London's Global University

UCL

2023 – 2027

Data Strategy

Cover image: Segment of the high-dimensional graph structure of UCL study modules. Graph of 6573 learning modules as individual nodes, interconnected by the module selection of 100,636 unique students across 2017-2021. Node colour is defined by the community structure of student module by the Bayesian generative stochastic block model, revealing a natural organisation of the modules students tend to co-select.

Image credit: Dr James K Ruffle, High-Dimensional Neurology Lab

About this strategy

Executive summary

We must increase the value we obtain from our data by using it more effectively and more efficiently to deliver our strategic outcomes in a challenging environment, while remaining secure and compliant. In the medium term this means achieving common understanding of our data priorities; aligning and investing effort and resources in people, tools and technology; and ensuring accountability for data. In the longer term this will create the time, space and capability to leverage more advanced artificial intelligence techniques, alongside a culture that develops data-driven solutions to our most-pressing problems. This will enable us to achieve our vision for data.

Preface

The UCL Data Strategy sets out the vision for how we will use data to achieve our goals within a culture that values it.

Our data is a key asset, providing the basis for insight that supports UCL's decision-making and the delivery of its strategic aims, and enabling the connected university experience that is a key objective of the emerging Digital Strategy. In addition, good management of our data will play a critical role in helping us remain secure and compliant.

This strategy outlines a high-level plan for achieving high-quality data with speed, consistency and trust.

Two sub-strategies (alongside other supporting ISD and Information Security strategies) complement this strategy:

- the UCL Data Governance Operating Model, which describes how our proposed operating model for data governance will embed accountability for data;
- the Reporting & Analytics Modernisation Programme, which sets out a vision for how we will use technology and process to improve the speed, efficiency and quality with which reporting and analytics are developed, delivered and consumed.

Scope and reach

This strategy underpins the delivery of data and insight and the mitigation of security and compliance risks and is a pillar of the Digital Strategy. It is essential to the delivery of the UCL Strategic Plan 2022-27.

This strategy applies to:

- all UCL staff
- third party stakeholders with access to UCL data and who supply UCL with data.

This strategy covers:

- enterprise (i.e. non-academic research) data in all its forms, including student, education and research administration, finance, HR, and estates data. (While academic research data is out of scope, the Research Data Strategy and this Data Strategy are complementary)
- structured, semi-structured and unstructured data.

Approval, duration and review

This strategy will be implemented over the period of the Strategic Plan 2022-27. Progress will be monitored regularly by the Data Strategy Steering Group. This group will also review the strategy to check that it remains relevant and appropriate two years from launch, in 2025.

This Data Strategy was approved by UMC on 2nd May 2023.

Challenges to which this strategy responds

Every day UCL staff experience challenges caused or intensified by data that lacks timeliness, quality or fitness for purpose. These data challenges make it difficult or costly to deliver our strategic aims, and are not consistent with the open, rigorous, and innovative university we aspire to be. Challenges include:

- our efforts to extract value from data are often loosely coordinated and not focussed on strategically important critical data. For example, we do not have a strategic view on what data is critical, so data quality remediation efforts may be misguided or dissipated
- data critical to decision-making is often of poor or unknown quality, increasing the likelihood that decisions are poorly evidenced and that outcomes are poor. For example, we may not be able to accurately count our staff or confirm which rooms in which buildings belong to which departments
- data is not available when needed. We cannot react quickly to opportunities or take decisive action, because data must be extracted, cleaned, joined and analysed prior to decision. For example, joining research grant data with staff data to deliver Equality, Diversity and Inclusion insight may require significant manual effort and time
- there are many local, 'shadow' copies of data because data is not trusted. For example, a department may be working to its own student intake targets because those held centrally are incorrect or untrusted, resulting in more staff time spent curating shadow data, coupled with inconsistent decision making because data sources are different. Strategic and operational reports may be based on different copies of the same dataset. This leads to multiple, inefficient cottage industries in data growing up across the organisation
- the user experience is poor because data is not joined up, not correct, and not standardised. For example, moving staff between organisational units may require changes in upwards of 20 systems, or a new survey may not use an up-to-date version of the UCL organisational hierarchy
- data pipelines are not automated and are often reliant on single-points-of-failure. For example, it may not be possible to supply staff data required by University Management Committee to make a decision in the same week as an external statutory deadline, because the same individuals manually provide both datasets
- staff may not have the skills or capacity to solve analytics problems themselves, or if they do, to share those analyses with other colleagues. This may mean that external consultancies are hired at high cost to deliver one-off, throwaway solutions to analytics problems
- we cannot meet our external statutory and compliance requirements. For example, we may struggle to submit accurate statutory HESA returns, resulting in reputational damage, rework and audit
- we cannot meet our internal information security requirements. For example, we may be at risk of GDPR breach because personal student or staff data is stored locally or because role-based access to data remains unimplemented.

How it will feel when the vision is realised

In 2027, a member of staff working within this mature data culture will have a very different experience from that of today. By implementing this data strategy, delivering its use-cases, developing more advanced capabilities and giving data due consideration during change, we will first move to become a proactive, then a predictive and optimising data organisation.

Access to data

The standard reports, analytics, KPIs, and management information products you need for your role will be easily accessible. Access to these data and insight products will be automated and based on your job role and new staff will have access (supported by training and guidance) from day one. These data products, their supporting metainformation, and underpinning datasets will be available through a single portal, and their use will be seamlessly embedded into organisational processes and context.

You will be able to access simplified new reports and insights which will be better aligned with your requirements. To help model scenarios and optimise decision-making, you will be able to access predictive analytics, which leverage machine learning models that have been rigorously tested and developed in collaboration with academics. Underpinning datasets will be available via intuitive self-service connection for analysis by data-literate staff.

Quality of data

You will trust the data used in all data, insight and analytics products, and you will find no inconsistency across products: admissions and enrolment data will be the same everywhere, because the same dataset will underpin and be reused in multiple products. Data will be high-quality because senior data owners have prioritised their most critical data and invested their time and that of their staff in improving its quality.

Governance of data

Senior staff, including UMC, will consistently reference data governance, question the provenance and accountabilities of new datasets, and model the behaviours aligned with a mature data culture. The Data Governance Group will be a critical, respected decision-making body for how data is governed at UCL and its advice will be sought by change portfolios creating new data products. Data owners will authorise access to data assets that they own and the data access spectrum will be clearly implemented and well understood.

Shareability and security

Data will be securely shareable. Because the reports and datasets you need are available for self-service analysis, with access managed by data owners and role-based, you will have less need to download and share datasets manually. As a result, the likelihood of GDPR (General Data Protection Regulation) breaches will decrease.

Career paths for analysts

The work of analysts will be better coordinated, with clearer career paths, and a centre of excellence will support the existing community of practice. Automation will mean that analysts spend less time maintaining manual data flows and more time learning and using innovative and advanced techniques to make better use of data. In addition, following the successful academic collaborations with the Department of Statistical Science and the Institute of Neurology, you may be part of cross-functional academic and professional service teams which will take forward critical use cases which leverage advanced analytics to optimise, predict and model outcomes. Our culture will allow us to accept and trust these models, even as their complexity increases, and our ability to leverage academic expertise and collegiality will differentiate UCL from its peers.

Implications for student and staff experience

Students and staff will enjoy improved experiences as analytics will help us understand how to optimise delivery of services important to them, just as decision-making and predictive analytics will help you identify problems and pain points before they become widespread issues. The benefits of AI and predictive analytics will be within our reach, allowing us to optimise how we use our resources, how we respond to change and uncertainty, and how we execute what our stakeholders demand of us.

Implications for UCL decision-making

This data culture will improve the speed and quality of our decisions. When new, critical, time-sensitive use-cases arise (perhaps provoked by external changes), analytical resource, capability and trusted data will be leveraged quickly.

We will deliver excellent data-driven outcomes to problems, facilitated by:

- the goal clarity created by effective governance and prioritisation
- time freed up through automation
- data governed by owners and made available through the data hub
- the analytical capability of staff developed through data and insight apprenticeships.

Introduction

The use of data determines how far we can achieve our operational goals and our strategic objectives, which comprise: an excellent education and student experience; world-leading impactful research and innovation; and more efficient operations.

In addition, harnessing our data can help us achieve these aims in a way that is equal, diverse and inclusive, and financially and environmentally sustainable.

The external environment is changing in ways that threaten our academic excellence. Challenges include:

- major external financial pressure on income and costs
- an increasing volume and complexity of research and innovation partnerships
- the need to balance student numbers, student expectations and regulatory demands
- intensified competition for talent and increasing staff workload
- high cost and increasing demand for space across education and research.

This Data Strategy outlines how our data can support the delivery of the UCL Strategic Plan 2022-27 within this challenging external environment, as well as for the long term. Data is described as a key enabler of this strategic plan, most obviously through provision of actionable data and insights to support strategic delivery and decision-making.

From 'Enabling our academic mission, (2022):

"All colleagues need access to relevant, accurate, logical, and consistent data and insights to inform and support meaningful planning and institution-wide strategic monitoring...Such data and insights will empower our leaders to make transparent decisions based on publicly available data."

Further, high quality data will enable:

- a connected, efficient student and staff experience as a key enabler of the "functions that are essential… but that are often hidden from view"
- the simplification and automation of process which "reduces costs, improves quality of service provision, and frees professional services staff involved to deploy their skills at a higher level."

Good management of data will also improve security and compliance (legal and regulatory). Reliable and attributable sources of data, alongside privacy-by-design, will improve information security, reducing the risk of loss, inappropriate disclosure and tampering.

In sum, increasing the value we get from our data by deliberately and strategically managing it will enable UCL's vision in three key areas:

Decision-making and insight – to support effective strategic delivery; to make high-quality decisions in faculties and in the professional services domains that support them; to provide support to Council, UMC, and the Provost by providing insightful answers to the key questions that face our university

Connected experience[1] – to deliver staff and student experiences around a single consistent view of data which is open, easily-accessible, comparable, and interoperable

Security and compliance – to reduce risk of data loss and tampering; to meet our statutory, legal and regulatory obligations for data; to increase trust and leverage data as openly as possible, and as closed as necessary.

1 The Data Strategy aligns closely with an emergent Digital Strategy, which has a connected university experience as a key objective and includes data as a core theme.

This Data Strategy outlines a vision for UCL's data, and describes the principles, priorities and enablers, alongside the outcomes it will enable.

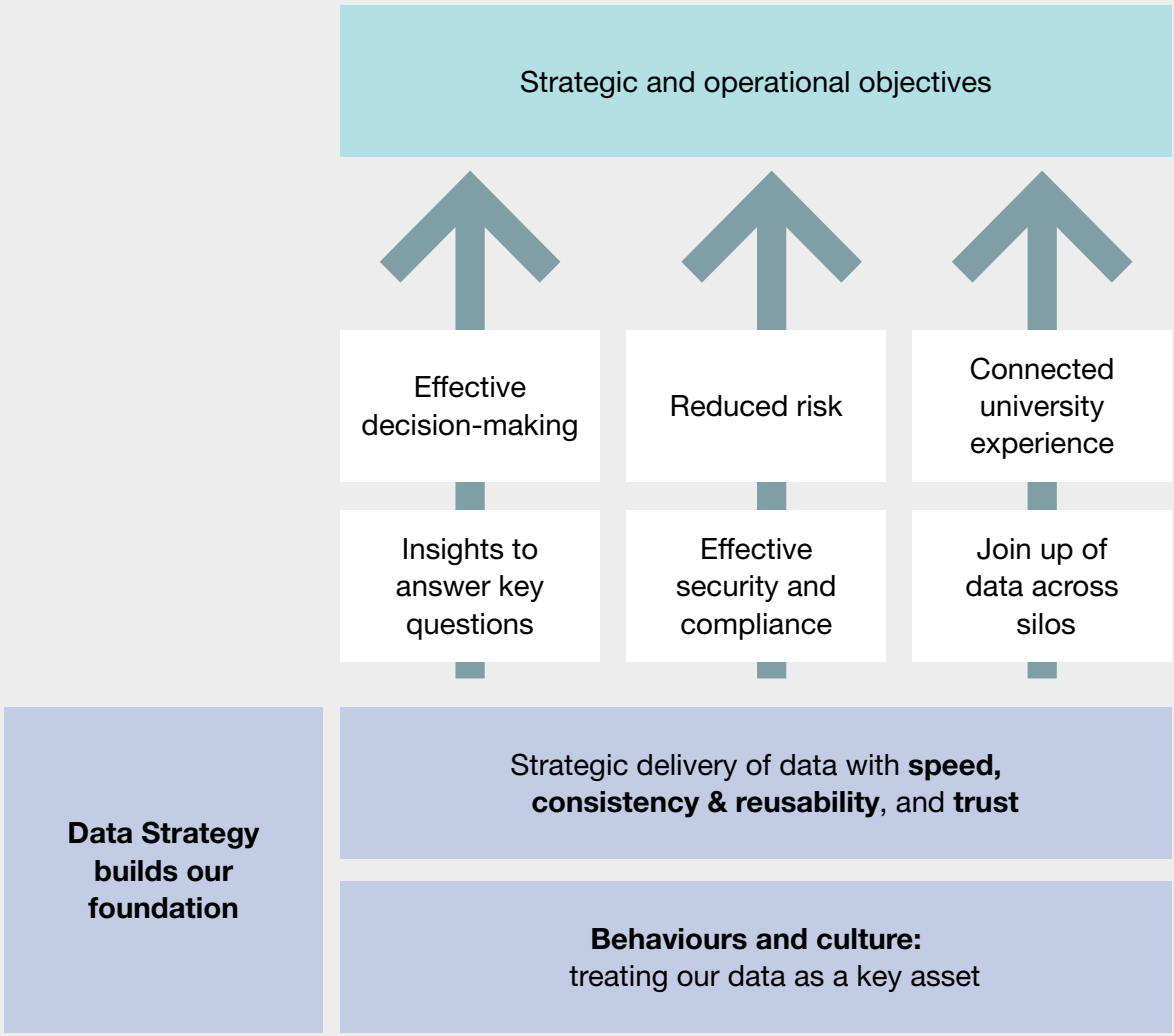

Fig 1. *How does our data foundation support UCL's aims?*

Vision for data

This Strategy proposes a vision for data that we seek to achieve over the next 2-5 years:

"At UCL, our data is a key asset, delivered with speed, consistency and trust, which drives insight used to support the University's decision-making and strategic delivery; student and staff experience; and security and compliance."

Strategic delivery of data will be based on five principles; enabled through five areas of focus across four themes; and delivered across three outcomes.

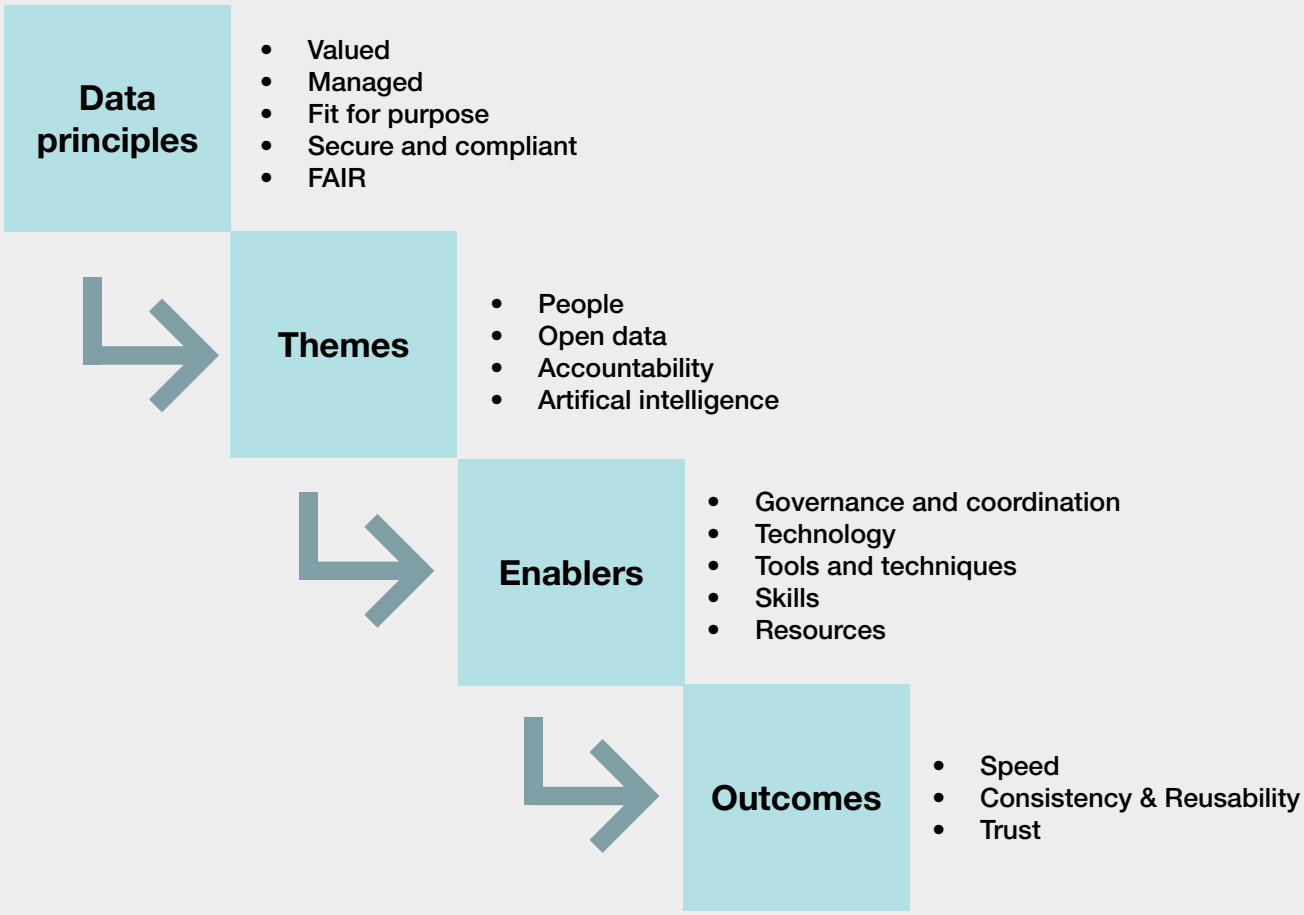

Fig 2. Approach to strategic delivery of data

Data Strategy

This section outlines our desired principles for data, and the mechanisms that will enable realisation of the desired outcomes.

Principles

These principles for data help improve the quality and consistency of our behaviours around data. Testing any organisational change or desired capability against these principles will contribute to the development of our data as a valued asset.

1. Data is valued

- Our data is a vital asset, central to our success. It is owned and governed in the same way as other valued assets such as buildings or money.
- Data belongs to the University, not to the individual[2]. Therefore, it is managed on behalf of the university by everyone interacting with it.

2. Data is managed

- Data is managed according to its importance, and throughout its lifecycle, to assure appropriate availability, integrity, avoidance of loss and continuity across change.
- There is clarity on who is responsible, and individuals are appropriately skilled.

3. Data is fit for purpose

- Data is the right quality. It is not perfect, but its quality characteristics are pragmatic, appropriate and transparent.
- This quality is regularly monitored with simple to understand data quality measures. Interventions ensure the data remains fit for purpose.

4. Data is secure and compliant

- The integrity of data is assessed, monitored and managed. Effective information security procedures allow the right people to use and update data and restrict inappropriate use.
- Availability of personal data is considered in terms of ethics, proportionality and need, while respecting GDPR principles of transparency and legitimate purpose.

5. Data meets the FAIR principles[3]

- **F**indable – data should be easy to find.
- **A**ccessible – data should be accessible through organisational authorisation and authentication.
- **I**nteroperable – data should be standardised and easy to work with.
- **R**eusable – data is reusable for many purposes, because it is findable, accessible and interoperable.

2 Data generated by researchers during their research is explicitly excluded from scope.
3 The FAIR principles seek to optimise the reuse of data: https://www.go-fair.org/fair-principles/

Themes

The four themes will support development of a mature and effective data culture. Taken together these themes also align closely to the strategic value clusters that underpin UCL's strategic plan delivery over the next five years, particularly openness, mutual accountability, and rigour and innovation.

1. People

We must invest in our people: those who create data and develop solutions based on it, and those who consume it. Our staff must have the capability and confidence to work with data if we are to make decisions and deliver outcomes according to our strategic goals. Collectively, we must understand our respective roles in managing data professionally as per other assets, and this change in behaviour must be seen as positive and sustainable, and not seen as taking time away from other activities.

As well as becoming 'data literate', we must understand that data is not infinite and cost-free, and that there is a need to prioritise the 'data budget' on activities best aligned with UCL's strategic goals.

We will continue the good work of the Data & Insight Community of Practice in sharing best-practice, and we must go further in providing training and meaningful career pathways for our data professionals.

2. Open data

UCL must become an explicitly 'open data' organisation. Data will become a strategic asset freely available to UCL staff, not a discrete resource for individuals or departments. It must be signposted and available to those who need it. This means that we understand our 'access spectrum'[4] of data and make deliberate choices about how and to whom data will be made available, internally and externally, given our obligations to privacy, security, and commercial sensitivity.

Data will be accessible, transparent and well-defined, and managed in accordance with risk and compliance. Datasets underpinning key performance indicators (particularly for UCL's strategic plan) will be made available in secure, governed and accessible forms for scrutiny. UCL is a large and complex organisation: striving for openness, transparency, and devolution of access will help us to build trust and improve data quality. We will consider data strategically as an enterprise-wide asset, not a local asset to be recreated independently in each department.

Data should meet the FAIR[5] principles for optimising reuse of data.

3. Accountability

We will embed accountability for data through roles and responsibilities and this accountability will be clear, visible and consistent. Data owners will collaborate across domains to resolve complex issues that have previously been considered too hard to fix: for example, how well our space data enables our efforts to optimise space utilisation.

Data owners will prioritise the governance of the data they own, and this will be evident in the time they and their teams invest in that governance. Identified data stewards will have an explicit mandate to improve data quality to reduce cost and improve the outcomes facilitated by that data.

There will be a clear data quality remediation process for all data and staff will be responsible for reporting bad data. Collectively, we will focus our efforts on improving the quality of our most important data. Data quality is managed according to all known use cases.

4. Artificial intelligence

We use the term 'artificial intelligence' to encompass analytics, machine learning, automation, and data analysis. As per the IBM definition[6], at its simplest, AI "combines computer science and robust datasets to enable problem-solving".

To realise the desired value from our data, we will strive to develop rigorous and innovative solutions to problems, leveraging AI tools and techniques, including through collaboration with UCL's academic experts.

This theme is deliberately placed last, because while potentially highest in value, it is also the most challenging, and cannot be leveraged independently of the previous three themes.

4 https://www.theodi.org/about-the-odi/the-data-spectrum/
5 https://www.go-fair.org/fair-principles/
6 https://www.ibm.com/topics/artificial-intelligence

Enablers

This section outlines the key enablers required to achieve the vision for data.

1. Governance and co-ordination

The strategic governance of data can be split into two categories: the governance of data itself through a data governance framework; and the governance of this Data Strategy.

Data governance is fundamental to an organisation's ability to use data strategically and increase the value of its data. This involves creating senior accountability for data, which is then leveraged to improve data quality and increase trust. This model is described in detail in the separate Data Governance Operating Model (see Annex III).

Separately, governance of this Data Strategy and its implementation is critical to its success. A Data Strategy Steering Group will be established to oversee the launch and implementation of this Data Strategy, and to monitor progress and success. Post-launch, the enabling work for this Data Strategy will be reported through the Strategic Plan 2022-27 governance structure, so that there is alignment and prioritisation at a senior level.

Leadership and co-ordination are required to position data as a true institutional asset. To do this will require strong governance and the collective adoption of two behaviours we have traditionally neglected at UCL:

a) the careful consideration of data during all change and transformation

b) strong prioritisation, so that we invest only in important data (as efficiently as possible).

Much of the enabling work will be co-ordinated and delivered through the existing Change & Digital portfolios, with the identified use-cases and required capabilities raised as explicit priorities within the People, Money & Insight portfolio, or other portfolios as appropriate.

2. Technology

Technology is a key enabler for the strategic vision for data because it provides solutions for how we manage, prepare, store, analyse and consume data. This data needs to be structured logically and made available to other services either for consumption or to ensure it is kept up to date. To achieve the vision, we need a sound underpinning of technological capability and whole-organisation data architecture.

UCL's strategic vision for digital transformation describes the desired outcome:

"…standardised and simplified enterprise data and services, which demand integrated and secure platforms, powered by modern cloud connectivity… [data is] owned by UCL-wide data owners and made available to self-serve on demand…secure data requires integrated and secure platforms, powered by modern cloud connectivity."

UCL Digital Strategy

The UCL Digital Strategy will focus on:

- Cloud - exploiting and adapting to more out-of-the-box cloud services and standard commodity platforms where they exist and building our own solutions in the cloud where appropriate
- Integration - joining together our services and data, flexibly and safely
- Identity - standardising process and data around identities and access
- Security – ensuring we maintain data confidentiality, availability and integrity.

One of the key themes of the emerging Digital Strategy is the delivery of a connected university experience, which envisions design of processes and data around students and staff: it aims to allow UCL staff to be better connected to their work, to each other, and to the running of the university. It is enabled by automated and efficient processes; better understood and more accessible data; and the equipping of senior leaders with timely management information. Underpinned by the Data Governance Operating Model, it will result in a more efficient, more agile and more resilient university better able to respond to challenges and disruption.

Finally, the way in which technology enables the development, delivery and consumption of data products such as reports and analytics will be outlined in the Reporting & Analytics Modernisation Programme.

Converging on fewer, reliable tools and techniques is desirable because this will reduce cost and allow us to grow internal expertise and resilience. We must work with ISD to ensure that self-service and enterprise solutions exist for the capabilities required, and we must leverage these wherever possible.

3. Tools and techniques

The tools and techniques we use to extract value from our data must be proportionate and fit-for-purpose. This means understanding the limitations of different techniques (for example, statistical models, algorithms or visualisation tools) and applying the appropriate one for a given use-case, recognising where additional expertise is required. We must also proactively monitor developments in the field, so that we can incorporate new data tools and techniques when appropriate.

Automation is a key enabling tool: just as other areas of UCL have successfully begun automating repetitive, manual processes to free up staff time to add value for users, we must continue to invest in automation of key data pipelines so that time can be spent on interpretation. At present, much critical analysis is still based on slow manual data pipelines by busy individuals, which makes it hard to react to opportunities and threatens the quality of data and speed of decisions. We must build on the foundation of cloud-data-warehouse, which contains automated pipelines for some but not all critical strategic and operational data. Striving for openness, transparency, and devolution of data will help us to manage, build trust in, and improve data quality.

As our institutional data maturity grows, and we aim for predictive and prescriptive analytics, data products will rely increasingly on AI techniques. For example, we might seek to improve the research grant pipeline by predicting, at a grant level, the likelihood of grant applications becoming successful. As use-cases become more advanced, the internal workings of reports and analytics will not necessarily be understood by users. The mark of a data mature organisation is that the value of these AI techniques can be translated for lay-audiences who trust their outputs, even as the techniques applied become more complex.

4. Skills

As outlined in the People section of the strategic vision for data, the capability of our staff in working with and understanding data is key to how we make decisions and deliver outcomes in support of our strategic goals.

As Sir Tim Berners-Lee advocates[7] in his position as President and Founder of the Open Data Institute; "We need to train both the people who are putting data and information out there, as well as those reading it, how to interpret and question it to ensure they understand it and are not being misled or deceived".

The need to improve the skills of our staff can therefore be summarised into two role types. Staff may fulfil one or both roles:

- Staff who use data
  Those colleagues tasked with using data to do their jobs. Every member of staff, at some point, will fulfil this role. These colleagues must be able to use and interpret data with confidence and understand its limitations.

- Staff who supply data
  Those colleagues tasked with developing data products or other solutions, including reporting, analytics or statistics. These colleagues must be able to wrangle, manage, govern, analyse, and visualise data, design user-centric data solutions, translate data for colleagues, and help colleagues understand its statistical limitations.

The emergence of self-service and the convergence towards fewer data platforms (for example, data visualisation tools) will help us identify and invest in the most effective development programmes for these staff. Recognising that the leading edge of data skills is constantly advancing, investing in staff data skills will require constant horizon scanning and continued investment, so that UCL's collective data abilities remain current.

7 https://www.iod.com/resources/blog/employment-and-skills/why-upskilling-a-generation-of-directors-in-data-literacy-is-key-for-businesses-in-a-post-pandemic-world/

5. Resources

The Data Strategy implementation will be enabled by people, financial and data resource, some existing, some of which will be required additionally, and all of which will require alignment.

Existing human resources include:

- The Data & Insight team in the Office of the Vice-President (Strategy)
- Change & Digital portfolio and product teams
- Data Protection and Information Security teams
- Reporting teams embedded in VP offices and professional services.

In addition, unlike many organisations, there is significant and advanced data expertise in academic domains themselves, as well as those departments that support them:

- Academic departments including Computer Science, Statistical Science and the Institute of Neurology
- Research IT Services
- UCL Advanced Research Computing Centre.

It is imperative to align this resource data expertise across the institution behind the most important use-cases. Collaboration across functional areas and data domains is key, especially as the most valuable data products increasingly contain more than one dataset (for example, research grant data from Worktribe and staff data from MyHR). More advanced capabilities, of the kind found in academic departments and Research IT Services, will be increasingly required to leverage AI solutions.

At present, important data initiatives (for example, preparation of important Estates space utilisation data) are often outsourced to external consultancies or contractors. As well as increasing the cost of delivery per data product, outsourcing increases risk, because solutions are not integrated into UCL's governance or technological infrastructure. In addition, expertise and know-how is lost, and the chance to cultivate and provide development experiences for UCL's data professionals is missed. Becoming a data mature organisation will mean outsourcing less critical data work. This will require increasing internal capacity and capability, including through efficiency and automation.

Where more capacity is required, or where skills do not exist and cannot be developed internally, we must attract and retain data talent. The data professionals Career Framework, developed by the successful Data & Insight Community of Practice, plays a role in retention by providing guidance on the skills and competencies required to develop data careers at UCL. Clear job descriptions and flexible career paths will be required to give UCL's data professionals options to further develop and help junior staff develop a data specialism over time.

Data itself is a resource: externally purchased datasets (such as Jisc, HESA, Uniforum or UCAS datasets) can help us understand the sector and benchmark ourselves against peers. These datasets should be purchased once, stored centrally and made available to those who need them.

Outcomes and KPIs

The success of this strategy will be assessed by measurable, quantitative key performance indicators, alongside an improvement to UCL's data culture: how it 'feels' to use data at UCL. The main outcome areas, along with associated KPIs, are described below:

1. Speed

To make timelier decisions and react more quickly to circumstances.

- o Reporting and analytics solutions for standard questions are automated end-to-end
- o All staff have appropriate, secure, automated access to standard reporting and analytics solutions
- o Users have skills to interpret and present data effectively to reduce lead time.

Data is available when it is needed. For new or urgent use cases, a framework is in place which helps us understand what data is needed, who is accountable for it, and what its quality is. Decisions can be made more quickly, because data does not

require sourcing, cleaning or defining prior to use, and because outputs are high-quality and fit-for-purpose.

KPIs for Speed outcome area			
KPI #	**Description**	**Measure**	**Target (baseline)**
1.01	Time saved	Total data management hours saved through data automation + total data management hours saved through data warehousing	**500 hours/year by 2025**
1.02	Speed of access and induction for new Deans, DoOs, Department Managers and HODs	% of new senior staff having appropriate access and induction to MI dashboardas within 5 days of starting	**80% by 2025**

2. Consistency and re-usability

To use the same data for many different purposes and to re-use data efficiently.

- o There are as few versions of truth as possible: single datasets are used for multiple purposes
- o Key datasets are stored in the data warehouse
- o Staff analysts can connect to centrally-warehoused datasets
- o Fewer key analysis tasks are outsourced.

Technology is leveraged so that staff across UCL share access to single sources of truth, while single sources are used for multiple purposes. Data pipelines become efficient and the cost of recreating data pipelines multiple times and independently is reduced.

KPIs for Consistency & Reuse outcome area			
KPI #	**Description**	**Measure**	**Target (baseline)**
2.01	Self-service connections to centrally-curated datasets by staff	n of connections to central data-warehouse/data lake datasets (3-month sum)	**100% increase by 2025**
2.02	Strategic dashboard views (overall)	n of views of strategic management information dashboards in D&I portal (3-month sum)	**100% increase by 2025**
2.03	Strategic dashboards engagement (key senior management group)	% of UMC, HODs, DoOs viewing strategic management information dashboards in D&I portal (in 3-month period)	**60% by 2025**
2.04	Datasets made available according to FAIR principles	n of datasets curated according to FAIR principles and made available to staff	**50% increase by 2025**
2.05	Outsourced data work	£ spent on data / analytics providers (12-month sum)	**50% reduction by 2024**

3. Trust

To underpin our insights with well governed and managed data.

- o Data is of quality sufficient for the intended purpose
- o Business terms are understood and well-defined
- o Data is as open as possible, and as closed as necessary; we are secure and compliant.

Accountability and stewardship for data is transparent and embedded. It is clear who is responsible for data and it is clear how to report data-quality issues to them. Data owners prioritise fixing high-priority datasets and collaborate to solve cross-domain data problems. Data is trusted and definitions are agreed and transparent, meaning that better decisions are made.

KPIs for Trust outcome area

KPI #	Description	Measure	Target (baseline)
3.01	Data criticality	n data criticality maps completed	**3 domains by 2025**
3.02	Data governance coverage	% data domains with identified and briefed Data Owners and Data Stewards	**95% by 2024**
3.03	Data quality	n of domains with automated data quality reports available	**3 key domains by 2025**
3.04	Data terminology	n of key business terms defined and published	**10 by 2024**

In addition, the following KPIs measure the success of the whole programme:

KPIs for Overall Data Strategy

KPI #	Description	Measure	Target (baseline)
2.01	D&I apprenticeship completion	Number and % of D&I apprentices completing	**90% of apprentices completing in cohort 1**
2.02	Data professional retention	Data staff (D&I apprentice + CoP member + CoE member) mean employment length	**Equal to staff mean by 2025**
2.03	Overall programme	n of priority use-cases delivered	**15 by 2025**
2.04	Overall programme	Staff survey % agree: I have access to the data and insight I need to do my job.	**80% agree by 2025**

Taken holistically, the delivery of data with speed, consistency and trust also has associated benefits for information security: reliable and attributable sources of data, alongside privacy-by-design, will improve information security, reducing the risk of loss, inappropriate disclosure and tampering, whether accidental or deliberate. The provision of readily available and authoritative data addresses the need to maintain local, often out-of-date, copies and to share these via email or other uncontrolled means. This strategy proposes the need for clear accountability and ownership, leading to deliberate application of confidentiality, integrity and availability; three tenets that underpin information security.

Implementation approach

The approach to implementation of this strategy will be user-centred and prototype-driven.

- **User-centred** means that the strategy will seek to address specific data-related use-cases and pain points that are of value to UCL's users, but for which the user is unserved. Users will be involved throughout design, implementation and evaluation, and the development process will be iterative and responsive to user feedback.
- **Prototype-driven** means that approaches to the capability goals of this strategy will be tested through delivery of small-scale prototypes. These prototypes will be used to foster learning and improvement of a given capability, which will then be rolled out to similar use-cases in other areas.

This approach will allow us to start small, without trying to change the entire organisation at once.[8] The use-cases have been identified through consultation with users, as part of the strategic plan consultation, or through the work of the Change & Digital portfolios, and are included in in Annex I.

The capabilities required to enable these use-cases can be categorised into the four themes of this strategy:

1. People

- Data & Insight Careers Framework
- Data apprenticeships
- Data & Insight Community of Practice
- Analytics Centre of Excellence

2. Open data[9]

- Data Hub
- Data pipeline automation
- Analytics sharing and governance
- Data democratisation and self-service
- Secure and automated access provision
- Insight Portal

3. Accountability[10]

- Data accountability
- Data definitions
- Data quality reporting
- Business change

4. Artificial intelligence

- Academic collaboration
- Forecasting and modelling
- Machine learning
- Statistical literacy
- Data communication

To ensure successful implementation of this strategy, it is critical that the capabilities (listed above) developed through delivery of use-cases are embedded in a joined-up manner. Generalisable outcomes from the use-cases will be applied at the institutional level where relevant, to improve the overall data experience for all users.

8 This follows the following the Gartner approach to implementing a successful data strategy: https://www.gartner.com/document/4014345?ref=solrAll&refval=354310469
9 Capabilities mainly delivered through the Reporting & Analytics Modernisation Programme
10 Capabilities mainly delivered through the Data Governance Framework

The Data Strategy Steering Group will ensure coherence and integration in the implementation of the Data Strategy, avoiding this risk of fragmentation to which the use-case approach might give rise. In addition, other mechanisms designed to achieve this include:

- the technical and architectural underpinnings, which are being developed consistently and in a connected way through the relevant Change and Digital Portfolios, coordinated through the Reporting & Analytics Modernisation Programme implementation
- the data strategy KPIs which incentivise single solutions
- the Data Governance Group, comprising senior data owners from every area of the institution (and which is already in operation)
- the principles for data, which guide and are embedded in all solutions. Specifically, the 5th principle includes ‘reusability’.

Approaches will be refined and further use-cases will be identified through Change & Digital portfolios and associated strategic plan governance and prioritised through the Data Strategy Steering Group. In this way, the required capabilities and culture will improve and become prevalent, as specific items of value are delivered quickly.

ANNEXES

ANNEX I: Priority use-cases

i. HESA Data Futures

Opportunity: HESA (the Higher Education Statistics Agency) is changing the statutory return of student data, moving in 2023 from an annual event to three times per year. Instead of thirteen months to correct any data issues, there will be a 4-6 week period per return. We need to ensure that the data is of acceptable quality for HESA, as this return is a condition of the university's license with the Office for Students and non-submission risks financial penalty and reputational damage.

Users: Student records, HESA

Capabilities: Data governance framework, automated data quality reporting, business change

ii. staff data quality

Opportunity: to improve the quality of staff data, which is highly valued (including for HESA, budgeting, planning, and to monitor KPIs and EDI goals). Reduce the time-to-insight based on staff data and remove the requirement for manual intervention. This will impact the Cubane Uniforum return, reducing the lead time before our data can be leveraged, and the resource invested in cleaning the Uniforum dataset.

Users: Users, stewards and owners of staff data

Capabilities: Data accountability, data quality reporting

iii. student enrolment data

Opportunity: reduce the number of sources of student enrolment data to improve trust in student enrolment data, raise its value, and reduce the time spent (for example in meetings) discussing why sources are not in agreement.

Users: Faculties, Departments, Finance, UMC, Planning

Capabilities: Data definitions, data communication

iv. offer targets and forecasting confidence

Opportunity: improve accuracy of admissions offer targets and student numbers forecasting, to heighten trust in the planning process and support operational and financial planning for student intake.

Users: Admissions, Departments, Faculties, Finance, UMC

Capabilities: forecasting and modelling, academic collaboration

v. Student Number Planning impact lead times

Opportunity: improve Student Number Planning processes by reducing the amount of manual intervention and hand-offs between departments. Consolidate multiple sources of truth for SNP data to reduce errors and increase efficiency, building trust in this critical process.

Users: Faculties, Departments, Estates, Planning, Finance

Capabilities: Data Hub, data pipeline automation

vi. Athena Swan application process

Opportunity: Athena Swan ratings are critical to UCL. Build a common data source for Athena Swan data to improve efficiency and support UCL to achieve its EDI goals.

Users: Faculties, Departments, EDI

Capabilities: self-service, data democratisation, data communication.

vii. data professional retention

Opportunity: develop career pathways for UCL's data professionals and establish retention mechanisms, such as professional development opportunities or apprenticeship programmes.

Users: Line-managers of data professionals, data professionals, D&I Community of Practice

Capabilities: Careers Framework, Apprenticeships, Community of Practice, Centre of Excellence

viii. award gaps

Opportunity: following a successful proof-of-concept to deliver programme level ethnicity award gaps, undertake further analysis of award gaps, in accordance with Office for Students requirements. Include confidence or statistical significance, as well as GDPR considerations, in this analysis, and present it simply and automated as far as possible, to best support departments to make interventions.

Users: Faculties, Departments, EDI, WP

Capabilities: Statistical literacy, data communication, data pipeline automation

ix. accessing Data & Insight portal

Opportunity: improve access to the Data & Insight portal, which contains key institutional management information (including student enrolment and admissions), for example, with faster access to critical dashboards for new members of senior staff starting at UCL. Automate access based on user role or job type. Consolidate delivery of all important reporting through this single portal.

Users: Staff users, Data Protection Office, ISD

Capabilities: secure and automated access provision, accountability

x. unifying organisational hierarchies

Opportunity: unify the versions of the organisational hierarchy, such as those used by Portico and HR/Finance, to eliminate mismatches at the point of consumption of any data using both sets, including enrolment vs targets.

Users: Users of student number planning data, Finance

Capabilities: data accountability, business change, data quality reporting

xi. master data source for estate and assets

Opportunity: Estates data is critical to delivery and measurement of the strategic plan. Establish a trusted data source for key Estates datasets, including buildings, assets and ownership. Data quality and data governance will be integral to this new source.

Users: Estates, UMC

Capabilities: data accountability, data pipeline automation, business change

xii. Key Performance Indicators and supporting data

Opportunity: Key Performance Indicators are being developed as part of the Strategic Plan 2022-27. Supporting contextual dashboards and underlying data, available through a single portal, are required to enable users and accountable parties to effect the necessary changes required to improve performance against each KPI. In addition, underlying datasets should be securely connectable so that skilled users (including academics) can connect to trusted data warehoused data sources to carry out their own analysis, using supported tools.

Users: UMC, KPI owners

Capabilities: Insight Portal, data democratisation, self-service, secure and automated access provision, data accountability

xiii. sentiment analysis for UCL brand

Opportunity: internal capability to deliver sentiment analysis is needed to move away from costly outsourcing to measure the success of the reputation elements of the strategic plan. In addition, analysis of survey free-text comments (for student and staff surveys) should be automated to realise the full value of this data.

Users: External Engagement, VP Strategy, consumers of strategic KPIs, VP ESE

Capabilities: machine learning, academic collaboration, analytics sharing and governance.

xiv. size & shape, timetabling and scenario planning

Opportunity: Given the uncertain external environment, student demand, and known space pressures, we need to build the internal capacity, capability and data for scenario planning, so that we can model the financial and operational impacts of possible student number futures. This will remove the need to rely on third parties deliver costly, throwaway analyses.

Capabilities: Data Hub, academic collaboration, forecasting and modelling

xv. realizing value of internal staff survey data

Opportunity: improve our survey design capability to ensure that data analysis requirements are taken into account. A multitude of suppliers are engaged to deliver surveys, each with unique technologies and variable quality of outputs . In addition, ensure that UCL owns its own survey data.

Users: UMC, HR, Faculties, departments

Capabilities: data pipeline automation, business change, academic collaboration

xvi. research grant network and success analysis

Opportunity: realise the value of the largest research grant dataset in the sector (UCL makes the most grant applications of any university). This dataset is ideal for advanced network analysis, which will provide insights into the real networks of academic collaboration that exist beyond the artificial hierarchies imposed by the organisational structure. In addition, this analysis could be extended to allow for identification of the factors contributing to research grant success, which could facilitate better forward planning (both physical and financial).

Users: UMC, Faculties, departments

Capabilities: academic collaboration, forecasting and modelling

ANNEX II: Timeline

Data Strategy: implementation timeline		2022-23		2023-24				24 -25	25 -26	26 -25
		T3	T4	T1	T2	T3	T4			
Use-case	Award gaps (viii)									
	Offer targets and forecasting confidence (iv)									
	Student Number Planning impact lead times (v)									
	HESA data futures (i)									
	Poor user experience accessing D&I portal (ix)									
	Staff data quality (ii)									
	Research grant network and success analysis (xvi)									
	Student enrolment data (iii)									
	KPIs and supporting data (xii)									
	Athena Swan application process (vi)									
	Realising value of internal staff survey data (xv)									
	Unifying organisational hierachies (x)									
	Data professional retention (vii)									
	Master data source for estate and assets (xi)									
	Sentiment analysis for UCL brand (xiii)									
	Size & shape, timetabling and scenario plannning (xiv)									
	Other use-cases identified by Steering Group									
Governance and associated strategies	Data Strategy approved									
	Data Strategy Steering Group stood up									
	and associated Data Strategy Review point									
	Reporting & Analytics Modernisation Programme									
	Data Governance Operating Model									

(As per the Digital & Change termly planning cadence, T4 = the summer period between term 3 and term 1.)

ANNEX III: Data Governance Operating Model

https://www.ucl.ac.uk/strategy-data-insight/data-governance