction between melanomas and the host	LICAP	£0.5m–£1m	75%–100%	MRC
GRACE (Effectiveness of the GRACE risk score compared to standard care)	LIDA / LICAMM	£0.5m–£1m	75%–100%	British Heart Foundation
SIPHER (System-science Informed Public Health Economic Research for Non-communicable Diseases)	Geography / LIDA	£0.5m–£1m	75%–100%	MRC
CPR for Cancer Outcomes	LIDA / LICAP	£0.5m–£1m	75%–100%	Macmillan Cancer Support
Creating a Digital Twin (SPF) SPENSER and other components	LIDA / Geography	£0.5m–£1m	75%–100%	Alan Turing Institute
STOP-ADENOMA (Colorectal cancer prevention study	LIMR	£0.5m–£1m	75%–100%	NIHR
IMforFUTURE (Innovative Training in Methods for Future Data)	Mathematics and Statistics	£0–£0.5m	75%–100%	EU - Horizon 2020
Treatment According to Response in Giant cEll arTeritis (TARGET)	LIRMM	£0.5m–£1m	25%–50%	MRC
Microbubble enhanced imaging and therapeutic delivery	Physics	£0.5m–£1m	25%–50%	EPSRC

As of 31/12/2019

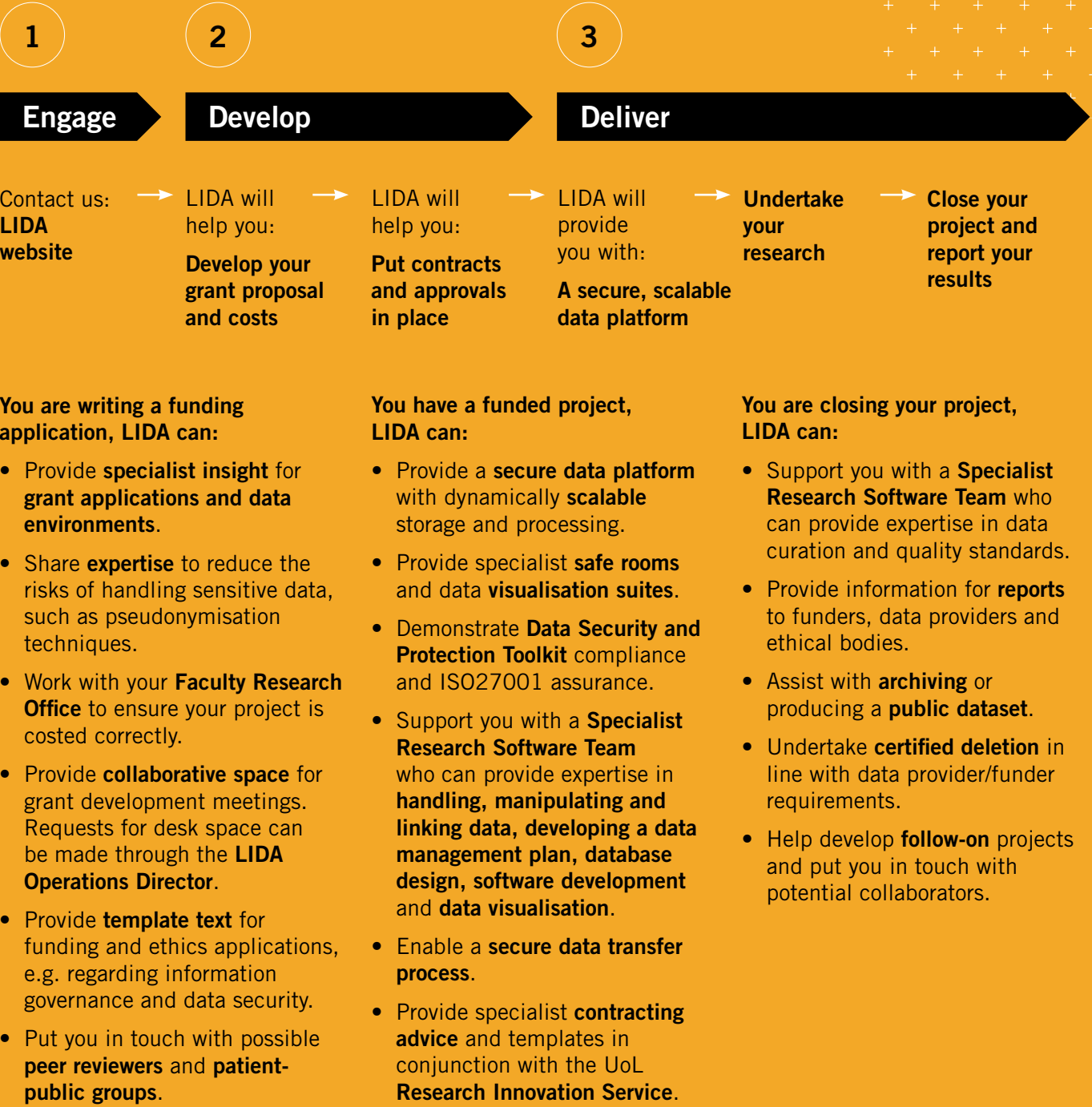
Project	Location	Total Value to Leeds (Price in KRISTAL) / £	Total LIDA Share / £	Funder
QualDash (interactive dashboard delivering healthcare audit information)	Healthcare	£0.5m–£1m	25%–50%	NIHR
BESTMAP (Behavioural, ecological and socio-economic tools for modelling EU agricultural policy)	Geography	£0.5m–£1m	0%–25%	H2020
CDRC Infrastructure Extension	LIDA / Geography	£0–£0.5m	75%–100%	ESRC
Leeds Biomedical Research Centre – Non-surgical treatments for osteoarthritis	LIRMM	£0–£0.5m	75%–100%	NIHR
Leeds Biomedical Research Centre – Drug repurposing and treatment toxicity	LIRMM	£0–£0.5m	75%–100%	NIHR
UKRI Innovation / Rutherford Fund Fellowship – Chris Carrigan	LIDA / LICAP	£0–£0.5m	75%–100%	MRC / HDR UK
MISTRAL (Multi-scale Infrastructure Systems Analytics)	LIDA / Geography	£0–£0.5m	75%–100%	EPSRC
Wellcome Trust Post Doctoral Fellowship	LIDA / LICAMM	£0–£0.5m	75%–100%	Wellcome Trust
Association of health related quality of life	LIMR	£0–£0.5m	75%–100%	British Heart Foundation
Bridging the Social City and the Smart City	LIDA / Geography	£0–£0.5m	75%–100%	Alan Turing Institute / Economic and Social Research Council
BioPASSPoRT (Biomarkers for Patients Assessment and Stratification Post Renal Transplantation)	Oncology and Cancer Research	£0–£0.5m	75%–100%	Innovate UK
Modelling the joint effects of sparse and dense temporal datasets on outcomes	LIDA / Statistics	£0–£0.5m	75%–100%	Alan Turing Institute
HELICAL (HEalth data Linkage for ClinicAL benefit)	LIRMM	£0–£0.5m	75%–100%	Marie Skłodowska-Curie Actions – European Training Networks
Yorkshire Specialist Register of Cancer in Children and Young People: Epidemiological and Applied Health Research	Epidemiology and Biostatistics	£0–£0.5m	75%–100%	Candlelighters
SPENSER (Synthetic Population Estimation and Scenario Projection Model)	LIDA / Geography	£0–£0.5m	75%–100%	Alan Turing Institute
Capturing relationships between individuals: Integrating Causal Inference and Agent Based Modelling	LIDA / Geography	£0–£0.5m	75%–100%	Alan Turing Institute
Understanding and Quantifying Uncertainty in Agent Based Models for Smart Cities	LIDA / Geography	£0–£0.5m	75%–100%	Alan Turing Institute
ActEarly: a City Collaboratory approach to early promotion of good health and wellbeing	LIDA / Psychology	£0–£0.5m	75%–100%	UK Prevention Research Partnership
New visualization methods for profiling datasets and analysis pipelines	LIDA / Computing	£0–£0.5m	75%–100%	Alan Turing Institute
Understanding bowel cancer in people aged less than 50 years – investigating changes to the microbiome	Pathology & Tumour Biology	£0–£0.5m	75%–100%	Bowel Cancer UK
		£60,489,164	£38,880,635	

This document, entitled ‘Leeds Institute for Data Analytics – A vision for data science and data analytics at the University of Leeds: A strategic proposal’, has been developed through a consultation exercise over a period of 5 months. It has built on the earlier ‘LIDA Strategy Plan 2018-2020’, which was developed through workshops attended by LIDA staff and Faculty academics, as well as the wide circulation for comment of the draft. The following have been consulted on prior to writing and /or have commented on iterations of document:

Professor David Beech	Dr Mark Iles
Professor Julia Bennell	Adrian Iredale
Professor Karen Birch	Professor Mark Kearney
Professor Mark Birkin	Sir Alan Langlands
Dr Daniel Birks	Professor Nora de Leeuw
Gillian Booth	Dr Robin Lovelace
Professor Julia Brown	Jane Madeley
Dr Luke Burns	Professor Sir Alex Markham
Professor Janet Cade	Professor Christopher Megone
Professor Philip Conaghan	Professor Eva Morris
Professor Andy Dougill	Dr Michelle Morris
Dr George Ellison	Professor Alastair Mullis
Francesca Fowler	Dr Tim Peakman
Professor Alex Frangi	Professor Lisa Roberts
Roger Gair	Stewart Ross
Professor Mark Gilthorpe	Professor Roy Ruddle
Professor Susan M Grant-Muller	Professor Steve Scott
Professor Geoff Hall	Dr Christopher Smith
Dr Marlous Hall	Professor Paul Stewart
Professor Alison Heppenstall	Professor Laura Stroud
Professor Jenny Hewison	Professor Andrew Thorpe
Professor David Hogg	Professor Tom Ward
Jo Holmes	Dr Philip Waywell
Dennis Hopper	Professor Robert West
Professor Jeanine Houwing-Duistermaat	Professor David Westhead
Dr Kahryn Hughes	Dr Jianhua Wu
	Professor Hai-Sui Yu

Research Management Process:

How LIDA can help support your research



Leeds Institute for Data Analytics

Level 11, Worsley Building
Clarendon Way
Leeds
LS2 9NL

Tel: **0113 343 9680**

Email: **lida@leeds.ac.uk**

Website: **lida.leeds.ac.uk**

Twitter: **[lida_uk](https://twitter.com/lida_uk)**

