Research IT Services Roadmaps

Jan 2020
Contents

eResearch Infrastructure ecosystem: Powering UCL’s ambitions ................................. 4

AI and Data Science ........................................................................................................ 6
  Current State .................................................................................................................. 6
  Plans for 2019/20 ......................................................................................................... 6
  Ambitious Goals .......................................................................................................... 6
  Costs ............................................................................................................................... 6
  The AI Roadmap in more detail ................................................................................... 7
  Timeline ......................................................................................................................... 8

Computationally powering UCL’s research ................................................................. 10
  Current State ................................................................................................................ 10
  Plans for 2019/20 ....................................................................................................... 10
  Ambitious Goals .......................................................................................................... 10
  Costs ............................................................................................................................... 11
  The RCP Roadmap in more detail .............................................................................. 11
  Timeline ......................................................................................................................... 14

Keeping Data Moving: the UCL Research Data Network ........................................... 17
  Network context .......................................................................................................... 17
  Research challenges .................................................................................................... 17
  Ambition and goals/technology changes ...................................................................... 17
  Estimated Costs ........................................................................................................... 19
  Research Data Networking Roadmap ......................................................................... 20
  Estimated roadmap timelines ....................................................................................... 21

Optimising the Research Data Lifecycle .................................................................... 24
  Current State ................................................................................................................ 24
  Plans for 2019/20 ....................................................................................................... 24
  Ambitious Goals .......................................................................................................... 24
  Costs ............................................................................................................................... 25
  The RDS Roadmap in more detail ............................................................................... 25

Enabling research using sensitive data .................................................................... 29
  Current State ................................................................................................................ 29
  Plans ............................................................................................................................... 29
  Ambitious Goals .......................................................................................................... 29
  Costs ............................................................................................................................... 30

Software to accelerate research ................................................................................. 32
eResearch Infrastructure ecosystem: Powering UCL’s ambitions

A research data rich, computationally intensive global research environment is essential to accomplish the ambitions outlined in UCL 2034 and the UCL Research Strategy. Technological advances have led to a huge shift towards computationally demanding research activity across all faculties. Data and its analysis can enable cross-disciplinary research between silos, create new disciplines and generate impact on society. Funders are increasingly supporting research that requires computation and data (e.g. BBSRC is moving funding 80% wet lab to 80% in silico research, UKRI calls aligned with the Industrial Strategy’s Grand Challenge of AI and Data, creation of UKRI eInfrastructure Directorate).

Major investment in the eResearch infrastructure ecosystem is imperative. A holistic approach is needed to develop the eResearch infrastructure ecosystem to avoid bottlenecks which compromise academic performance. We are leading in some areas but significantly behind competitors in others. eResearch infrastructure is primarily based within Research IT Services but essential services are also provided by other departments (e.g. ISD Architecture, UCL Library Services, Research Contracts etc). Failure to create this system will compromise UCL’s ability to generate world class research, teaching and socio-economic impact.

Bottlenecks identified during the development of the roadmaps include:

- The disconnect between investment in UCL’s core network and the rapid rise in research data volume and type, leading to difficulties in transferring high volumes of data. This risks the viability of research projects and partnerships.
- The failure of recurrent operational budgets to match increases in the number of users and the number of services provided through capital expenditure. This risks limiting the research output that can be created from UCL’s capital investments.

Our vision is to create a world class ecosystem which optimally supports the breadth of UCL’s academic disciplines and prior experience of using eResearch infrastructure. This ecosystem will:

- Host or facilitate access to compute and data storage resources to meet UCL’s research needs, including national and international facilities (e.g. ARCHER2, PRACE, SUMMIT).
- Create a pipeline to develop hardware and software services to provide experienced users with early access to cutting edge tools and resources for their research.
- Provide low-friction access to computational and data storage services for open and sensitive data, reducing the barrier for researchers to use these resources.
- Upskill and build research communities with the knowledge and knowhow to exploit UCL’s eResearch infrastructure tools and resources.
- Support career pathways and the building of communities of research software engineers, research data professionals and system administrators to enable UCL to better recruit and retain these key personnel. It is evident that UCL’s eResearch infrastructure relies on technical experts.
- Effective coordination across the wider ecosystem (e.g. UCL Professional Services, ISD, UCL Library Services, OVPR) to support and facilitate researchers’ use of data, compute and software.

The specific roadmaps below (see table) provide further detail regarding UCL’s future needs across the ecosystem and they are intended to be updated annually. They are aligned with UKRI’s eInfrastructure roadmaps (currently unpublished).
<table>
<thead>
<tr>
<th>Making the right investments:</th>
<th>Underpinning and enabling activities of capital spend:</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI and Data Science</td>
<td>Software to accelerate research</td>
</tr>
<tr>
<td>Computationally powering UCL’s research</td>
<td>Engaging UCL research communities</td>
</tr>
<tr>
<td>Keeping Data Moving: the UCL Research Data Network</td>
<td>Underpinning applications to support research activities</td>
</tr>
<tr>
<td>Optimising the research data life cycle</td>
<td>Support and training for eResearch skills</td>
</tr>
<tr>
<td>Enabling research using sensitive data</td>
<td></td>
</tr>
</tbody>
</table>
AI and Data Science

This field suffers from an overload of buzzwords. “AI”, “data science”, “machine learning”, “deep learning” – all are hot topics, and all are related. To align with the new pro-Vice Provost for AI role, we use “AI” as a catch-all term for activity in this area, making use of large amounts of data and computational resources to derive insight.

Current State

UCL does not have a coherent AI strategy (although the new pro-Vice Provost is working to change this). There is a team within the UCL Research Software Development group called the “AI Studio” which provides code and algorithmic consultancy for researchers. The Research Computing group have a number of GPUs deployed as part of the Myriad High Throughput Computing service and maintain a number of applications for AI researchers (Tensorflow, PyTorch etc) but the existing service is very much a traditional HPC-style service with GPUs, rather than a service designed specifically around the needs of AI researchers. The Research Data Service provides data storage and repository services for all researchers but contains no provisions for managing common shared AI datasets or any potential regulatory issues with some such data. Similarly, there is a lack of datacentre network performance within the Slough datacentre which may cause issues for the very large datasets used by AI researchers.

UCL is part of the Alan Turing institute but RITS doesn’t make any formal use of this, although we do have informal links.

UCL RITS provides no training around AI.

Plans for 2019/20

1. In late 2019 we will begin the process of designing a coherent set of services for AI researchers:
   - RCNIC will begin evaluating hardware and software solutions to form part of an AI focused compute service with the help of key stakeholders in the research community and locally the AI Studio.
   - The Research Software Development Group will recruit more roles into the AI Studio.
   - The Research Data Services team and the Research Computing team will look at issues around data transfer between the services.

Ambitious Goals

1. A world leading AI platform for researchers that researchers choose to use because it is easy to use, easy to access and cheap.
   - This platform will drive industry engagement with RITS, providing a service for local SMEs to get AI research done.
2. A world leading team of experts within the AI Studio to help researchers to exploit these methods, build new workflows, and incorporate ML methods into existing workflows.
3. Develop open-source ML and AI solutions and services that can apply and scale cross UCL departments.
4. Data storage services which are capable of meeting the performance and regulatory needs of researchers.
5. Fast datacentre networking so that data from goal 3 can be ingested into goal 1 without the network being a significant bottleneck.
6. The best regime for training researchers in ML/AI methods in the UK.
7. A set of data interoperability standards for medical work.

Costs

- World leading AI platform: £2m
• Three additional RSD posts for the AI Studio: £207k pa
  o Do development work as the seed for direct funded posts.
• Two support posts within Research Computing to support the hardware and software of the AI service: £138k pa

The AI Roadmap in more detail

World leading AI platform

At the moment, for complex historical reasons (mostly centred around funding and requirements gathering) UCL lacks a compelling business case and hence platform for AI research, and in reality, it’s not clear that it is well understood as to what this should look like. Currently researchers (including the RITS AI Studio) prefer to use expensive cloud solutions like Azure and Google Cloud when it would be much more cost effective as an institution (4x+) to be using local resources. Building this platform will require engagement with researchers to identify what their needs are in the machine learning space before putting a bid in for funding in 2020/21.

As the intention is for the RCNIC environment to be a test-bed of technologies for new services it may be that the best way to proceed is to design some of its projects such that they can evaluate vendor AI software platforms (e.g IBM’s WLMA, nVidia NGC + DG-X, Cray Urika etc.) and the associated hardware platforms for integration into the Myriad services on “AI accelerator nodes”.

To build a truly world class AI compute platform will require a significant investment in hardware and software technologies.

As well as recouping funding from researchers at UCL and central funding sources, this platform gives us the opportunity to work with local SMEs and create an industrial engagement programme to get external funding from them. This is possible because unlike HPC stacks, AI stacks are almost entirely open source. As well as the research opportunities this provides, activities like this look very attractive to national government sources of funding which could be leveraged to further expand the service.

The software stack and user interface will be as important as the hardware performance. Part of the attraction of cloud services is that they offer a full ‘algorithm in a box’ solution allowing researchers to try different methods on their data via a web interface with minimal expertise – which may often result in sub-optimal solutions! We will need to provide a similarly low barrier to entry, but with expert assistance on hand. The range of tools provided would grow steadily according to the skills of the team and as new solutions are developed, covering areas such as general-use NLP solutions for performing search on textual data, OCR solutions for scanned documents, image processing and recognition, etc.

As well as the compute side this will have to provide suitable storage for working data for AI/ML software both hardware and software presentation with the possibility of supporting Hadoop-like filesystems, tools like Globus for data management and so on. This will also need appropriate interconnects and presentation layers for those, with the potential for tools like rCUDA (which abstracts CUDA calls over an HPC interconnect fabric) to be adopted if suitable.

World leading human support layer, including training

The existing team within the AI Studio will need to grow to help users make use of these new methods. Partly this will be direct funded, but it is likely that this will require some central funding to bootstrap it, as was the original model for the Research Software Development Group. These posts could overlap with the data science support posts proposed for the Data Safe Haven.
The duties of this team would include working closely with the other teams within RITS to glue the services together in a way that allows researchers (and potentially external partners) to use them efficiently. This should include work and consultancy on best practice both technically and in the field of algorithmic ethics.

Work will need to be done developing training materials and courses along the lines of those provided for general research software development. This could include “AI Carpentry” style workshops to orient researchers getting started using AI, as well as longer more in-depth training, particularly making use of local facilities and tools, in collaboration with UCL academics. Care will be taken to supplement rather than compete with existing provision, also bearing in mind support available through the Alan Turing Institute.

**Data interoperability**

UCL works with a number of partner institutions, particularly in the field of medical research. A programme of work will be constructed to design interoperability standards to ease the sharing of data (as appropriate) between these groups and projects, in collaboration with the UCL Clinical Informatics Collaborative. This will then form the basis of part of the data sharing best practice model used by the new AI focused members of the Research Software Development Group.

This will also include work on standards for recording data provenance as it moves through the services.

**Local mirroring of datasets**

Within machine learning, particularly in the image processing sphere, there are a large number of common data sets used to train algorithms. The local mirroring of these would speed up the training and development of algorithms locally as this data would no longer be brought in repeatedly by researchers slowly across the Internet. This service will be integrated into the AI compute service such that its operation is largely invisible to users.

In accordance with the eResearch domain data strategy, we will also look to host centrally commercial datasets for which UCL has licences, and make these available in a controlled fashion to researchers.

**Timeline**

<table>
<thead>
<tr>
<th>PERIOD</th>
<th>AREA</th>
<th>TASK</th>
</tr>
</thead>
<tbody>
<tr>
<td>FY2019/20</td>
<td>AI Platform</td>
<td>RCNIC evaluations of hardware and software technologies</td>
</tr>
<tr>
<td></td>
<td>AI Studio</td>
<td>Recruit posts to help devise the technology stack with RCNIC</td>
</tr>
<tr>
<td>FY2020/21</td>
<td>AI Platform</td>
<td>Deploy initial production AI platform</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hire posts into RC to support platform</td>
</tr>
<tr>
<td></td>
<td>AI Studio</td>
<td>Initial call for projects to kick start the service</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Start process of hiring sustainable, direct funded posts into the team</td>
</tr>
<tr>
<td>FY2021/22</td>
<td>AI Service</td>
<td>Initial Industrial projects with London SMEs</td>
</tr>
</tbody>
</table>
2019/20
- Ongoing HTC GPU expansion on Myriad
- AI Service design
- Hire into the AI studio

2020/21
- AI Service Implementation
- Launch
- Initial projects

2021/22
- Production AI Service
- Industrial engagement
- Direct Funded projects
Computationally powering UCL's research

Current State

We provide a diverse range of research computing services that support users across a range of disciplines from traditional users running large parallel simulations like Chemistry and Physics on our large HPC service (Grace), to Biosciences users running Next Generation Sequencing workloads and new users in the Digital Humanities on our high throughput service (Myriad). In addition to these services we run the Aristotle interactive Linux service for teaching, and an RStudio service for the department of political science. We have limited capacity for GPU users which limits UCL’s work in machine learning.

As well as running UCL services, we run two national Tier 2 services, the Materials and Molecular Modelling Hub (known as Thomas) and the compute service for the Faraday institution (Michael).

The Research Computing Platforms provided centrally consist of:

- Grace: 11K cores, HPC, 1PB of storage.
- Myriad: 3K cores + GPU and high memory nodes, 1 PB of storage.
- Thomas: 17K cores
- Michael: 7.5K cores
- Aristotle teaching cluster is run unfunded on hardware retired from previous services.

Training is currently provided, somewhat ad-hoc, by the operational teams who are overstretched and understaffed.

Plans for 2019/20

1. In late 2019 we will replace Grace with a new service “Kathleen” costing £1.8m.
2. Leveraging a framework agreed as part of the ‘Kathleen’ tender process by which future HPC can be purchased with our supplier HPE without going out to tender.
3. £1M funding to provide a significant upgrade to the Myriad service in terms of storage and core count.
4. Perform a joint R&D exercise with Computer Science (RCNIC).
5. Establish a booking system for Aristotle, to alleviate some, but not all of the issues with the service.

Ambitious Goals

1. Flagship (bring prestige to UCL):
   - Short term:
     - An AI capability that attracts world-class researchers to UCL, our affiliates and to the AI Studio
       - Year 1 start team
       - Year 2 established team with success stories
     - “Top 500” Tier 3/2 system
     - Recognition (of UCL) as a meaningful centre of HPC expertise (based on RCNIC)
   - Long term:
     - A further national (Tier 2 level) deployment.
     - Establishment of an HPC4AI group.
     - >£1M Philanthropic support for HPC.
     - Multiple entries in Top 500 style lists e.g. 1x Top 500 and 1x Green 500
2. Transformative (improve the environment for researchers):
   - Short term:
     - “Low friction” access to compute for researchers
     - Expert trainers + training to match.
     - A world leading support team made of acknowledged experts in their field
   - Long term:
     - “Zero friction” access to compute – it’s just there and works.
     - HPC teaching facility and training.
     - A sustainable facility for development of technologies for new services in new application areas.

Costs

- A Top 500 system costs upwards of ~£8M.
- A Green Top 500 systems costs upwards of ~£2M.
- ‘Low Friction’ access requires mostly integration work but also a small number of servers and will cost ~£0.2M.
- Classroom, online and hybrid Training courses cost on average ~£0.1M each.
- To provide excellent levels of support, the RCP team needs a significant increase in headcount with an additional two RC analysts (1*G8, and 1*G7) and 2 more system administrators (another 1*G8 and 1*G7).

The RCP Roadmap in more detail

World leading HPC

As a world leading institution, UCL should be visibly leading the world in terms of HPC provision. Previous flagship systems at UCL have debuted in the Top 500 list, a list of the fastest 500 recorded scores in the Linpack benchmark. Grace’s replacement, “Kathleen” will be the first UCL flagship service to fail to do so, partly due to slightly reduced funding at UCL for HPC, but mostly due to the rise of Chinese systems in the Top 500 dramatically raising the barrier for entry in terms of cost. While Grace entered costing £3.2M at around #400, the bar has now raised to in excess of £8M capital spent on a system.

Funding will need to be secured either centrally or externally, for both capital and operational costs. If successful, current bids for Tier 2 systems are unlikely to receive the required level (estimates are around £4M for a replacement MMM Hub service). UCL is also chasing further involvement in DIRAC which may bear fruit in terms of additional funding.

An alternative option is to look at metrics which may be cheaper to attempt but more relevant to the needs of researchers. The Top 500 is based around the Linpack benchmark which is not representative of workloads run by actual researchers as it heavily favours pure numerical performance while most problems are memory or I/O bound. As an example, Cambridge recently got considerable positive press from hitting the #1 spot in the IO500 (which measures storage performance) on their Tier 2 service, ahead of national systems costing hundreds of millions of dollars in the US. There are similar lists for Green HPC (the Green 500) which may be of interest and cost considerably less to compete in relative to Top 500 clusters. This would tie in with UCL’s ambitions around becoming carbon neutral and the UK’s ambitions to lead on tackling climate change as well as more generally making our HPC services more sustainable in future.

UCL as a meaningful centre for HPC research
A second pillar of being world leading in the world of HPC is to be recognised as such in developing new technologies and systems. This will be somewhat enabled by the RCNIC initiative with CS but as an institution we need to significantly amplify our promotional efforts about what we do. (An example is that the department of statistical science have an actual FPGA code “VOLNA” which we should be making more noise about and more of an effort to visibly support.) Some of the projects planned around RCNIC, DIRAC and Tier 2s include looking at alternative interconnects, alternative compute architectures and better software provision.

RCNIC will be used as a vehicle for getting external funding for the kind of industrial and nationally funded projects which are difficult to do currently in RITS, for example: hosting hardware for industry funded PhD students outside of the production services. This will increase our value to vendors and these better relationships will lead to early access to hardware, software and potentially funding from the vendors for particular projects.

World leading AI platform

At the moment, for complex historical reasons (mostly centred around funding and requirements gathering) UCL lacks a compelling business case and hence platform for AI research, and in reality, it’s not clear that it is well understood as to what this should look like. Currently researchers (including the RITS AI Studio) prefer to use expensive cloud solutions like Azure when it would be much more cost effective as an institution (4x+) to be using local resources. Building this platform will require engagement with researchers to identify what their needs are in the machine learning space before putting a bid in for funding in 2020/21.

As the intention is for the RCNIC environment to be a test-bed of technologies for new services it may be that the best way to proceed is to design some of its projects such that they can evaluate vendor AI platforms (NEC’s vector systems, IBM’s PowerAI, nVidia NGC + DG-X, Cray Urika etc.) for integration into the Myriad services as “AI accelerator nodes”.

This will link in to related work in the Research Data roadmap and the AI roadmaps.

HPC on sensitive data

A significant hole in capacity within the UK is a lack of platforms for doing effective HPC within a protected environment, suitable for working on sensitive data (such as patient data). Efforts in this regard are discussed as part of the “Data Safe Haven Roadmap”.

Low friction access to compute resources

While UCL’s HPC services are, compared to many other institutions, much easier to use and very well managed, existing systems are designed around our existing long term userbase in the physical sciences. These systems are not friendly or easy to use for new researchers, particularly those from fields like the biosciences and digital humanities where there is not an established base of expertise. There remain issues around providing easy access to resources and so we will bundle resolving these into one vision: “Low friction access to compute resources”.

A new researcher starting at UCL should be able to get access to HPC in a way that removes current technological frustrations, for example access to the researcher’s files. We are working on integration between the HPC/HTC services and the UCL Research Data Storage Service, but this is just a starting point. Users should be able to transfer files to and from the services effortlessly, for example, mounting their /home (and /scratch) directories directly on their local machines to transfer
files rather than needing to learn an SFTP or SCP client. This is technically feasible, but requires us to invest effort implementing Samba services within our clusters.

Similar to file management, the current primary access method to HPC and HTC resources is command-line access via an SSH session and a set of command-line queue management tools. While this is a comfortable environment for experienced users of HPC services (and must be retained), we must also support new users who should not need to become Linux experts in order to get access to HPC resources. There are technologies available to allow us to run virtual desktops, presenting users with a simple drag and drop interface such as Jupyter or Rstudio (combined with file system mounting). We already support some of this in a limited form in the ability of some applications like MATLAB to submit workloads into the clusters on behalf of users and in the provision of Rstudio as a login environment on Myriad for researchers in the Economics department.

We plan to streamline the process by which researchers get access to grant funded equipment, to provide speedier purchasing for HPC equipment to be added to the services, but also a more efficient and cost-effective approach like hosting VMs within ISD’s infrastructure and in the Cloud, for example through the Cloud Labs project, and pre-packaged VMs for data analysis for researchers.

As part of RCNIC and in collaboration with the Research Data team we intend to run a project to evaluate Hadoop, Apache Spark and similar technologies for a future data analytics service.

Excellent training

The current state of training in the use of HPC and HTC services at UCL is somewhat ad-hoc and fairly minimal. In order to address this, we need to look at alternative ways of getting access to trainers with expertise. This might be making use of existing training staff within ISD, or it might alternatively mean bringing in external resource.

In addition, it is necessary to move the existing training platform, Aristotle, from a pilot into full service. As part of this in FY19/20 we have funding for a booking system to allow lecturers to use the platform for undergraduate teaching.

Engagement

As part of being recognised as a world-leading team providing HPC services we need to build recognition of our work both within and outside of UCL. Part of this effort will involve engaging students so that they consider HPC a worthwhile career to build towards. This will involve developing a programme of activities with vendor support to engage students.

We also need to bring in or build expertise in HPC fields which are lacking within the current team, which is heavily Chemistry based.
## Timeline

<table>
<thead>
<tr>
<th>PERIOD</th>
<th>AREA</th>
<th>TASK</th>
</tr>
</thead>
<tbody>
<tr>
<td>FY2019/20</td>
<td>Low Friction HPC</td>
<td>File system mounting + RDS integration</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evaluation of UIs in RCNIC</td>
</tr>
<tr>
<td></td>
<td>Centre of Expertise</td>
<td>Set up RCNIC with initial projects</td>
</tr>
<tr>
<td></td>
<td>AI Platform</td>
<td>Use RCNIC to evaluate AI software stacks</td>
</tr>
<tr>
<td></td>
<td>Training</td>
<td>Aristotle HW refresh and booking system</td>
</tr>
<tr>
<td></td>
<td>HPC provision</td>
<td>Deployment of Kathleen</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Upgrade of Myriad</td>
</tr>
<tr>
<td>FY2020/21</td>
<td>AI Platform</td>
<td>Bid for AI platform + delivery</td>
</tr>
<tr>
<td></td>
<td>HPC provision</td>
<td>New HTC hardware contract for Myriad</td>
</tr>
<tr>
<td></td>
<td>Low Friction HPC</td>
<td>Delivery of UI services based on prototypes deployed in RCNIC</td>
</tr>
<tr>
<td>FY2021/22</td>
<td>HPC provision</td>
<td>Kathleen mid-life upgrade co-design with HPE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Myriad expansion</td>
</tr>
<tr>
<td></td>
<td>AI Platform</td>
<td>Expansion based on demand</td>
</tr>
<tr>
<td></td>
<td>Low Friction HPC</td>
<td>Production HAC VMs for lightweight data analysis</td>
</tr>
<tr>
<td>FY2022/23</td>
<td>Green HPC</td>
<td>Design and build of a “Green HPC system”</td>
</tr>
<tr>
<td>FY2023/24</td>
<td>HPC Provision</td>
<td>Design and build of Kathleen’s replacement, based on output of RCNIC evaluation projects</td>
</tr>
</tbody>
</table>
Tier 2 Phase 2 Project

- Aristotle booking
- Aristotle HW

FS Presentation
- HPC UI prototypes

2019/20

Set up RCNIC collaboration
- AI Service design

2020/21

Ongoing production HPC/HTC expansion
- Myriad framework
- HPC UI prototypes

2021/22

Second year of projects
- AI Service Implementation
- Work with RSD on HAC VMs

Speculative Tier 2 Project
- Launch
- Aristotle HW

Sustainable RCNIC
- Green 500 design

Production HAC VMs
- Launch
2019/20

FDM

Work to retain existing talent

2020/21

FDM

Fill current vacancies

Recruitment

Expanded team

Make case for new posts

ISD Apprentices

2021/22

Student engagement
Keeping Data Moving: the UCL Research Data Network

Network context

UCL Network services provide support to services provided by ISD:RITS such as HPC and Research Data Storage, local research activities and basic commodity supporting services. However the needs of researchers as a whole dwarf those of their professional services colleagues.

The challenge is to maintain consistency for researchers who undertake both research specific activities and other tasks but have much greater demands on scale and speed of provision.

Research challenges

SCALE - Delivering to research scale
Research needs dwarf those of standard use. Recognising research may need 10 or 100x faster, larger provision.

SPEED - Delivering to research timescales
Research requires fast turnaround of experiments and ideas. Understanding that high throughput research requires different approaches when new needs are anticipated.

AGILITY - Delivering to research agility
Research needs are unpredictable. Acknowledging new needs require flexibility and potentially expert consultation quickly.

NATIONAL – Delivering to a bigger picture.
Discerning and responding to the changing national picture and the shifting location of shared services and collaboration.

Ambition and goals/technology changes

<table>
<thead>
<tr>
<th>Ambition</th>
<th>Goals and technology changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>The highest speed network link for researchers in the country at UCL. – get the research data where it is needed, quickly, on the core network and out to the world</td>
<td>Shift to 2x100Gbps for UCL to the Internet – opening the front door to data in and out of UCL</td>
</tr>
<tr>
<td>A Research community prioritised upgrade programme to provide improved and increased capacity to the underlying networking starting across the backbone and to the edge. Using forming national standards as a baseline for all research.</td>
<td>Replacement of backbone and local infrastructure to support 100Gbps/10Gbps where it is needed – opening the internal doors to research data around UCL and between research partners</td>
</tr>
<tr>
<td>Super high speed for research to any desktop or lab – Solving the last mile problem in departments</td>
<td>Audit and replacement of slow less than 1Gbps capable connections - killing off research bottlenecks</td>
</tr>
<tr>
<td>Development of a Network Services catalogue with associated SLAs, including service patterns for specialist</td>
<td>Audit of current areas of high demand where insufficient quality is delivered taking input from researchers. - Targeting the initial roll outs to the right areas</td>
</tr>
</tbody>
</table>
| Requirements to provide clear options for researchers that will meet their needs. | Standard setups for 10Gbps to local HPC or research data generators, available via self-service. Targeted initially at areas of high need and research value.  
- **The fastest speed for equipment for local HPC and data capture equipment**  
  1 Gbps as standard for all researchers with reduced contention at the building or location level  
- **Improved connectivity for all researcher areas**  
  Device support. Ability for devices and “things” to connect to the network  
- **Providing a network for Internet of Things (IoT)** |
|---|---|
| **Research data transfer zones**  
- Networks for HPC and Research data  
Planning for and serving data movement at the > 100Gbps  
Using Researcher journeys, develop improved or new network architecture designs to facilitate large data transfer capabilities at the cutting edge end where current technologies do not satisfy demand. | **100Gbps Zones for data drop off and collection (Globus)**  
- **Planned services for research data transfer**  
Standard “off the shelf” setups for 10Gbps+ to local HPC or research data generators. Split firewalling and supporting local network setups.  
- **Local zones for high bandwidth using researchers**  
Data centre fabric upgrade to provide <100Gbps backbone.  
- **ability to create specific paths to RITS services from UCL locations** |
| **Self-service research networks**  
– speed up ideation to experimentation for research  
Federated management capabilities for local teams for networking services. | **Open up configuration. Through software defined networking allow previously complex network configurations to be self-serviced.**  
- **Deliver research scale and developmental networks quickly**  
Open up monitoring through dashboards and federated reporting.  
- **Deliver improved assurance of quality by publishing real-time speed and quality information** |
| **Research data network consultancy and support**  
– Skilled staff able to consult on the most challenging problems and solutions | Improved network expertise and availability to research departments.  
- **Additional focus on research network** |
Consultancy/skills in local support and central support teams

<table>
<thead>
<tr>
<th>Monitors and dashboards</th>
<th>Consultancy/skills in local support and central support teams</th>
</tr>
</thead>
<tbody>
<tr>
<td>– Be sure the network is operating as you expect</td>
<td>Quality of service established</td>
</tr>
<tr>
<td>Comprehensive monitoring as part of a Network Operations Centre function to provide performance, fault and capacity data to the network teams and relevant information for consumers of their services.</td>
<td>SLAs in place for data transfers with improved or altered SLA for prestigious or high value research locations.</td>
</tr>
<tr>
<td>- Linking network dashboard metrics with Research priorities</td>
<td>Open dashboards to report on network metrics</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Flexible research spaces</th>
<th>Consultancy/skills in local support and central support teams</th>
</tr>
</thead>
<tbody>
<tr>
<td>– research where you want and moved freely from lab, to desk</td>
<td>New campus wireless access points</td>
</tr>
<tr>
<td>Pervasive, fast network services across all UCL buildings and open areas providing both services to UCL community and simple self-service guest service capabilities.</td>
<td>New guest wireless</td>
</tr>
<tr>
<td></td>
<td>Wireless bandwidth monitoring and guarantees</td>
</tr>
</tbody>
</table>

**Estimated Costs**

Note this funding needs to be sought from a number of different sources both locally to UCL and at the national level. The numbers noted below are intended to help the planning of that funding activity.

**Highest speed network**

- Core upgrade and shift to 200Gbps for UCL to the Internet - £3.5m over 2 years
  - Links (Internet and datacentres)
  - Firewalls and other core services
- 100Gbps capable core/distribution upgrade - £5m over 3-5 years

**The last mile to lab or desktop/HPC**

- Audit and replacement of <1Gbps capable switch equipment - £1m over 2 years
- 1 Gbps as standard for all researchers - £2m over 4 years
- Standard “on the shelf” setups for 10Gbps to local HPC or research data generators - £3-5m over 5 years
- Variant option: 10Gbps option presented to every building at UCL – £15-20m over 5 years

**Transfer zones**

- 100Gbps Zones for data drop off and collection (Globus) - £2m over 3 years
- Datacentre backbone upgrade for >100Gbps - £2m over 2 years
• Network digitisation and automation environments with automated workflows - £1m over 3 years

Research consultancy

• Additional skills and roles for local research focussed support professionals
  – *This needs to be assessed as part of other change works*

Monitoring and dashboards and Self-service networks

• Quality of service established - £500k
• Network Operations Centre - £500k over 2 years
  o Open dashboards to report on network metrics
    Opened up monitoring, configuration and access control for local IT specialists

Flexible research spaces

• New campus wireless access points - £2m over 3 years
• Guest Wireless - £100k over 2 years

Research Data Networking Roadmap

UCL has a large network with diverse and demanding business activity. This means it has varied requirements for network services.

It is recognised that digital activities are increasing in many areas within research domains, leading to increased datasets, flows and computational analytics, for instance in the increase in 'Dry Lab' research. UCL Researchers are also continuing to seek out and be part of wide collaboration initiatives.

The implications for the networking service are to provide options for more specialised or higher demand services to the research community in addition to overall improvements to the service.

In order to provide increased capacity and capabilities to research areas and to keep pace with the overall increased demand on the service, an upgrade programme will need to initiate across the core infrastructure and links within the UCL network. The edge connectivity provided by JISC at the campus edge and datacentres will also require corresponding upgrades.

Once this has been progressed to enable greater capacity, a tactical contention programme, targeting specialist requirements for higher connectivity options, e.g. for research equipment that has large data requirements, can commence. We will prioritise this deployment using information gleaned directly from asking the researchers, our own improved monitoring capabilities and the locations of UCL’s strategic research communities. This will not be a static activity but an ongoing programme of improvement and growth. This will prioritise rollout of improvements from the backbone across UCL’s estate to the local networking provided in buildings including hardware and physical cabling.

In addition, large data transfer requirements will be considered, looking at the researcher pathways, so that alternative approaches from standard network architecture, e.g. Research Data Transfer Zone, can be assessed and trialled. A designated zone comprising of the Research Data Transfer
Zone components could be trialled, comprising of high speed data transfer nodes utilising optimal software toolkits, separate firewall/security infrastructure and accompanying monitoring for the service.

An IoT Platform will be developed to support connecting limited function devices on the network with an aim for providing self-service capabilities and suitable security measures.

The wireless service is heavily in demand and is not consistently available across UCL’s sites as deployment has been approached in an unstructured way so an expansion programme has already started with the aims for 100% coverage across UCL buildings and open areas. A self-service guest wireless service is also in the process of being deployed to replace the overly complex previous service.

Monitoring currently is limited in scope and capabilities which hinders the network teams’ and communities ability to pre-empt capacity issues and faults. Networks will collate data on use and capacity of networks into a single place, developing a Network Operations Centre function. This will utilise network performance management solutions to provide baseline data, not only to assess the stability of existing services but to adequately plan for future growth patterns and design scalable systems to support these. This will be used to proactively identify any impacts to research activities and the approach to rectify these including changes to existing or new designs. The networking monitoring service will also deploy real time network monitoring around the campus to display in real-time the health of network consumers and relevant details to support teams and those using the network for research.

Linked with this is the development of comprehensive vendor agnostic management tools for all network services, looking to provide federated, self-service capabilities for local teams, supporting the local but central operating model at UCL.

A review of typical and more specialist network usage patterns and requirements will be undertaken to enable the development of a comprehensive Network Service Catalogue. These will subsequently be used with monitoring and QoS technologies to assure and monitor the quality that is delivered to these types. There will be associated, measurable, defined Service and Operational Level Agreements. The intention is not just to develop a structured, consistent, standard approach to network design and services but also to allow researchers to have understanding of the available services so that they are able to use it to their full advantage.

Underlying these roadmaps are improvements to ways of working, regular evaluation of new technologies and focus on automation to provide more efficiency, agility and flexibility which are key capabilities for meeting the requirements for Research Data Networking.

**Estimated roadmap timelines**

<table>
<thead>
<tr>
<th></th>
<th>Programmes elements starting</th>
<th>Programmes elements completing</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018/2019</td>
<td>• Wireless Expansion- New campus wireless access points</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Guest wireless</td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>QoS Configuration</td>
<td>High Speed Core Network Programme - Internet connectivity, Edge Firewall, Backbone and local distribution zones</td>
</tr>
<tr>
<td>-------</td>
<td>------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>2020/2021</td>
<td>State of the Art Datacentre - 100Gbps support for data transfer/HPC</td>
<td>Prioritised High Speed rollout for Researchers - Standard 10Gbps setups to desktops/research equipment</td>
</tr>
<tr>
<td>2021/2022</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2022/2023</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2023/2024</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Optimising the Research Data Lifecycle

Current State

Research Data Services (RDS) provides two services at present, the Research Data Storage Service (RDSS), and the Research Data Repository (RDR). The latter is provided in conjunction with the Research Data Management team in UCL Library. Although not a formal service, the RDS also offer advice and guidance on good data storage practice.

Research Data Storage Service (RDSS)
The RDSS provides researchers with a large-capacity resilient and backed-up shared storage space to be used for research data. Researchers apply for a storage quota via a web form, which also captures basic metadata about the project to assist with future data curation. The storage may be mounted as a local drive. Non-anonymized personal or sensitive data must be encrypted before being added to the RDSS.

The RDSS is intended to be used during the active phase of a research project whilst data is being worked on. Important data that needs to be published or preserved over long time periods should be moved to the Research Data Repository.

The RDSS currently holds about 2.5 PB of data before duplication.

UCL Research Data Repository (RDR)
Launched in June 2019, the UCL Research Data Repository provides a data publication and preservation platform for significant research data. Researchers can upload datasets via a web interface or app and describe their data using a formal metadata schema. New records are reviewed by Librarians to ensure they meet minimum standards. The Repository enables data sharing in a FAIR manner (Findable, Accessible, Interoperable, and Reusable). It is a key component of UCL’s Open Science infrastructure. The service is not yet certified for sensitive data.

Plans for 2019/20

1. Expand the capacity of the Research Data Storage Service by approximately 2PB to cope with growing demand
2. Improvements to the Research Data Repository to enable sensitive data handling and restricted access (over the next two years)
3. Customized metadata at the Faculty / Departmental level for the Research Data Repository
4. Addition of preservation tools to the Research Data Repository, to ensure the long-term accessibility of data (20+ years)
5. Integration between the Research Data Storage Service and the Research Data Repository, to make it easier to move files from the former to the latter
6. Scoping projects to consider whether UCL would benefit from a centrally provided research database service or an Electronic Lab Notebook service.

Ambitious Goals

1. Full-lifecycle research data provenance - via the compilation of rich metadata and an audit trail of data processing activity we will improve research reproducibility.
2. An integrated suite of services to improve research data management from planning to re-use, underpinned by advice and training.
3. Cost-effective long-term curation and preservation of all kinds of research data, including sensitive data.
4. Enabling new insights from existing research data via the easy discoverability and visualization of data.
5. Combining improved data security with greater research data re-use, via anonymization services and appropriate storage and access mechanisms.

Costs

Research Data Management is a rapidly evolving field in which ‘best practice’ is still being considered and defined. Prof. David Price, UCL’s Vice Provost for Research, has described it as the “major agenda item over the next 30 years of research”. As such, the levels of investment required in order to achieve our ambitions in this field are still becoming apparent.

One inevitable cost associated with Research Data Services is the cost of data storage. At present, we can purchase 2x1PB of storage for approximately £150,000 (our storage is mirrored across two separate sites for resilience), and a storage infrastructure refresh is required every 5-6 years. The next major refresh will be required in 2021/22, when we are likely to move to a hopefully more cost-effective tiered storage model.

Estimated capital funding required over next three years:

- RDSS costs (minus income): c. £1.8m
- RDR costs & development: c. 1.5m
- New projects, integrations, and services: c. £500k - £1.5m

TOTAL: £3.8m to £4.8m spread over 3 years between 2020/21 capital funding round and 2022/23 capital funding round.

Besides capital costs, the implementation of new services may require up to 2.5 additional FTE by 2022 to be paid from recurrent costs. 1.0 FTE at grade 7 is already required to support the technical aspects of the Research Data Repository (along with 2.0 FTE in the Library’s Research Data Management team). A further 0.5 FTE at grade 7 is likely to be required to provide anonymization support; 0.25 FTE to support the putative Research Database Hosting Service; 0.25 FTE to support an institutional ELN system; and potentially another 0.5 FTE at grade 7 to support external data cataloguing and access, customer relations management, and visualization services. More precise estimates of ongoing staffing requirements will be produced during the relevant capital projects.

The RDS Roadmap in more detail

The Research Data Services’ vision is that UCL will offer a suite of services enabling best-practice research data management from project planning through to long-term preservation and re-use. RDS services will be integrated with those of other Research IT Services teams, enabling data to flow between storage and compute and be processed by code developed by the Research Software Development Team. Metadata will be generated and associated with the data at each phase of the research lifecycle, providing rich documentation with minimal overheads for the researchers producing the data.

Metadata will be registered with the services managed by the Research Applications Team, enabling data downloads and citations to be tracked and used to support REF submissions and impact statements.

The integrated services will meet the most stringent funder requirements and enable researchers to produce new and innovative research. The scale and capacity of the services will enable
straightforward big data management; metadata will enable data discovery and access; tools and services built upon the core services will facilitate visualization, re-analysis, and re-combination of data sets to unlock new insights and understanding. UCL researchers will have the tools to manage externally acquired data in the same manner as data generated within the institution – legal rights permitting.

Once the core services are in place, a series of projects will consider the specific requirements of various data use-cases, including sensitive data handling (working in conjunction with the Data Safe Haven team), and ensure that each in turn can be supported centrally by ISD rather than requiring duplication of local departmental IT environments, hardware, and support.

An enhanced research data management consultancy and training service, provided in conjunction with the library, will assist researchers with building good practice into their research from the planning phase onwards, helping inform workflows, technology choices, documentation practices, and ensuring appropriate long-term data preservation and access. Departmental and research-group processes will also be integrated with RITS data storage and archiving services.

The following diagram illustrates the manner in which the RDS team intends to provide and connect data services across the research lifecycle:

In order to realize the vision, RDS has established a number of milestones that need to be reached. Several of these have already been passed, such as the launch of the Research Data Repository; progress towards others is underway; and some remain quite distant. We regularly review the order in which we should aim to meet each milestone based on feedback from researchers, legislation, and the demands of the wider academic environment (funder and publisher requirements for instance).
Milestones that we intend to meet over the next three years include:

**Research Data Storage Service**

- Mounting of RDSS on Myriad (2019)
- Mechanism for transferring data and metadata to the Research Data Repository (2019/20)
- Implementation of ‘facility projects’ alongside standard research projects – for collecting machine-generated data (2020)
- Metadata transfer between RDSS and Myriad – linking input code with output code, the software run on the code, and metadata about the environment in which the code was run (2020)
- Access for non-UCL research collaborators (2020)
- Implementation of Research Data Transfer Zone for fast large-scale data transfers to UCL (using Globus technology) (2020/21)
- local storage cache mechanism for research groups needing to analyse data locally, with automated data feeding into central storage (2020/21)

**Research Data Repository**

- Implementation of data impact metrics (working with Library and OVPR) (2019/20)
- Implementation of selected subject-specific metadata templates (2020)
- Federation of Repository with EUDAT Service (a European research data management service forming part of the European Open Science Cloud) (2020)
- Data curation tools enabling format migration (2020/21)
- Harvesting of datasets by UCL researchers from external repositories & automatic addition of relationships between data and published articles citing that data (2020/21)
- Establish process for periodic review of long-term archived data (2021)

**Services contingent upon scoping studies and/or other groups**

- Addition of a ‘menu’ of data services and chargeable datasets to the Worktribe costing process, (working with the Research Applications Group and the Data Safe Haven team) (2020/21)
- Data for AI platform (working with Research Computing) (2020/21)
- Implementation of pilot research database hosting service (2020/21)
- Implementation of pilot centrally-hosted Electronic Lab Notebook Service (2020/21)
- Hadoop big data platform (working with Research Computing and the RCNIC project, and the Data Safe Haven Team) (2021/2)
- Dark archive (offline, for highly sensitive data) (2021/22)

**Other research data services**

- Development of an anonymization service, consisting of tools, advice and guidance (2020)
- Catalogue of 3rd party datasets to which UCL has access (2020/21)
- Customer Relations Management system for managing strategic data providers (2021)
- Development of a data visualization service, facilitating new insights from existing data (2021+)
Enabling research using sensitive data

Current State

UCL Data Safe Haven service is a secure research network encompassing infrastructure, services and tools for the management and analysis of UCL sensitive-category research data. This service is certified for ISO27001 and meets the requirements of the NHS DSP Toolkit.

The service hosts over 600 UCL research projects/studies, and is a key platform for research data for UCL and partner institutions. Initially scoped to store patient identifiable data handled by the School of Life and Medical Sciences, the service now has additional applicability for UCL research, holding data ranging from pupil-level information arising from educational interventions, to data from industry released to UCL for research purposes, under data sharing agreements.

Plans

In FY19/20, ISD is running a capital project to refresh and enhance functionality of this service, with £1M budgeted to renew infrastructure and expand the service components to include HPC/HTC, data science tools, and GPU in support of emergent machine learning workloads.

By virtue of this project, UCL will achieve a significantly improved service, made of modern components and addressing many of the additional services that our researchers require, placing the service on an excellent footing for the next five years, while adding horizontal scalability to be able to grow the service in response to research demand and funding.

Ambitious Goals

There are the ambitions of the capital project through to July 2020 which will introduce a service that is world-class today. However, the infrastructure and services sides of platforms for sensitive data are a moving target, so this is an area that will continue to evolve rapidly, particularly in the software space and in the respect of data storage requirements.

1. Develop and grow the service components introduced under the capital project in the next year, providing plentiful data processing capability and petascale storage for sensitive data, doubling or trebling current computational capability and active storage capacity.
2. Introduce Hadoop/HDFS services for advanced analytics, and develop a new service in this area, aligning the design with any service to be offered on the Research Data Storage Service platform for non-sensitive Big Data projects.
3. Within the next two or three years it is anticipated that machine learning products will be more freely available to be consumed in a ‘productised’ form, with likely general-purpose applicability for UCL workloads involving sensitive data. In addition, signs point to the robotic process automation (RPA) market maturing and it may therefore also be valuable to investigate this product segment for users of Data Safe Haven.1 Further, we will ensure that due consideration is given to Deep Learning, and UCL’s ability to exploit AI developments in this area for our varied research community handling sensitive data. We expect deep learning to become increasingly important for clinical research, in particular.
4. Introduce rapid provisioning and deprovisioning mechanisms, to a similar level of responsiveness to public cloud, and at a similar cost to the researcher, but with a

higher level of assurance provided by virtue of the detailed, research-specific Information Governance process already embedded in UCL’s Data Safe Haven service.

5. Implement an archive/data-mover service to support the long-term retention of sensitive data sets. Introduce the same kind of data classification and meta-data as the UCL RDSS (Research Data Storage Services). The data management capability of the Data Safe Haven service will need to be developed, to include the same kind of meta-data as RDSS repository, either through investment in automated data classification products, or potentially through a significant specialist staffing function involved in statistical disclosure control functions akin to those operated by the ONS or the UK Data Service2. These activities will support UCL’s ambition to be a leader in data preservation, scientific reproducibility, and advancement of FAIR data principles.

6. Mapping UCL research activities and associated grant funding and publication/open science metrics, enabling a detailed picture of the value of the service in respect of the UCL research portfolio. If data were available on other UK HEIs and their equivalent services for handling sensitive data, this would also provide a meaningful point of comparison. In addition, where funding applications are being made to funding councils such as ESRC, guidance should be developed in order to help researchers leverage funding for data ingestion costs associated with sensitive category data.

7. Investigate the feasibility of applying ISO27001 expertise to other UCL ISD research related platforms in order to support a continuum of research data at different levels of sensitivity.

It is recommended that these additional considerations be addressed in the 3 years following completion of the Data Safe Haven Refresh project, which runs to July 2020. The strategic value is closely aligned with the UCL Research Strategy:

- Objective 2.2 – Build and support partnerships with other HEIs, hospital trusts, and data providers
- Objective 3.1 – Champion a broad definition of research impact. There is strong evidence of public value derived from research made possible by the suitable handling of sensitive data, particularly in the spheres of healthcare, and education research
- Continuing to invest in developing the UCL Data Safe Haven as a service, in both staffing expertise, hardware, and software capabilities. This supports the UCL 2034 theme D ‘Delivering excellent systems and processes’, and this platform is critical to the operational research activities that contribute to UCL’s Research Excellence Framework submission in 2021

**Costs**

- Approximately ~£2 M of capital for additional hardware could be consumed in delivery of the ambitions outlined above
- Additional spend on software may be required, likely as an increasing proportion of the solution (licensing makes up a small proportion of the current cost of running the service, since much off software tooling is covered under existing campus-wide licenses). As the Data Safe Haven service expands, so may its software portfolio. An additional investment of around £150,000 per annum (Autumn 2020 – Autumn 2022) could be considered to cover enterprise licensing for a variety of software that has been investigated during the project, but was not within-budget (e.g. Anaconda)
- As the DSH platform grows as described above, it is estimated that up to 3*FTE would be required to deliver the services, costing an additional £180,000 per annum
The potential savings arising from this would be in services and staffing used in the delivery of external services that provide similar functionality to UCL Data Safe Haven, which may move as workloads to DSH. In addition, there is an opportunity-cost element to be considered, were sufficient data to be available to indicate those research grants that would not have been won, were it not for the service benefits delivered arising from this investment.
Software to accelerate research

The RSDG helps researchers to produce high quality research software, from the simplest scripts to complex simulations running on state-of-the-art supercomputers. Our Research Software Engineers (RSEs) aim to make UCL computational research world-leading in its impact and reproducibility, supporting the open science agenda, through training researchers and through direct collaborations.

Current State

- The RSDG was the first such group in the world supporting an entire university, and remains at the forefront of provision of this service to researchers.
- We have developed a scalable model of delivery aimed at meeting the large amount of currently unmet demand in an agile responsive manner. The group has grown from 3 FTE in 2013, through 8 FTE in 2016, to 19 FTE in December 2019, with additional recruitment ongoing as demand continues to rise faster than supply.
- With the UCL Centre for Data Science we launched a Data Science consultancy role in 2018, branded as the “AI Studio”. This will be expanded alongside other specialist research software development provision, to provide UCL with an efficient means by which researchers are able fully to exploit the data landscape.
- We deliver in-person training covering from the basics of programming for research to more advanced research software engineering practices, and have collaborated with Digital Education to develop self-directed, online versions of our most popular courses. In partnership with relevant parties within and beyond UCL we will continue to review and expand our portfolio.

Ambitious Goals

1. To establish ourselves as a recognised world-class RSE operation, giving UCL a reputation as a reliable source of high-quality research software that other researchers can build upon, and that achieves measurable impacts (e.g. in REF returns).
2. To seed ‘satellite’ groups based in faculties, including: SLMS, MAPS, Engineering and Digital Humanities.
3. To establish a new ‘HPC for AI’ specialist group in collaboration with research computing and research data services. (See also the AI and Data Science chapter above.)
4. Including these groups, RITS to provide 100 FTE research software and data science specialists by the end of 2022.
5. To partner with (a) leading external institution(s) for sharing RSE skills, resource and support, providing the greatest possible cost benefit to UCL researchers.
6. To provide easily deployed and used end-to-end solutions for researchers in specific domains (e.g. molecular dynamics, data science, clinical trials, digital humanities), together with other RITS service groups.
7. To dramatically expand the training provision available to UCL researchers, in collaboration with academic partners across UCL.
8. To develop world-leading career opportunities for research software specialists at UCL, blending professional services and academic roles, able to migrate smoothly between research and industrial worlds.
Costs

- An advantage of the Research Software Development service is that it requires little core funding, being primarily supported through research grant income from the projects on which we collaborate. In addition, indirect income associated with these posts can (and does) further subsidise other RITS activities, including administrative overheads, free ‘drop-in’ support, training, taster projects for groups with limited access to funding, and joint projects with other RITS teams, according to strategic priorities.
- Growth within RSDG thus has potential to reduce demand on central funding. The key challenge is allocating physical space to enable growth.
- However, capital funding can accelerate delivery, particularly in areas such as creating new training courses and developing a consistent and interoperable ‘RITS platform’ of end-to-end solutions.
- We recommend at least the creation of a new permanent grade 9 ‘Head of Research IT Training’ position to lead activity in that area (see also the Training chapter below). This would cost £81,832 per annum at the current pay scales (spine point 49).

The RSDG Roadmap in more detail

The Research Software Development Group has seen dramatic growth in demand since it was created in 2012, with an average annual income growth of about 50% during that period – despite no growth in core funding. If anything, this rate of growth is increasing; however, we are only forecasting a risk-averse 10% future growth at present for hiring purposes. Assuming an average 30% success rate on current grant proposals, we expect to receive at least £1 million in the 2019/20 financial year.

This growth is largely organic, driven by increased awareness of our existence and word-of-mouth advertising from pleased collaborators. It reflects a recognition across UCL that our involvement on projects improves the reliability, reproducibility and impact of research software, reducing the bar to translation. We will continue to innovate and improve our ways of working to ensure that we can provide a world-leading service to UCL researchers, and thereby ensure that UCL research software ranks highly in exercises such as REF (goal 1). Measuring ‘world leading’ in this space is difficult, with all metrics proposed to date being rather poor proxies and too easy to game – there isn’t an equivalent of the HPC Top500 chart (which itself has limitations)! Essentially it means that others are looking to copy us (which they are) and that the research we support is ranked as world-leading, including UCL research software itself being highly reused and cited.

As part of this goal, we are working with the Research Applications team in RITS on the open science strategy and architecture design, both to promote the concept of research software as a REF output,
and to ensure its place in UCL’s research management toolset. The prototype Research Software Dashboard we developed in 2015 looked at features that could be useful in this space, and we envisage retiring it in favour of a new integrated approach post REF 2021 (see also the Research Applications Roadmap).

Being world-leading further includes facets such as: representation on the trustee board for the newly launched Society of Research Software Engineering; visiting RSEs coming from across the world to learn from and work with us; securing RSE fellowships; being a leading player in RSE international conferences and initiatives. We would also like to be recognised internationally as the maintainers for some core software underpinning significant areas of research. As a first step in this direction we are seeking funding to lead maintenance of the libSBML software that supports much computational systems biology research. While commensurate with the Open Science agenda we favour open source licensing whenever possible, this may still bring opportunities to offer commercial support options for research software through UCL-B.

**Goals 2–4** relate to managing the ongoing growth in demand effectively. This requires novel approaches, not simply hiring more staff within a generalist central group. We are therefore developing strategic collaborations to create satellite groups and sub-groups, focused either on particular research domains or technical specialisms. Potential examples are of the former kind are medical informatics (where we have a close collaboration with UCLH on the Experimental Medical Applications Platform development and related research projects) and medical imaging. We are in the process of creating a sub-group in this area along with the Centre for Medical Imaging Computing, the Clinical Research Informatics Unit, and others. In contrast, we already have a technical specialist sub-group in the AI Studio, which will continue to work closely with UCL’s AI Network, the pro-Vice Provost for AI, and local partners in supporting researchers to employ modern data science approaches in their work. We will expand this team with the additional skill sets required. Working jointly with Research Computing we will create an HPC/HTC focused sub-group within two years, with a particular emphasis on AI applications.

We are also investigating partnerships with external providers who can deliver the more mundane aspects of projects more cheaply than local staff, and/or can provide specific technical expertise that is not available in-house. Web development may be an area where this approach is particularly effective, with RSDG eliciting research requirements and building computational backends, UI/UX design done in collaboration with ISD Digital Presence, and front-end website builds by external companies. There may also be scope to partner with (an) other academic institution(s) elsewhere in the world to share in the load (**goal 5**). This would need to be handled extremely sensitively given the bad press typically associated with outsourcing! Careful management would be needed both in selecting the partner(s) and assigning work, to ensure quality of outcomes. This would therefore not be pursued significantly without extensive investigation of track record and cautious trial periods. As an alternative to ‘outsourcing’, RSE groups within the UK have been discussing for some time the idea of work sharing or pooling; exploiting specific expertise developed by particular groups or passing work to an undersubscribed group to cover exceptionally busy periods. There are considerable contractual challenges associated with such arrangements, as the Turing Institute have found – including the practical difficulty that we are all currently oversubscribed!

As a knowledge-based team, our impact is largely driven by the quality of the staff we hire and the expertise they both bring with them and develop while here. Expanding the team increases what we can do, but we want to do better than this purely linear growth. Several avenues are being explored.
One is to improve mentoring and training for more junior members, through enhanced capacity and coaching/mentoring skills in the senior team. Reusing knowledge or tools from previous projects is another mechanism to make progress faster. The bespoke nature of much of our work is a challenge here, but we are always on the lookout for opportunities to extract and share common tools, libraries, or just ways of working on particular problems. Collaboration with researchers working at the forefront of software engineering, AI, etc. is important here.

Our research software development expertise is a “force multiplier” enhancing the benefit to UCL from investment in hardware and other RITS services, through enabling researchers to make more impressive use of these services. Together with other teams we aim to develop a seamless researcher experience across the ‘RITS platform’ of interoperable services, providing a sector-leading Research IT environment (goal 6).

One aspect of this is the planned “Automated Cloud Hosting Service for Research Applications” service, which will assist researchers to build and deploy web applications. In the 2019/20 financial year we will undertake detailed planning to scope and cost the service, prior to launch the following year. This will aim to build on ISD’s new Hybrid Automated Cloud platform to provide cost-effective long-term hosting for research outputs designed to be interactive online resources. There are thus links with both Research Computing and Research Data services here.

Increasing impact also comes through training researchers themselves to build better quality software (goal 7) – and indeed this has the greatest long-term impact relative to effort invested. UCL provision generally however is often piecemeal and delivered by many different groups. We have done significant work already in this area, e.g. through short Software Carpentry workshops and our longer Research Software Engineering with Python course, and have ongoing collaborations with some academic departments and UCL CDTs to provide training, as well as getting involved in new initiatives such as Clinician Coders and cross-London shared Software Carpentry workshops. During the coming period we will work with these and other interested parties (e.g. the e-Research domain, ISD Digital Education) to develop a comprehensive shared curriculum in computational and data-intensive research methods, where our students benefit from hands-on experience. These will be targeted more broadly than the current focus on CPD for research staff and students. There are opportunities to stratify training provision according to different levels of need, for instance running discipline-specific courses, summer schools, etc. This activity will also involve RITS staff beyond RSDG (and staff beyond RITS). The outcomes from this process should include (i) the creation of courses that count as formal credits for UCL students, both at MSc and doctoral levels (e.g. through the Doctoral School and CDTs; this would also represent another funding stream), and (ii) selling courses externally to raise income and enhance our reputation. Exploiting the full RITS expertise will enable us to position these courses uniquely within the market (for instance AI for healthcare using HPC within the Data Safe Haven). We aim to develop at least one new course each year.

This is the area where enhanced capital funding can really make a difference to prioritise these activities, and academic buy-in from across the university can ensure that solutions are joined up, exploiting relevant expertise from across UCL’s many departments, rather than working in isolation. To ensure progress is strategic and coordinated, we recommend at least the creation of a new permanent grade 9 ‘Head of Research IT Training’ position to lead activity in this area.

Finally, goal 8 underpins our ability to continue to attract and retain the best staff in this extremely competitive area. One challenge in creating satellite groups jointly with other departments is being
able to offer equivalent terms and conditions to staff, particularly recruiting to permanent contracts. Our track record within RSDG and robust financial forecasting model means we can get approval to hire permanent staff on the basis of projected future income; to date this has not been possible for jointly appointed staff. The group head is therefore working with Fiona Ryland and colleagues in central professional services (PS) units to export our model for use by other ‘collaborative service groups’ (both those directly engaged in research and more widely in UCL) and so streamline collaborations. We also expect that not all RSE-like activity within UCL will be undertaken by our group; some will be appointed by departments, and some will be primarily aiming at an academic career track but with significant software development activity (and indeed, as the creation of software becomes more prevalent and software training improves, this category is likely to expand). We want to ensure that RSE-like roles can benefit from the same job security as RSDG provides, and that RSEs wherever they are based can benefit from the same progression opportunities. This supplements engagement with UCL’s Career Pathways initiative. We aim to provide world leading RSE career development options, with staff able to move flexibly between teaching, research and PS roles (or a combination thereof!) in any department, into industry and back, with progression possible on technical excellence not just management responsibility.
Engaging UCL research communities

Develop a rich, vibrant and engaged community across ISD and academic departments ensuring that the eResearch infrastructure ecosystem is tailored to the researchers’ current and future needs. The engagement staff across UCL should be seen as the first point of contact for all UCL’s research IT needs. To enable this, three streams activities have been identified;

**INFORM:** Engage with potential and existing stakeholders, through personal interaction, group events and online channels to identify and cultivate new networks and partnerships to raise awareness of RITS services (and beyond) and increase uptake of these services

**SUPPORT:** help researchers to solve research IT problems, providing a holistic overview to service provision, sign-posting support where appropriate, gain knowledge of researcher projects and aspirations, and broader strategic plans, to help improve existing services and develop new ones.

**CONSULT:** Engage with the academic community to understand future research needs and to gain valuable feedback providing opportunities for continual service improvement, gaps in service and opportunities

The engagement roadmap takes into account the fact that opportunities exist to improve and widen communication and engagement across the wider RITS team, facilitation roles and governance established at UCL.

**Current State**

- RRIISG engages with researchers via its members and indirectly via the RIIISG Subgroups: Research Delivery/e-Research Domain, Research Outputs & Research Management.
- RITS proposes capital project ideas which may be developed into full proposals by the respective committee and these proposals are then prioritised by the sub-group committee and then RRIISG. There is little consultation with the wider
- Research community regarding capital budget investments, or communication about the progress of the capital projects. Researchers are only engaged by project managers to provide input during the requirements gathering phase and user testing phase of capital projects.
- With regard to engagement activities, the RITS engagement team (RITS and IT for SLMS facilitator) are school aligned and work closely with other professional services staff to support research activities across UCL. The role of the RITS facilitators is seen to;
  - Direct communication with researchers, with a particular focus on increasing awareness, uptake, impact and support for RITS and Library Services. The team work closely with the Library Data Management team across the services.
  - Communications channels include a mailing list, printed material, two websites (one serving an external audience, and one providing user support)
  - Attending and networking at key events (internal and external), understanding and promoting external services that researchers have access to, liaising with research IT professionals from other HEIs to share best practice
  - 1:2:1 Support for researchers include fortnightly drop-in sessions and email support via Remedyforce
  - Targeted presentations are delivered to a range of audiences (e.g. at student inductions, departmental meetings, special interest groups) to inform and support researchers with research IT related activities.
  - A range of training courses are provided to support RITS services e.g. software carpentry workshops. The training is responsive based on the current need and the
team works across UCL to engage with those who would not normally attend these courses and would not have historically attended e.g. SHS

- Translating user needs by signposting to relevant IT services, identifying opportunities and prospective users for new services.
- Communications are also effectively channelled through the IT Service Partners, Heads of Faculty IT, and all RITS staff in the course of their usual activities.
- The RITS Annual Survey attempts to monitor awareness and satisfaction with our services, but this needs reworking to make the information gathered more relevant and actionable to enable Service Owners to respond, plan service improvements and report on as part of their metrics.
- Ongoing challenges retaining and recruiting staff for PS and research with e-Research skills.

**What has changed**

- UCL has experienced an unprecedented degree of growth, with more than 12,000 academics and researchers, compared to fewer than 6,500 in 2007.
- Technological advances have expanded the number of disciplines which use data and associated e-Research methodologies. The different academic disciplines have varying levels of maturity in using HPC, data science, research data management and software development.
- The introduction of GDPR has increased awareness of issues such as information governance, research data management, and security across the board, resulting in increased demand for joined-up advice and technological solutions.
- Research increasingly involves e-Research approaches, often difficult to identify and quantify, however still lead to increased demand for Research IT Services. Researchers need support in defining, justifying and costing these services in research bids. RITS have introduced cost recovery mechanisms, augmenting investment from central funding. Although cost recovery mechanisms have been introduced, more work is required to clarify the internal processes for handling these requests and developing the online guidance.
- The introduction of the ISD digital masterplan has a number of implications for engagement:
  - Objective 2 Research IT: “ISD will continue to meld together services at the national, regional and discipline specific level as well as new commercial cloud offerings to provide an easy to use one-stop shop for researchers”.
  - Objective 2 Research IT: Promoting world-class research through IT education. Focuses on students, but there’s no reason staff couldn’t also benefit.
  - Objective 2 Research IT: Supporting data-driven science and open science.
  - Objective 2 Research IT: Managing research from Opportunity to Impact.
  - Objective 3 Digital Transformation: proposes an integrated, task-based UCL intranet supporting PS functions, whilst also aiming to have a world-class external web presence. This will impact our decisions around how our web presence can be managed in the future.
  - Objective 4: Proposes a move towards well integrated systems that will enable a more evidence led approach to decision making.
  - Objective 5 Information Security: Increase awareness and provide platforms to help researchers handle sensitive research data.

**Ambitious Goals**

1. The ultimate goal, is to develop a research engagement strategy across UCL to outline how we will inform, support, and consult with the UCL research community. Aim is to include engagement/facilitation roles across UCL in one cohesive service offering with structure and
governance developed to enable this. The success of the strategy will require effective consultation with research community to understand business processes and support of the research lifecycle. (look to incorporate findings from the current Research Support Transformer)

2. Before a UCL wide research engagement strategy can be delivered a review of the current RITS engagement function defining what success looks like, redesigning the operating model and how the effectiveness of engagement approaches will be monitored and assessed.

3. Define a communications strategy that will underpin the engagement activities, defining the levels of engagement and the appropriate activities. The three identified streams will be improved upon by;

**INFORM:**
Communicate and promote levels of investment in capital projects and services demonstrating commitment by senior management at UCL.

- Provide monthly updates on all RITS projects via the RITS website, showing deliverables and business benefits.
- RITS facilitators will act as senior user on ISD project boards, engaging with academic senior users, communicating on and championing the services e.g. help coordinate and publicise good news more effectively. It will enable us to develop reputation for investment in innovative and world-leading IT systems that underpin research and impact.

Promote Research IT Services internally and externally finding opportunities to expand into new areas.

- Continue to showcase online and through newsletters the impact of services, and collaborations between researchers, departmental staff and central PS staff in data, research networks, hardware and software research projects. This will inspire the community to use the services and infrastructure.
- Grow UCL’s internal and external reputation that it is leading in the development of the next generation of research IT infrastructure and related services by encouraging members of RITS to contribute regular opinion pieces showing “Thought Leadership” in their field. (Buy-in is needed from the RITS heads of making it becomes part of expectations, e.g. to write up their view of recent conference, etc)
- Develop coherent website and aligned with TOPS transformer website project and the ISD Digital Masterplan. This will involve collaboration with ISD Webcomms and the Intranet project to prepare content for the new platforms.
- Develop and publish a communications strategy aligned with the engagement roadmap which also addresses questions like - how do we promote the RITS brand.

**SUPPORT:**
Review and improve the support for research activities finding opportunities to access the PI earlier on in the research lifecycle.

- Develop working relationships with other facilitation/support teams e.g. professional services, library, local support to develop new forms of joined up support e.g. PI induction sessions, training and web content.
- Extend drop-ins to other ISD colleagues’ e.g. Hybrid and compute and other professional service teams focused on data, compute and software.
- Alignment/collaboration with faculty run drop-ins e.g. SLASH, IOE
- Cross promote other PS drop-ins e.g. research services.
• Review the current format of the drop-in sessions with a view to increasing uptake; consider for example, the hours, frequency, coverage and marketing. Promote the space a resource to all researchers work alongside RITS in ensuring all IT needs are being meet.

**Identify the appropriate support for the entire research community taking into consideration those who are not traditionally RITS service users, e.g., “low friction” HPC or new and emerging centres/depts, e.g., UCL Centre for AI**

• Define a stakeholder engagement strategy that provide an inclusive and continuous process for engagements, encompassing a range of activities and approaches to UCL’s research community

• Develop a framework for building a stakeholder map covering the entire UCL research community, showing which communities need which type of support. Populate the map with insights gained through feedback exercises such as the annual survey and direct engagement with researchers.

• Tailor communications to specific communities based on the stakeholder map. (Inform activity)

• Issue quarterly Research IT briefings aligned to researchers needs, e.g., Big Data, discipline specific alongside the eResearch domain.

• Start a series of ‘How to’ to help researchers get the best out of their IT understanding the benefits and the ‘why’ e.g. blogs or lunchtime sessions.

**Provide additional support and advocate for research communities which have experienced a significant recent technological advancement in their field. Enable the researchers to adapt and address data or computational challenges quickly**

• Facilitate sessions with RITS service owners to map needs and translate into mini projects

**Sharing best practice and engaging with technical experts communities**

• Engaging with relevant Communities of Practice (CoP) to gain ideas and feedback from the wider community and to engage staff in related change projects (e.g. Data and Insights CoP, Research IT CoP and the Research Management subgroup),

• Continue to run the Research Programming Hub e.g. tech-socials, SLACK, coffee mornings.

• Exploring additional technical communities such as Data Professionals, Data Managers, Systems Administrators/operations and Special interests groups e.g. AI, HPC, Research Data

**CONSULT:**

**Standardise processes for collecting, reporting and acting upon feedback received from the research community**

• Create and promote point in time avenues for on-line/fac to face feedback, look at good practice in this area e.g. you said, we did approach.

• Conduct stakeholder interviews with a research focus (joint Library Services & RITS) e.g. early career researchers.

**Improve the quality of feedback across RITS portfolio of services**

• Redesign the existing annual survey in collaboration with Service Owners to improve the quality of feedback received, whilst ensuring backwards compatibility with previous survey results as far as possible. Aim to improve circulation, inclusiveness, and relevance. Look to consult with researchers more generally about their activities and how fit for purpose our services are when it comes to supporting those activities.

• Collaborate with JISC on the research on-line survey, contributing to the design of and take part in the pilot for 2019/20

**Pro-active development and agile delivery of new services.**

• Capture customer service requirements throughout the year. Have an annual process/ to elicit capital project ideas from the research community that can feed into the planning cycle in the
Respectful RIISG subgroups. Provides the opportunity to identify new opportunities and challenges and to identify a community who will champion the capital projects as it is developed.

- Host systems thinking workshops (involving users and professional service teams) to identify capital projects, changes to processes or services that could increase the benefit of existing services and infrastructure. These may need incentives and expertise from other ISD colleagues to facilitate. Review tools that can support this activity e.g. thought exchange - https://www.thoughtexchange.com
- Communicate on contingencies / risks delivering capital projects. Run a ‘co-creation’ activity with a friendly research group on one or more new services, e.g. DMZ and Research Computing and Networking Working Group (RCNIC)

Horizon scanning activities with faculties, departments and other initiatives (i.e. Institutes, science technology platforms, RICS, bibliometrics) to identify their strategic direction and future needs in response to their priorities and anticipated technological advances.

- Translate relevant research ambitions from the Faculty Operating Plans (SoP) into Faculty Research IT Strategic plans
- Produce a horizon scanning report per faculty/dept
- Increase the number of faculties and departments that have included RITS in their multiyear strategic plans, and vice versa.

Over the next year (2020), redefining the RITS engagement model and with the support from UCL senior stakeholders to provide a cohesive support model across all of UCL’s engagement/facilitation roles, will see this roadmap evolve and more detail on the specifics of the ambitious goals will be provided.

Costs
- With an ambitious goal of merging engagement/facilitation roles across UCL in one cohesive service offering and with the work to redefine the RITS engagement model this is likely to see an increase the number of research facilitators across UCL. This number currently sits at 3 FTE but plans to see at least 6 facilitators to accommodate the vast numbers of researchers across multiple disciplines.
- Website redevelopment (including imagery) and case studies – 25K
- One G8 (1FTE) 2yr post to support and develop new processes outlined in the roadmap including comms and web skills that would be embedded into business as usual.
Underpinning applications to support research activities

Current State

The Research Applications group provides several services in conjunction with OVPR and Research Services to support UCL's research management activities. These include the UCL Library to support the management and discovery of Research Publications (RPS and UCL Discovery) and Profiles (IRIS); Research Services to support the entire project lifecycle (Worktribe), The Doctoral School and HR to support doctoral students' project activities, skills assessment and training (Research Log and Research Skills). There is also some legacy systems; Equipment (KIT) Catalogue and Research Impact Curation (RIC) Database.

- **Research Publications (RPS and UCL Discovery)** - an online system providing a central internal hub for details of all UCL research publications. RPS is also the principle route for depositing full text to UCL Discovery (eprints), one of UCL's institutional repositories.
- **Research Profile (IRIS)** - is the research portal for UCL. One-stop shop for information on research activity at UCL. It brings together data from Human Resources, Finance, Student Records and Research Publishing systems.
- **Research Grant Costing (Worktribe)** - a holistic research management system that supports the whole project lifecycle from proposal development to post-award management. This service fully integrates with UCL finance system and provides rich management data to the Data and Insight programme to provide new research planning and forecasting tools.
- **Research Management Log and Research Skills** - used for recording details related to a student's graduate research programme, scheduled supervisory meetings and activities concerning the development of academic and key skills. It is a mandatory component of all UCL research degree programmes. The Research Skills system is an online booking system for the Doctoral Skills Development Programme; a series of online modules delivered in Moodle covering a range of research skills important to the early career researcher.
- **Equipment (KIT) Catalogue** is a searchable catalogue of UCL's significant research equipment. This catalogue encourages equipment sharing and collaboration by making visible our research technologies/capabilities across all faculties at UCL.

The Research Applications roadmap will align closely with Research Data and Research Software in particular the management of outputs and support the improvements to metadata capture across the lifecycle as defined in the Research Data Services roadmap.

The research applications portfolio is governed by the Research Applications Operations Group (RAOG) comprising of ISD and senior representation from each business area. The meeting provides oversight of current work and provides an opportunity to set priorities.

Plans for 2019/20

We have funding for 5 specific areas;

- **Research Applications Architecture aligned with Open Science** - Develop a target architecture model for the research applications domain to underpin a roadmap for the digitisation of research management process that integrates and simplifies to deliver an excellent researcher experience. This suite of applications will also support and align with Open Science policies and practices. The remit will also extend beyond the original scope of Outputs and Profiles to include all research applications, providing a fuller picture and steer for possible changes to the current research application portfolio.
- **REF Preparedness** - Accountable for the technical deliver of REF 2021. This includes managing all RPS upgrades, addressing all reporting requirements and responsible for the data submission to Research England in November 2020. There activities are governed by a set of REF technical working groups reporting to the REF Strategy Group chaired by Vice Provost for Research, David Price.

- **New Ethics and Data Protection Service** - Creation of a new service to support UCL’s Ethics Review and Data Protection business processes. This year will see the tender and procurement of a new system, development and rollout to the existing central Research Ethics Committee and Data Protection team.

- **Digital Accessibility for Research IT Applications** - fix inaccessible digital content on the website, online learning and Research IT systems in response to the Public Sector Accessibility Regulations 2018. This work will be aligned with the Digital Education project, whose aim is to provide a steer and champion this work across UCL.

- **Open Science Online Training course** (on hold) – providing support to UCL Library to adapt existing provision into an online course, covering aspects of Open Science.

**Ambitious Goals**

1. Support UCL as being the recognised global leader of Open Science and Open Data practices providing the right capabilities/tools/environments to support this;
2. Support Open Access and Plan S policies and practices e.g. promote the use of publications standards i.e. ORCID for global identification. Although the future of Plan S is still not clear there may be opportunity to align services with suggested good practice
3. Providing transparency and high levels of reproducibility of research outputs in particular Research publications (alongside research data and software). Agree a new research applications architecture to meet required capabilities, ensuring the plan is funded and delivered. These would include but not exhaustive (yet to be fully defined and agreed)
   a. Agreed single deposit for all research outputs. Showcase and connecting UCL research and researchers (research staff and students), providing a means of connecting through a search engine optimised route.
   b. Surface management information to include tools for analysis and data visualisation and presentation
   c. Provide self-service enablement - single point of entry for editing data
   d. Support collaborative working, finding collaborative opportunities as well as funding opportunities.
   e. In principle ensure all tools are interoperable, quick, have built-in visualisations & analysis tools, as well as API endpoints so that others can program against them to do e.g. machine learning/mining e.g. for collaborations
4. Ensure UCL makes an optimal submission to REF2021 that will be reputationally and financially beneficial to the institution e.g. ensure no technical ‘unclassified’, correctly identify our most excellent independent researchers, go above and beyond OA requirements, smart use of 5% tolerance and submit best research of former staff
5. Support UCL in meeting statutory reporting requirements e.g. REF, KEF and funder requirements i.e. researchfish. This includes supporting data identification and collation, and data manipulation and visualisation and reuse of data from external sources.
6. Support IT data capture at pre-award stage to help with IT planning and forecasting. Include costs for HPC, RDS, DSH and IT generally. Work with the engagement team to use this as a means of first point of contact to all PIs
**Costs**

For capital funding over the next 3 years

- **REF - testing and submission to Research England (Nov 2020)**
  - 2020/21 £150-200k
- **Research Applications landscape beyond REF - for procurement activities, development and delivery of a suite of services**
  - 2020/21 £200k
  - 2021/22 £1m
- **Ethics 2020/21** – dependant on the approach, additional development and rollout to faculties
  - 2020/21 – £250K
- **New Projects and Initiatives** – £500k

Increase in services will require additional FTE to support additional services.

**The Research Applications Roadmap in more detail**

Outside of the current project scope detailed above, the emphasis of the next 18 months is to;

*Maintain a steady state clear of REF 2021 whilst still progressing the business needs, leveraging the current systems to improve capabilities without causing too much impact on the research processes prior to REF. This will include;*

- Extending current capabilities of existing services e.g. Review of existing modules within Symplectic Elements e.g. grants and impact. Scope business requirements through a dedicated task-group and understand how these modules will align with requirements. Rollout and deliver would include identification of business ownership and a model for delivery and support.
- Develop and enhance IRIS based on user feedback and documented enhancements, with an emphasis on improving our central home for showcasing UCL research and improving user satisfaction. IRIS has for some time has been supported by ‘keeping the lights on’ where dedicated time and development is required. This work will also support the current REF exercise.
- Provide support and steer for the Worktribe Steering group, formed to agree/prioritise and secure funding for UCL-wide requirements for future development/deployments of Worktribe modules where appropriate. As well as Research Services, key future drivers include KEF, UCLB/UCLC, MTAs, and feeds into the data warehouse.
- Fully understand the reporting requirements across the suite of research applications and provide working models for each business area. Work closely with internal ISD teams and existing suppliers to ensure business needs are met e.g. support Research Services to implement a reporting solution using Worktribe data to create and manage the operational and management information needs.
- Fully understand the IT requirements for the Doctoral School and Organisational Development (OD). Review the ambitions of both business areas and understand how existing services currently meet these needs and how best to service these needs in the future. This work will look at existing services and future service in the RITS pipeline e.g. Electronic Lab Notebooks (align with Data Services roadmap)
- Understand the usage and requirement for the existing legacy systems in the portfolio e.g. Kit Catalogue and the RICs database. Where services are no longer fit for purpose plan, ensure current requirements are being addressed e.g. through other projects/services and plans for possible decommission.
- Replatforming of current services e.g. UCL Discovery (eprints) and ensuring existing operating environments are stable and consistently managed e.g. RPS
- Implementation of a ‘devops’ ways of working across the ISD operational teams. Benefits of which include faster delivery of features, more stable operating environments, improved communication and collaboration and more time to innovate (rather than fix/maintain)

Research Applications Beyond REF (2021/22)

The outcome of the target architecture model for the research applications domain will see a major change to the way research applications are delivered at UCL. The project will have agreed future system capabilities and sort feedback and buy-in from the UCL research community. At this point will be in a position to manage the procurement of a suite of applications through a formal procurement process and the ultimate aim is to deliver the right services, supporting research activity, to UCL’s research community.

The majority of the roadmap looks at Pre-REF activities. This will be updated in the future to align with the architect review activities (yet to be defined).
Training

Support and training for eResearch skills

One of our key priorities is to support the development of early career researchers – ‘eResearch’ skills are no longer desirable assets for our early career researchers; they are essential to their learning and growth. Equipping them with core skills in data management and programming fundamentals sets them up for success in an increasingly complex digital environment and are as vital to their future career as writing or a familiarity with ethical considerations.

We will ensure our researchers thrive in the ‘eResearch’ environment by providing the access they need to training and specialist support. We refer again to the illuminating feedback from the eResearch Domain launch event which found that “increased access to expertise and support” was the factor that would most greatly accelerate their research, over “more funding” and “faster more available compute”. It is vital therefore that UCL’s significant ongoing investments in eInfrastucture are accompanied by skills-building initiatives, to promote both researcher success and return on this investment. The increasing complexity of the data environment adds to the urgency of this imperative, particularly in the social and medical sciences. The benefits of this activity go further; researchers with a high level of proficiency in research programming, or in using HPC systems, are more likely to seek and gain access to national and international facilities beyond UCL (see the last bullet point in Plans for 2019-20, below). In many fields of further study or employment, these skills are now considered to be core professional competencies. As the size and diversity of our eResearch community grows and demand for skills-based training increases, new approaches are needed to scale provision to meet this at every stage of a researcher’s career. Through collaboration with the academic community, UCL’s Centres for Doctoral Training (CDTs) and our delivery partners, we will foster the integration of teaching, training and support that enables and enhances the effectiveness of our researchers at every stage of the lifecycle. From data management planning to publication, and technology selection to data analysis, research students and staff at all career stages will benefit from delivery models that reflect changing and varied learning patterns and needs.

Current State

Research IT Services has run well-attended classroom-based training courses each term for the 8 academic years that the department has been in existence, having developed new courses in Unix Shell, Python programming for research, and more advanced Research Software Engineering practices, in addition to long-standing hands-on training in High Performance cluster computing. UCL-wide Software Carpentry workshops have been run on at least a termly basis since 2013, along with a pilot Data Carpentry session with a view to running this for domain-specific groups in the future. Version control training (which is typically delivered as part of Software Carpentry) has also been delivered as a standalone session for a few departments and groups, with a view to running these more frequently; this will also help with training a larger pool of Software Carpentry instructors. These courses have been supplemented by advanced training offered by NAG and ARCHER (EPCC) as part of the national HPC service. The UCL Library offers hands-on support and training for IRIS and RPS on a formal and informal basis. RITS has also run several IRIS/RPS awareness sessions for particular departments/groups, including the IOE. The Research Data storage and repository services are relatively recently established and hands-on training does not yet exist (there is some online help via the RITS website). It is envisioned that the Research Data Management support officers (based in the Library) will help to identify and develop the training required and...
coordinate and promote its delivery, as well as continuing to help staff drop-in advice sessions and enhance the existing online materials on both the ISD and Library websites. For the past 3 years, we have collaborated with Digital Education to develop self-directed, online versions of our most popular courses. This year, funds have been allocated to the eResearch domain for Research IT training, with the eResearch Domain’s Training Working Group acting as academic governance for the project.

What has changed

- UCL has experienced an unprecedented degree of growth, with more than 12,000 academics and researchers, compared to fewer than 6,500 in 2007.
- Technological advances have expanded the number of disciplines which use data and associated e-Research methodologies. The different academic disciplines have varying levels of maturity in using HPC, data science, research data management and software development. They consequently have differing training needs.
- The introduction of the ISD digital masterplan has a number of implications for training:
  - Objective 2 Research IT: Promoting world-class research through IT education. This focuses on students, but staff could also benefit.
  - Objective 2 Research IT: Supporting data-driven science and open science.
- The demand for training across UCL (both for existing and new courses, for classroom based and online) is increasing, but RITS does not currently have the staff resource to expand the classroom training to more than 1-2 instances of each course per quarter. Self-help resources such as Lynda.com/LinkedIn, although of good quality, could be better signposted, and some are incomplete. Some of the unfulfilled demand for classroom training could possibly be shifted to self-help/online routes by better signposting (e.g. from Moodle) and developing our existing materials where appropriate, as well as providing curated lists of externally provided training (e.g. EPCC, Software Carpentry, Lynda/LinkedIn). The RITS online training project has already developed three self-paced online courses (Introduction to Unix Shell and Introduction to Python, and Research Software Engineering with Python). Of course, we recognise that there will still be demand for classroom training (in particular, from research students who tend to prefer it, and people who are completely new to research computing) and also understand that users who choose self-paced learning will still need support.
- Increasing demand for both the pilot Aristotle interactive Linux service for teaching, and an RStudio service for the Department of Political Science. The Aristotle teaching cluster is run unfunded on hardware retired from previous services; the RStudio server is currently funded from the IT for SLASH budget.

Plans for 2019/20

- Continue to work with RITS service users, the Doctoral School, CDTs, OD Research, the eResearch Domain Training Working Group and other interested parties to identify unfulfilled training needs. Continue to develop links with teaching units regarding CPD skills in data- and computationally-intensive research, to promote join-up.
- Continue to work with the ISD Digital Skills Development team to address training demand and unfulfilled requirements via regular catch-up meetings; carry out an annual training review, performing a gap analysis of user needs and available resources. Continually review and improve existing classroom and online training resources, ensuring that these are in sync and complement each other, allowing users to select the most suitable source of training.
• With the aim of coordinating Research IT training across UCL, work with the Research IT Community of Practice Common User Training sub group (Research IT COP CUT), identify and curate a set of additional training resources in Moodle, e.g. articles, self-paced online learning (LinkedIn/Lynda.com), and formal training courses (both classroom and online, internally and externally provided), which can be referred to in staff and student inductions and linked to from the RITS website and other sites across UCL where staff and students might expend to find this information (for example, the Doctoral School, OD Research and ISD’s training pages).

• More effectively raise the awareness of the training on offer, by targeted advertising of particular courses to staff or student groups that are likely to benefit, especially for training that has been designed to be domain-specific such as Data Carpentry.

• Define a set of “user stories” (e.g. see Software Carpentry’s learner profiles: https://software-carpentry.org/audience/) and develop a user training path flowchart and tools for self-assessment of existing skills, linked from both the RITS website and Moodle.

• Continue to liaise with Library Services to plan and deliver joint Research Integrity training, contribute to central and local induction sessions and additional training for departments across UCL as appropriate.

• In 2019/20, develop a Research IT training course offering formal credit for UCL students.

• Partner with the Library to develop Open Science online training.

• Develop the RITS Introduction to HPC/HTC classroom course in line with the principles of Software and Data Carpentry (see HPC Carpentry https://software-carpentry.org/blog/2017/06/hpcarpentry.html), and develop complementary online introductory HPC training.

• Establish a booking system for Aristotle, to allow it to become a full service used by lecturers in undergraduate teaching. This, along with a hardware refresh will alleviate some (although not all) of the issues with the current pilot service.

• Develop the support available from RITS technologists at our existing fortnightly drop-in sessions to include self-paced learning cohorts, and facilitate peer support. This could include hands-on HPC taster sessions, an introduction to GitHub, RDM hints and tips, for example.

• Support the UCL research community to access local, regional, domain-specific and national resources (both facilities and training) through engagement and communication via local and national channels (e.g. ARCHER Champions network).

Ambitious Goals
Increasing research impact comes through training researchers to build better quality software. UCL training provision however is often piecemeal and delivered by many different groups. The Research Software Development Group has already done significant work in this area, e.g. through short Software Carpentry workshops and the longer Research Software Engineering with Python course, and have ongoing collaborations with some academic departments and UCL CDTs to provide training, as well as getting involved in new initiatives such as Clinician Coders and cross-London shared Software Carpentry workshops. During the next few years, we will work with these and other interested parties (e.g. the e-Research domain) to develop a comprehensive shared curriculum in computational and data-intensive research methods, where our students benefit from hands-on experience. This activity will also involve RITS staff beyond RSDG (and staff beyond RITS). The outcomes from this process should include (i) the creation of courses that count as formal credits for
UCL students, and (ii) selling courses externally to raise income and enhance our reputation. Exploiting the full range of RITS expertise will enable us to position these courses uniquely within the market (for instance AI for healthcare using HPC within the Data Safe Haven). We aim to develop at least one new course each year. **This is the area where enhanced capital funding can really make a difference to prioritise these activities, and academic buy-in from across the university can ensure that solutions are joined up, exploiting relevant expertise from across UCL’s many departments, rather than working in isolation.** To ensure progress is strategic and coordinated, we recommend at least the creation of a new permanent grade 9 ‘Head of Research IT Training’ position to lead activity in this area.

1. Dramatically grow training and skills development capacity and capability to meet diversity of needs relating to UCL researchers’ ability to fully exploit available services and infrastructure. Become a global leader for open science/open data practices and reproducibility of research outputs, training all of our researchers in best practice in managing the research data and software underpinning their publications. This will require additional staff resource (see costs below).

2. Establish a local network of software and data carpentry instructors (of which there are now significant numbers across UCL) to seed the foundations of a research IT training community/hub (along the lines of the Research Programming Hub [https://www.ucl.ac.uk/research-it-services/software-development/ucl-research-programming-hub](https://www.ucl.ac.uk/research-it-services/software-development/ucl-research-programming-hub), or linked to it).

3. In 2021, establish a dedicated teaching cluster facility to replace the Aristotle service, using the same batch job scheduler as the research clusters. Work with the ISD training team to make the case for a UCL-wide RStudio server for hands-on training. In 2020/21, develop and pilot an AI for Healthcare course.

4. In 2021/22, launch the AI for Healthcare course to an external audience.

5. Best practice research data management training, from project planning to long term preservation and re-use, provided in conjunction with the library. This will assist researchers with building good practice into their research, helping inform workflows, technology choices, documentation practices, and ensuring appropriate long-term data preservation and access.

**Costs**

- Classroom, online and hybrid training courses cost on average ~£0.1M to develop; capital funding can therefore accelerate delivery in areas such as creating new training courses. In order to expand the range of training (including developing advanced HPC courses), the Research Computing team needs a significant increase in headcount. As proposed in the Research Computing roadmap, an additional two analysts (1 grade 8 and 1 grade 7) would relieve some of the pressure on the support teams and allow more time to develop and deliver training, but ideally we would have funding for an additional grade 7 post to work with the Research Computing and Research Software Development teams on advanced training. We recommend at least the creation of a new permanent grade 9 ‘Head of Research IT Training’ position to lead activity in this area. The intent this year is to develop HPC Carpentry with staff from the RC and RSD teams, along with an introductory HPC/HTC online training course. With an additional member of staff with the right expertise, advanced classroom training could be developed using the same model, e.g. an AI for healthcare.
course. Funding required for hardware for a dedicated teaching cluster (to replace Aristotle – as part of Myriad phase II)

- £50K funding to be secured for the Open Science online course
- RDM best practice training to be developed with the Library under BAU, until the end of 2021. The launch of a new anonymisation service (see Research Data Services roadmap) may require additional training to be developed for which funding would also be needed.