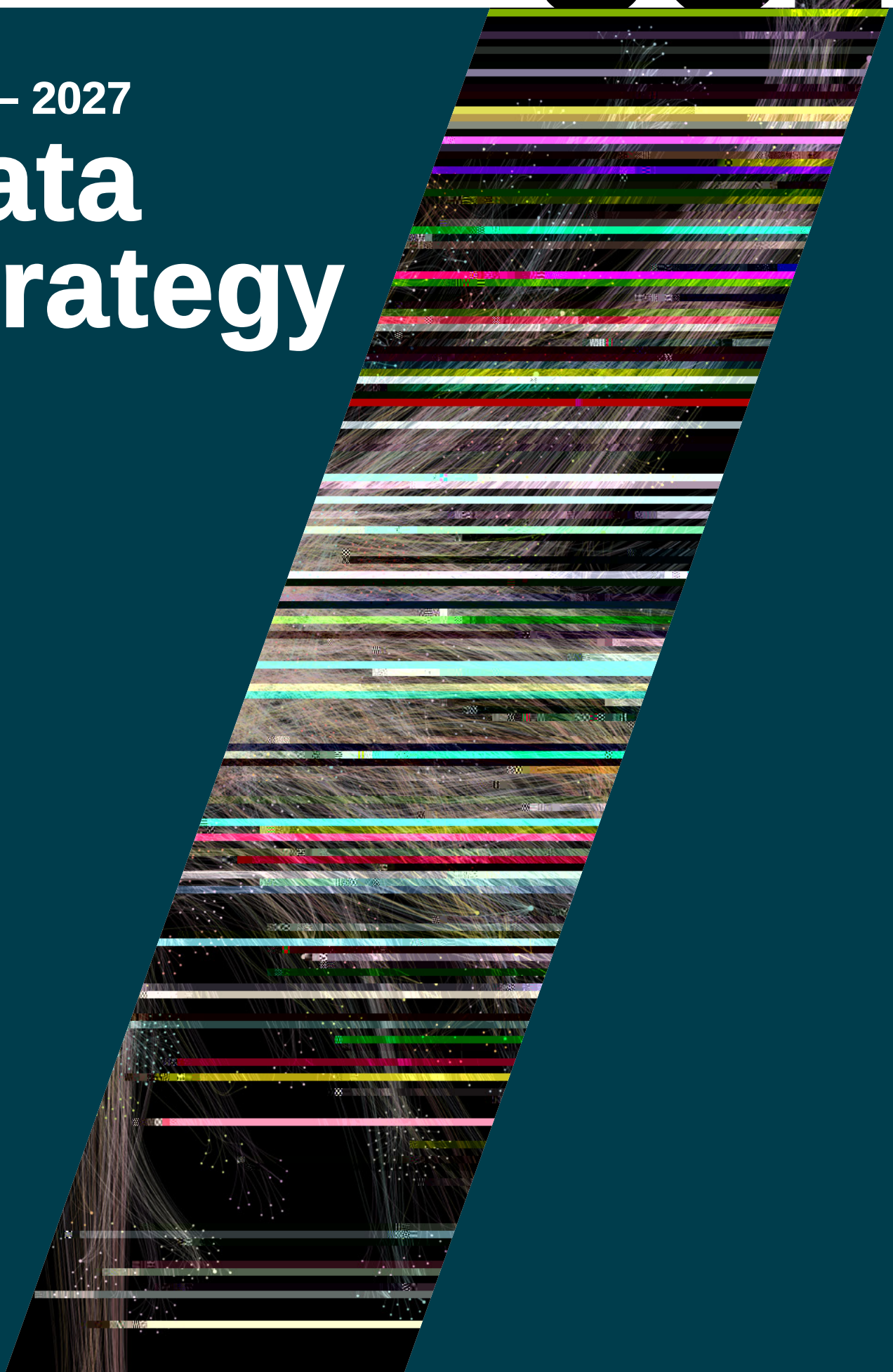
UCL

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 de□ned 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 Ruf□e, High-Dimensional Neurology Lab

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. Idata T0.01 Tw 0 -1.2 T

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 rst 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 simpli ed 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.

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 ows and more

The use of data determines how far we can achieve our operational goals and our strategic objectives, which

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

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

3. Tools and techniques

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.

2. Consistency and re-usability

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

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

Technology is leveraged so that staff across UCL

3. Trust

To underpin our insights with well governed and managed data.

- o Data is of quality suf□cient for the intended purpose
- o Business terms are understood and well-de□ned
- 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 rehdmata and it is clear

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. Speci□cally, the 5th principle includes ‘reusability’.

Approaches will be re□ned and further use-cases will be identi□ed 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 speci□c 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

(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-gT/GS1 gs1stra9