1. Evolution in an age of technological revolution

After a number of years in which the stock response to the accelerating data deluge was simply to throw larger and larger volumes of storage capacity to the university research community, the provision of planned and structured research data management (RDM) services has at last begun to gain a foothold across the higher education sector. So what has been going on in those most recent decades of the information age, since around 1990, and why has it taken so long for the sector to accept that it needs to introduce a new kind of support infrastructure, both technological and human?

A full nine years after Tim Berners-Lee launched the world wide web, an act that introduced the first truly public data search and retrieval service to the internet, the UK government’s spending review of 2000 delivered its original and unprecedented e-science initiative. This was designed to encourage the development of an IT infrastructure sufficient to support the increasingly global research collaborations emerging from science and engineering disciplines. Such collaborations, in fact e-science itself, were characterized by the shared use of some combination of very large computing resources, enormous data collections and fast and ubiquitous access to remote facilities or sensor data. Three years later, in its Circular 6/03 (Revised) Digital Curation Centre,¹ the UK’s Joint Information Systems Committee (Jisc) called for proposals ‘to pilot the development of generic
support services for maintaining digital data and research results over their entire life-cycle for current and future users’. The analysis that Jisc used to promote this initiative was that the ‘current generation of “e-Science” experiments and computations will create more scientific data in the next five years than has been collected in the whole of human history. Properly curated, this data will form a major resource for future generations of scientists.’

From these beginnings, one may assume that the active curation of primary data had been recognized as a core requirement of the e-science community. The corollary was that some new kind of service provision would be devised to manage the challenges inherent in not only the long-term preservation of electronic documents, databases and publications, but also to ensure that data users now and in the future could discover the data they need and make effective use of their preferred data manipulation tools and methodologies. As Circular 6/03 most perceptively determined, data curation ‘includes all the processes needed for good data creation and management, and the capacity to add value to generate new sources of information and knowledge’; the consequence of which, inescapably, must involve a sustained ‘interaction between creators and providers of data, the archivers of data, and most importantly the consumers of data’.

So far, so good. A further nine years on from that formative observation by Jisc, the Royal Society report Science as an Open Enterprise: open data for open science\(^2\) referred to the ‘rapid and pervasive technological change that has created new ways of acquiring, storing, manipulating and transmitting vast data volumes’, a revolution that has stimulated ‘new habits of communication and collaboration amongst scientists’ and challenged ‘many existing norms of scientific behaviour’. Yet in the view of the report’s distinguished authors, six further changes in research practice are still necessary before there can be a truly successful exploitation of these new technologies and collaborations:

1. A shift away from a research culture where data is viewed as a private preserve.
2. An expansion of the criteria used to evaluate research in order that credit can be given for useful data communication and novel ways of collaborating.
3. The development of common standards for communicating data.
4. The mandating of intelligent openness for that data which is relevant to published scientific papers.
The creation of a strong cohort of data scientists to manage and support the use of digital data.

The development and use of new software tools to automate and simplify the creation and exploitation of datasets.

According to the Royal Society’s report, the ‘means to make these changes are available’ but they have not yet been achieved due to the lack of ‘an effective commitment to their use from scientists, their institutions and those who fund and support science’.

Whilst acknowledging that the Royal Society’s particular focus in this report is to promote the principle of openness and to generate new momentum in open science, this assertion could be taken by those who have made real advances with standards, tools and the foundation of openness as the cause for some alarm. If one infers that between the late 1990s and the present day there has been little progress by the higher education community in its response to the transformational changes being experienced in research data generation, sharing and curation, then a number of significant initiatives have most certainly been overlooked. The Society’s key message, however, is expressed by the phrase ‘effective commitment’, which we shall interpret as the coherent and pervasive adoption of those measures necessary to the implementation of RDM processes, protocols, standards and services; further, this potentially universal adoption has to demonstrate that institutions have accepted these measures as fully funded, established and sustainable components of their institutional structures and business. In that case, yes, the Royal Society is spot-on, as the distance between the few who occupy the vanguard and the majority that has yet to address what is meant by a serious commitment to data management and sharing appears to have widened in the past few years. This is not an unexpected state of affairs; nor is it in any way a static one. In the large, diverse and complex organism that is a research-led university, the pursuit of such a course of action will be associated with considerable cultural change. This will be the transformational process that the Royal Society recognizes has still to be secured, but which even in the most agile and risk-disposed of institutions cannot be achieved in haste.

The development of institutional RDM services has already been likened to ‘an iterative cycle similar to business process redesign’ (Jones, Pryor and White, 2012) where, borrowing terms from the business sector, one would expect the introduction of any new mission-critical philosophy, principles and structures to be grounded in a fundamental rethinking of existing
processes, in order to deliver greater productivity and value to the customer. Typically, in business, this will involve the radical and rapid simplification of organizational layers and the elimination of unproductive or redundant activities. Not so in higher education. In all but the most ‘managed’ of universities, the prevalent culture will generate significant resistance to the unchallenged achievement of these goals. Here, the ‘iterative cycle’ that defines the process of redesign will prove to be exactly that, with an inevitable need for extensive repetition, in numerous forums, of the argument for RDM; the consequent intromission of recurrent adjustments to the parameters for change; and the persistent necessity to advocate change when you may have felt that your manifesto had already received the community’s assent. We do not mention this merely to paint a negative view of higher education or lay blame, but to point out how important it is to be aware of these challenges from the outset and not to be discouraged by the prospect, since the underlying movement for change that will eventually strengthen your cause is slowly but surely under way.

2. Six necessary changes?

Take, for example, the first of the Royal Society’s six necessary changes: a shift away from a research culture where data is viewed as a private preserve. This is already well established as a necessary practice in some of the ‘big sciences’, like physics, where the analysis of huge volumes of research data depends on its being shared internationally; and in the field of genomics, the rapid access to sequence and other kinds of genomic data has been a guiding principle of the Human Genome Project. In the UK, in 2009, the Panton Principles\textsuperscript{3} were written by a group of scientists convinced that ‘for science to effectively function, and for society to reap the full benefits from scientific endeavours, it is crucial that science data be made open’. The Principles were subsequently refined and taken to a broader public platform with the help of the members of the Open Knowledge Foundation Working Group on Open Data in Science,\textsuperscript{4} a diverse group of scientists, publishers, students and others with an interest in promoting open science. So, whilst this is all still indicative of a culture of self-regulating scholarship and not the large-scale public enterprise being promoted in the Royal Society’s report, it is nonetheless a real representation of the shift in attitude and practice that will in time demand platforms and structures for support on an institutional, national and global scale.

It is as feasible to provide evidence of progress in each of the other five
areas of necessary change. In the area of research evaluation, for example, Main Panel C of the UK’s Research Excellence Framework (REF), the new national mechanism for assessing research undertaken in higher education institutions, plans to accept a wide range of research output, including digital artefacts such as datasets and multi-use datasets. This is a small beginning in the move away from the traditional submission of research publications, and it applies to only one of the four main assessment panels, but it is a very significant change. In a consultative document issued on 25 February 2013, the Higher Education Funding Council for England (HEFCE) prepared to take further steps after the 2014 REF by seeking views on the submission of scholarly publications delivered as open access. Speaking on behalf of the four UK funding bodies for higher education, and with an objective to increase considerably the proportion of research outputs published in open-access form, HEFCE declared their intention to require that outputs meeting the REF open access requirement shall be accessible through an institutional repository. They are of course talking here about publications, rather than data, but they make clear that in practice each submitting institution will ‘maintain a web facility through which all relevant outputs might be identified and accessed (including items available through a link to another website)’. Even more significantly, having made explicit the requirement for all submitting institutions to provide enabling infrastructure, there follows particular reference to the role of data in making the research process more effective, with respondents to the consultation being invited to comment on whether sufficient progress has been made to implement a requirement for supporting open data as well as open publications.

The growth in specialized national services is also an indication of gathering momentum. We have already remarked the creation of the Digital Curation Centre (DCC) and, if you are pondering the need for standards or other authoritative guides to data management frameworks, a simple search on the Centre’s web pages will quickly reveal the extent to which benchmarks for metadata and data communication protocols have advanced internationally. If help in deciding what software platform best meets your needs is what you are seeking, OSS Watch, funded in the UK as a free, national service, provides unbiased advice and guidance in the use, development and licensing of open-source software. If cloud services are going to feature in your infrastructure portfolio, the JANET Brokerage Service provides independent guidance, collaborative purchasing power and due diligence to help institutions move towards cloud storage and external data centre services.
The success of the Jisc’s Managing Research Data Programme is further evidence of the determination of national bodies to persuade and assist institutions in meeting the data challenge. In particular, its 2011–13 programme has enabled 17 large institutional projects to focus on developing and extending RDM infrastructures. An additional eight projects are helping research groups, projects or departments fulfil disciplinary best practice and the requirements of research funders by implementing data management plans and supporting systems.

These are indeed all significant footholds and, as more initiatives are sustained, there is already evidence of increasing cross-fertilization of ideas, experience and the exchange of good practice right across the sector. Yet projects are notorious for their inability to gain the critical mass and momentum that will ensure their sustainability before project funding is exhausted. Although the Jisc Managing Research Data Programme expects that the data management projects it has funded will each be part of a recognized institutional mission to provide high-quality support for research, in an environment where there is heavy competition for limited resources that shift is not everywhere guaranteed. But the kind of change envisaged in Science as an Open Enterprise, which has been further invested with the weight of government resolve in the Open Data White Paper: unleashing the potential (Cabinet Office, 2012), including Tim Berners-Lee’s proposed Five Star Scheme for data standards, presupposes that universities will themselves have committed to sophisticated mechanisms of support and unrestrained measures for delivery of their research output. After several years of best efforts on the part of national agencies, in order to move more nimbly to that happy state now requires something of revolution rather than the long, slow particulate motion of evolution. History is witness to the fact that such a transition is normally enabled by the intervention of some fresh external stimulus.

3. Policy frameworks, mandates and expectations

Major stimuli were in no short supply during the year prior to publication of the Royal Society’s report. Research Councils UK (RCUK) set the tone for 2011 with its Common Principles on Data Policy, which although lacking teeth was unique in giving common cause to the seven councils’ commitment to transparency and for the first time described a coherent approach across the research base. Publicly funded research data was declared a public good, ‘produced in the public interest, which should be...
made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property’. Actions for institutions were implied by the assumption that they would have data policies and plans in place and that measures would be taken not only to preserve data of acknowledged value but that it would also be assigned sufficient metadata for other researchers to understand and reuse it. It was even recognized that the use of public funds to support the management and sharing of publicly funded research data would be legitimate, although no new funds were being offered for that purpose – but, after all, these were principles, not plans.

More instructional policy frameworks were soon to follow, however. Having released a revised Research Data Policy in September 2010, the Economic and Social Research Council (ESRC) brought a new requirement into force in spring 2011, whereby research grant applicants would be required to submit a statement on data sharing and provide a data management and sharing plan. What softened this new obligation for institutions was the knowledge that the ESRC already offered extensive support to its funded researchers in the shape of the UK Data Archive (UKDA) at the University of Essex. As explained in the revised policy document, data centre staff will happily assist researchers to plan their data management and sharing, and providing ongoing support throughout their project, including final deposit and reuse. As we shall explore later in this book, the availability of national data services should always feature in deliberations concerning the deployment of institutional resources for RDM service provision.

But if one is looking for the moment when university managers woke up to a more revolutionary flavour in research council data policy that, too, occurred in 2011, when on 1 May the Engineering and Physical Sciences Research Council (EPSRC) published its Policy Framework on Research Data. The Framework described seven core principles, which were directly aligned with the core RCUK principles, two of which were deemed to be of particular importance: first, ‘that publicly funded research data should generally be made as widely and freely available as possible in a timely and responsible manner’; and second, ‘that the research process should not be damaged by the inappropriate release of such data’. Both imply actions to be taken by institutions and their researchers, both assume that mechanisms do or will exist to enable those actions. Neither sounds extraordinarily radical except that, unlike the ESRC, the EPSRC provides no support infrastructure, meaning that universities in receipt of EPSRC research grants
would themselves have to supply the appropriate services and support.

By itself, that realization was not likely to persuade institutions that they must commit to the development of RDM services; after all, higher education is awash with principles of one kind or another. The moment of epiphany came, though, with the accompanying nine ‘clear expectations’ that the funder lay at the door of the 100-plus organizations in receipt of its funding. Most rousing of them all was expectation number nine, which plainly stated that those organizations ‘will ensure adequate resources are provided to support the curation of publicly funded research data; these resources will be allocated from within their existing public funding streams, whether received from Research Councils as direct or indirect support for specific projects or from higher education Funding Councils as block grants’.

To ensure that, having gained everyone’s attention, the attention did not immediately slip away again, the EPSRC had written the previous month to forewarn all university vice-chancellors and principals of the EPSRC policy timetable. As reiterated in a second letter, sent in February 2012, the timetable was dependent upon two deadlines: by 1 May 2012 each institution was to have a clear roadmap in place to align policies and processes with EPSRC expectations, which by 1 May 2015 should have led to full compliance with those expectations. If by that later date any institution is found not to be fully compliant and it can be shown either to be deliberately obstructing the proper sharing of research data or otherwise seriously failing to comply with EPSRC’s expectations, this will initiate a process that could ultimately lead to it being declared ineligible for EPSRC support. Since the EPSRC supports about 8000 academic researchers from a £4 billion portfolio, this could not be ignored.

What, then, do the other eight EPSRC expectations mean for universities required to ensure there are ‘adequate resources . . . to support the curation of publicly-funded research data’? The self-analysis involved in the creation of a roadmap will enable an institution to discover its current condition and the steps it needs to take to achieve compliance. In other words, it will need to ask what actions are necessary to support research data curation, who will take them, when should they be taken and how much is this activity expected to cost. Effectively, as demonstrated by Table 1.1, the creation of an EPSRC roadmap could be the kernel of a larger RDM service plan.
### Table 1.1 EPSRC expectations and university responses

<table>
<thead>
<tr>
<th>EPSRC expectation</th>
<th>Potential university response</th>
</tr>
</thead>
</table>
| 1. Research organizations will promote internal awareness of these principles and expectations and ensure that their researchers and research students have a general awareness of the regulatory environment and of the available exemptions which may be used, should the need arise, to justify the withholding of research data. | • Provide publicity and guidance, possibly through workshops and training including induction for new researchers and postgraduates.  
• Provide a mechanism via the institutional repository (where one exists) to add a statement / link to data.  
• Reflect funder and regulatory obligations in policies for research and research data.                                                                                                      |
| 2. Published research papers should include a short statement describing how and on what terms any supporting research data may be accessed.                                                                                                                                                                                                                   | • Publicize this requirement, include it in induction for new researchers and reflect it in policies for research and research data.                                                                                                                   |
| 3. Each research organization will have specific policies and associated processes to maintain effective internal awareness of their publicly-funded research data holdings and of requests by third parties to access such data; all of their researchers or research students funded by EPSRC will be required to comply with research organization policies in this area or, in exceptional circumstances, to provide justification of why this is not possible. | • Implement measures to record and make discoverable all publicly-funded research data that cannot be treated as an exception. This could have far-reaching consequences – e.g. the creation and maintenance of a data catalogue that also records the details of third-party access requests. |
| 4. Publicly-funded research data that is not generated in digital format will be stored in a manner to facilitate it being shared in the event of a valid request for access to the data being received (this expectation could be satisfied by implementing a policy to convert and store such data in digital format in a timely manner). | • Undertake an audit to identify such physical material.  
• Consider the cost, benefits and risks involved in conversion to digital format, compare with alternative ‘physical’ measures and select preferred option.                                                                                         |
| 5. Research organizations will ensure that appropriately structured metadata describing the research data they hold is published (normally within 12 months of the data being generated) and made freely accessible on the internet; in each case the metadata must be sufficient to allow others to understand what research data exists, why, when and how it was generated, and how to access it. Where the research data referred to in the metadata is a digital object it is expected that the metadata will include use of a robust digital object identifier (for example as available through the DataCite organization – http://datacite.org). | • Identify appropriate and acceptable metadata standards and protocols.  
• Implement processes and responsibilities for adequate metadata assignment.  
• Establish and sustain a mechanism to enable data discovery on the internet.  
• Advise researchers of requirements and the support to be provided.  
• Provide guidance and training to data custodians and/or researchers in the assignment of metadata.  
• Include requirement in research data policy.                                                                                                                                          |
| 6. Where access to the data is restricted the published metadata should also give the reason and summarize the conditions which must be satisfied for access to be granted. For example ‘commercially confidential’ data, in which a business organization has a legitimate interest, might be made available to others subject to a suitable legally enforceable non-disclosure agreement. | • Provide guidance, particularly in induction for new researchers.  
• Include requirement in research data policy.                                                                                                                                                                                                   |
Such an analysis makes plain that meeting these not-extraordinary expectations for data management will involve not only the research process but a diverse range of associated activities, roles and organizational constituents. At least three key perspectives would have to be represented and agreements forged between them for a fit for purpose service to emerge: (1) the internal community engaged in research practice, (2) institutional management and (3) the providers of information systems, services and support. Each of these perspectives will in turn be multifaceted, potentially adding to the complexity of the undertaking. Most heterogeneous will be the research support perspective, which could embrace staff from the library, information technology services, research administration and the records management functions, not forgetting the various kinds of specialist support staff located within individual research teams. As a rule, these groups will not have worked together previously as coherent teams and they may know little about each other’s activities; it is equally unlikely that they are already pursuing shared routines and objectives. Thrown together by the force

<table>
<thead>
<tr>
<th>EPSRC expectation</th>
<th>Potential university response</th>
</tr>
</thead>
</table>
| 7. Research organizations will ensure that EPSRC-funded research data is securely preserved for a minimum of 10-years from the date that any researcher ‘privileged access’ period expires or, if others have accessed the data, from last date on which access to the data was requested by a third party; all reasonable steps will be taken to ensure that publicly-funded data is not held in any jurisdiction where the available legal safeguards provide lower levels of protection than are available in the UK. | • Long-term measures will have to be decided in order to archive, curate and enable discovery and access.  
• Need to assess the extent to which available options will be deployed (i.e. an institutional repository, national data centres, cloud services) and investment made.  
• Identify roles and responsibilities for the preservation of data, its periodic review and methods for secure and appropriate disposal. |
| 8. Research organizations will ensure that effective data curation is provided throughout the full data lifecycle, with ‘data curation’ and ‘data lifecycle’ being as defined by the Digital Curation Centre. The full range of responsibilities associated with data curation over the data lifecycle will be clearly allocated within the research organization, and where research data is subject to restricted access the research organization will implement and manage appropriate security controls; research organizations will particularly ensure that the quality assurance of their data curation processes is a specifically assigned responsibility. | • Effectively, expectation number 8 is requiring institutions to create a comprehensive research data management service!  
• As implied by the second half of this expectation, the service is likely to involve a broad range of actors and activities with new roles and responsibilities to enable the gamut of data management planning, the assignment of metadata for discoverability and access, long-term preservation, et al.  
• Extensive planning and investment will be required, with the potential for new recruitment where essential skills are lacking, together with the introduction of new routines for governance and communication. |
majeure of funder mandates they will have to face a wholly new set of interpersonal subtexts and organizational dynamics that will need to be addressed, in addition to the main challenge of familiarization with the tenets of research data management.

4. Supporting modernization and institutional focus

Meeting that challenge is not a prospect that has to be faced in isolation. Also in spring 2011, with funding from the Higher Education Funding Council for England (HEFCE), the DCC rolled out the first of its institutional engagement programmes. This initiative was but one feature of yet another funding body’s attempts to tackle current issues in higher education. Although HEFCE’s initial focus here was not on research data, its Universities Modernisation Fund, set up to support 10,000 extra student places in 2010–11 and encourage universities and colleges to adopt efficiencies and savings through shared services and other innovative practices, provided the DCC with sufficient resource to embark on a series of unique collaborations with 21 UK universities. These collaborations, called institutional engagements, were designed with the sole purpose of improving institutional capabilities (their ability to articulate and achieve RDM objectives) and capacity (the creation of sustained and effective infrastructures) in the conduct of effective research data management. These free, 60-day engagements have provided the selected 21 universities with assistance in making the case for RDM services and infrastructure, in re-engineering the roles of support staff and equipping them with new skills, and in the transfer of knowledge about available tools and techniques. Moreover, whilst this chosen clutch of institutions represents around only a fifth of the research-led universities in the UK, a condition of each engagement was that the experience and outcomes would be openly promulgated and actively shared with the wider community.

Whilst the first phase of that programme has drawn to a close, the DCC has maintained its links with the candidate institutions. It has also commenced a second tranche of engagements that reflects the needs of a more mature data management environment, whilst at the same time recognizing the growth of a rearguard ‘long tail’. This refined engagement process now offers a series of thematic modules covering specific technical solutions to policy and infrastructure questions, which are likely to appeal to institutions that have already begun to shape their RDM service. But in addition there is a package of complementary products designed to assist
latecomers in conducting advocacy, building strategies and re-engineering the skills base. With this broadening base the Centre remains a key source of advice and guidance, acting increasingly as a medium for exchange between multiple layers in the emerging community of data practitioners.

5. Global perspectives

Membership of that community is increasingly international, although the pace outside the UK has been set mainly by Australia and the USA. Both have witnessed pressure brought to bear on researchers and institutions from funders, from government and from within the research community: pressure to share and make data open as well as to select, appraise and preserve data for specific periods as a condition of the compact between funders and grant holders. The tension between these aspirations and mandates and the ability to resource appropriate responses has been addressed rather differently in the case of these two nations.

The national research data strategies of Australia and the USA have already been contrasted (Treloar, Choudhury and Michener, 2012) as responses to two very different government and research sector environments. The approach taken in Australia has been to develop a national data service, ANDS, which exploits a collaborative framework of data stores, federations and services across these sectors. Through the provision of guidance and instruction on internal institutional data management, ANDS's institutional partners are expected to move forward together in the implementation of data management planning frameworks, tools for data reuse and the development of teams with data management skills.

Taking this common approach, with considerable central funding and a three-year high-level project plan, raised the profile of RDM through ‘a combination of national services and coherent institutional research data infrastructure, combined with the ability to exploit that infrastructure with tools, policy and capability’ (Treloar, Choudhury and Michener, 2012). In Chapter 8 we provide a closer look at one of those institutional infrastructures with a case study from Monash University.

Although ANDS was created to drive improvements in the Australian research environment, the data registry solution that has been developed under its auspices, Research Data Australia,13 which provides a discovery service for Australian research data collections, has emerged as but one element of the ANDS service that is influencing service development elsewhere in the global data community. As UK universities become more
involved in the management of research data and their capacity to contribute develops, the case for a UK Research Data Discovery Service is also growing, adding a further potential building block to the transformation of research practice. Recognizing that the ANDS approach has demonstrated success and that the software is relatively mature, Jisc and DCC will in 2013 complete a pilot registry service that aggregates records of research data held by UK universities and by key national and discipline-orientated data centres.

Notwithstanding the National Science Foundation’s requirement from 2011 for grant submissions to include data management plans, the movement for change in the USA has come about less as a consequence of mandates or dictates than from initiatives arising within the research support community itself. There, far-sighted librarians in particular have noted the implications for their services from the rise of data-intensive research and have responded quickly by re-engineering their technical and human infrastructures. Whereas in the UK the gap between researchers and the library has been widened by the availability of the internet, which has nurtured a research culture of increasing self-reliance in terms of information retrieval and use, in the USA libraries have seized the initiative by retraining staff as data scientists and by taking the lead in such multi-disciplinary, multi-partner programmes as Data Conservancy, a $20,000,000 project involving ten partner institutions. Notably, Data Conservancy, with its objective of scientific data curation, perceives a library-led organizational framework delivering cross-disciplinary discovery. In February 2013, President Obama’s administration required research funding agencies to ensure that the public can access publications and data produced as an output from Federally funded research. This unusual high-level intervention will give further strength to the Data Conservancy mission to preserve data and make it available for scientific use. Any such pronouncements are there to be exploited by enterprising librarians keen to reinvent themselves as more active partners in the research agenda.

This and other data enterprises are by no means inward-looking and will provide a further source for the designers of RDM services of exemplars in infrastructure and service provision. The DataONE project, for example, with a team sourced predominantly from the digital library world, is charged with delivering a federated data network to preserve and improve access to data in the biological and environmental sciences. The DataONE plan talks in terms of implementing global infrastructure by allowing a number of nodes to be added to the network on continents outside North America.
In simple terms, DataONE will provide a place for researchers to store data and its associated metadata but, to equip them with the means to use this service effectively, the project is addressing a raft of interrelated activities. Data management planning, the techniques, protocols and methods for data acquisition and back-up, data analysis and workflow are but a small example of the areas under development and any institutions engaged in the design of RDM services would sensibly add this and similar international projects to their knowledge base.

6. Unfamiliar territory

Planning for the creation of RDM services not only brings together groups and functions that previously were separate and distinct, it very probably brings them into contact with a whole register of management and financial terms and procedures that previously were remote. Sometimes these can become confused and certainly they can seem confusing. We have therefore included in this introduction a brief explanation of some of the terms you are more likely to encounter. Their inclusion here does not necessarily imply that each of them will feature in your own start-up processes; for example, you may choose to develop a business case to secure the resources necessary and elect only to use a simple roadmap in place of full business and operating plans. Your approach will depend on the organizational structures and procedures already in place. What follows is meant to clarify terminology not dictate the manner of its use.

Strategy

Logically the starting point, an RDM strategy should be an aspirational statement setting out your long-term goals and objectives and the courses of action necessary to achieving them. It may include reference to the human and financial resources needed for meeting objectives but, essentially, it is more to do with setting the scene than setting detailed budgets. Endorsement of a strategy can in most circumstances be taken as approval to proceed within an institution’s overarching business mission. It also legitimizes the development of policy and serves as an invitation to submit a business case.
Business case

The business case will formally present the argument for resourcing the RDM project or programme. Typically, it will be a detailed but accessible document submitted for consideration by senior management, although it may also be delivered as a presentation. It will describe the background and context to implementing RDM infrastructure and services, the anticipated benefits, options considered (including the reasons for their adoption or rejection), predicted costs, a gap analysis and expected risks (including the costs and risks of doing nothing). By formally endorsing a business case, senior managers are consenting to the deployment of resources and the preparation of forward budgets. From this point it is also appropriate to commence work on a policy to underpin the strategy.

Business plan

As a summation of the overarching direction for RDM development and the activities to be resourced and undertaken, within a specified timeframe and well defined context, the business plan serves as the main reference document for the programme. Whilst business plans may be externally or internally focused, in terms of developing institutional RDM services it is probable that the principal focus will be internal, although external interests could exist on global scale in the form of consortial partners, research funders, commercial investors, publishers and discipline collaborators.

The business plan will reflect the terms of engagement approved by management, by drawing together all the components of the RDM programme into a coherently structured process and timetable. It will include the RDM vision, a summary of strategy, critical success factors and a description of perceived benefits, together with detailed plans for financial and human resources, the roadmap and operational plan, an explanation of measures to comply with legal, statutory and funder obligations, and a sustainability and exit plan. Although the RDM business plan is likely to cover a period of five to ten years, it should be reviewed annually to ensure that the identified critical success factors remain appropriate (and are being achieved). Resubmission within the institutional planning cycle is recommended to ensure that confirmed budgets are carried forward.

Roadmap

The term ‘roadmap’, recently associated with EPSRC’s data management
expectations, has been used to describe the series of actions necessary to achieving a particular level of RDM effectiveness. Ideally it will explain key expectations, current arrangements and gaps in provision; the milestones to be achieved over a specific timeframe to meet expectations and mend those gaps; the roles and responsibilities of individual actors and groups; and the costs. Whilst it has been used specifically to identify the measures that will deliver compliance with EPSRC policy, the roadmap is likely to feature as the baseline for a broader development of RDM service strategy and, as such, it could be treated as an element (and potential surrogate) of the business case and of the business plan.

Operating plan
An operating plan works as a subset of the business plan and will describe the goals and activities of the contributing parts of the RDM programme or organization (e.g. the working group or project team, the library or IT department) over a given operational period, typically a budgetary year. It is usually constructed (or revised) annually and is the basis for the annual revenue budget request. However, whilst the focus will be upon the next year’s budget submission, operating plans should include the predicted activities and budgets for between one and three years, in order to allow the institution to develop rolling forecasts and balances. The operating plan is also the mechanism for highlighting (and seeking ratification) of modifications to the business plan that may have arisen from changes in policy, movement in the fiscal environment, or adjustments to the strategy.

Since the implementation and growth of RDM will involve a diversity of stakeholder groups within an institution, the operating plan is the means by which they can each prepare for resource allocation at a team or departmental level (where the development of RDM is likely to depend on separate departmental budgets), at the same time acting as the medium for enabling a coherent and inclusive request for funding. The need for significant cross-departmental dialogue in the development of operating plans is therefore crucial. In practice, where RDM services are being grafted on to existing units and functions, each party may present its own contribution to the RDM strategy as a component of a discrete departmental plan.

Policy
An RDM policy describes the principles that an institution has agreed will
guide the decisions and the actions necessary to achieve desired outcomes. It is not to be confused with either a strategy or a plan, since it does not explain what will be done to achieve those outcomes. Nonetheless, an RDM policy is valuable as a public statement of intent and an expression of the commitment of management and senior stakeholders. Furthermore, it identifies such decision-makers as accountable for not only the policy but also the implementation of measures to deliver the resources and infrastructure required to enable the policy.

**Capital budget/expenditure**

Capital budgets are prepared to enable long-term investment in major items such as new or replacement equipment (e.g. servers and networks) or physical facilities (e.g. a data centre or repository). Other non-recurrent costs may be eligible for capital funding; it is advisable to seek advice on this matter from the finance department. Some funders allow research grants to contribute to capital budgets for the development of data centres and RDM services. It is recommended that clarification of funder policy is sought before funds are committed in this way.

When considering proposals for capital expenditure, an institution is likely to rank submissions to determine which will be the most financially rewarding. It is therefore important to have laid the groundwork in the endorsed business case for demonstrating the cost-benefits of the RDM development programme.

What constitutes a capital item can differ between institutions and it is always worth checking whether designation as a revenue item is both achievable and more beneficial. It is also important to remember that capital items will depreciate in value and will require maintenance. Both aspects of purchase should be described here but the actual budget for maintenance will appear in the revenue budget.

**Revenue budget/expenditure**

The revenue (or operating) budget covers the cost of those annual activities necessary to the development and delivery of RDM services, including estimates of the cost of human and other resources but excluding any capital items. As acknowledged above, the potentially diverse nature of an RDM team can call for significant cross-department collaboration in the creation of revenue budgets.
7. Summation

In this chapter we have introduced the provision of RDM services as an unfolding patchwork of challenges. For the higher education sector in the 21st century the development of digital technology has irrevocably altered research practice; it has also changed the context, having been instrumental in enabling new research methods that have accelerated the globalization of research. The impact of the digital age on traditional research support services, which were primarily understood in terms of the university library, is a leading motivation for change; some might say almost a case of do or die for the library profession. Certainly, for the time being, there is a gap between service provision and customer needs, although this is not only to be seen as a gulf between the library and the research agenda but is also evident in the marked impotence of some IT services when faced by enormous data scale and diversity. But as we have pointed out, the challenge is not only for these more obvious contenders, since it embraces a wider and uniquely heterogeneous ensemble of service and support.

This challenge demands a whole new way of thinking. In his summary of the report *Data-driven Infrastructure*, which a number of approaches are suggested for managing both corporate and research data, the author Max Hammond explains that, while universities have begun to address the growing demands for data management by creating a variety of architectures, ‘one concept that is emerging is that of data-centric architecture which focuses primarily on organizational data rather than systems, and is designed to facilitate the sharing of data between the processes handling institutional data’. It is not just a technology challenge that we are facing, although it is a socio-technical one, but principally it is a data challenge. That is what is new.

In a context of increasing regulation by government and greater levels of intervention by research funders, institutions can no longer procrastinate and must act to find a sustainable approach to RDM services that is fit for their own purposes. This book provides a step by step guide to the components of an RDM service, concluding with case studies that describe several innovative approaches to the development of data-driven infrastructures. There are a number of approaches to consider and many decisions to be taken. We wish you well as you embark upon them.
8. References

8.1 Websites
6. www.dcc.ac.uk.
7. www.oss-watch.ac.uk.
12. www.epsrc.ac.uk/about/standards/researchdata/Pages/default.aspx.

8.2 Citations

