

Plan M

Review of Stakeholder Input

December 2019

Introduction

In association with key strategic partner organisations and in consultation with libraries and suppliers, Jisc has focused a lot of energy over the past few years designing and implementing an important new element of national library data infrastructure. The [National Bibliographic knowledgebase](#) is a very large aggregation of data contributed by national, academic and specialist libraries which provides the foundation for a growing number of [Jisc Library hub services](#). Those currently available are: [Discover](#), [Compare](#) and [Cataloguing](#).

Discussions around this initiative have been wide-ranging but one strand has consistently focused on the issues and challenges in relation to bibliographic metadata and how we might collectively use the NBK as a catalyst for transforming the way that the data ecosystem/marketplace functions in the UK. Jisc convened a number of [Community Data Groups](#) during 2018 which helped to dissect some of the issues and propose ways forward. In Spring 2019, representatives of those groups convened a meeting to bring together a variety of stakeholder organisations including publishers, aggregators, data suppliers, standards bodies, library suppliers, industry advocates, procurement consortia, and libraries. It was at this meeting (May 22nd, 2019, Senate House, London) where the term 'Plan M' was first proposed.

(A definition of this term and a draft description of its purpose is set out in the related document – "[Plan M – Definition and Direction.docx](#)")

A follow-up meeting was organised (September 27th, 2019, Jisc London) to discuss the setting up of Task & Finish Groups to work on specific aspects of the challenge (e.g. data quality, data standards, open licensing). In parallel with this, strategic governance groups featuring senior managers and library directors with oversight of Jisc Library Services took the view that the ideas being surfaced by the 'Plan M' discussions required discussion at a more strategic level and two Focus Groups were convened to give senior library stakeholders the opportunity to scrutinise the direction of travel and steer future activity. (September 30th, 2019, Jisc London; and October 21st, 2019, Jisc Manchester).

What follows sets out a summary and a synthesis of the discussions relating to the challenges and issues that have necessitated the formulation of a plan. It draws particularly on some of the key points that emerged from the two senior stakeholder Focus Groups but does not attempt to record the rich discussion that has taken place throughout 2018-19. Its purpose is to pick out some of the most important levers for future action and to acknowledge the foundational concepts upon which Plan M needs to be taken forward.

Issues & Ideas

This section is divided into 5 areas: Concepts, Records, Workflows, Collaboration and Economics. These categories encompass most of the discussions that have taken place at the various meetings and provide a framework for setting out some of the headline points.

Concepts

- Trying to galvanise action and bring people together under the banner of ‘improving metadata’ as an end itself will not be effective and does not do justice to the scale of change that is necessary. The motivating focus needs to be on how it relates to the role that libraries want to play in 10 years’ time and beyond (see *10-year Vision* section below); and the positive impact that changing practices will have on end users and research.
- There is a clear appetite in the community, at all levels and across sectors (academic and commercial) to aim for ambitious change. There is broad agreement that Jisc is well-placed to facilitate and coordinate the required programme of work and that the NBK provides the central infrastructure around which workflows and agreements can be designed.
- Libraries want to be able to easily acquire data from various sources, amend it if they choose to, and freely re-share it with other libraries. Assertions around whether data should be ‘owned by’ or ‘licensed to’ the library sector are of less importance than what can practically be done with the data.
- There is broad acknowledgement that the routine creation and delivery of bibliographic data at scale has largely been outsourced to third-party services and that those services need to operate on the basis of a reasonable revenue model, either to support data creation and/or data delivery.

Records

- Discussions around ‘data quality’ sometimes imply that there is a gold standard of data that might suit all purposes if only common agreement could be reached and quality maintained. Libraries have different needs and acquire data with different ends in mind. It is more useful to talk about data that is ‘fit-for-purpose.’
- The point above notwithstanding, it will be useful to identify and agree the parts of bibliographic records which can be defined as being automatically in the public domain and freely reusable in any context, including for commercial purposes (i.e. equivalent to CC0).
- Whilst there are a variety of use cases around bibliographic data, there will be commonalities of interest amongst libraries and these should be mapped to provide some clustering of requirements. Not only is this important for understanding fitness-for-purpose, it is important for designing viable business models for data supply (see ‘*Economics*’ section below).
- It is important to record and retain the provenance of data where it has been created or managed by an entity that is seeking revenue for its activities. It would appear technically unrealistic to expect all changes to every record by every actor in the ecosystem to be recorded in a provenance trail.
- There is palpable frustration from the library sector about data formats and cataloguing standards. This relates to: inertia around moving on from MARC; the incompatibility between MARC and ONIX; the sluggish adoption of RDA; the uncertainty around BIBFRAME and linked data; the siloed nature of KBART. These are all issues that need resolving but Plan M is not the right vehicle to do that – at least in the short term. (see ‘*Collaboration*’ section below)

- EBook data is seen as a particular challenge and will need to be addressed by a focused strategy within the framework of the Plan M activities.

Workflows

- The principle that Plan M should work towards is 'create once and share with everyone'.
- From the library perspective it is frustrating that the creation of high-quality records can't be pushed as far back in the pipeline as possible, ideally all the way back to the publisher and the point of publication. There may be benefit in including an advocacy work strand in Plan M that reinforces messages to publishers about the potential for higher visibility for their products in discovery systems and the consequent effect that may have on sales.
- There is support for putting the NBK at the centre of the ecosystem and making it a central distribution hub for the dissemination of appropriately licensed metadata. It is a simple concept but has a lot of implications and poses many questions.
- The concept of designating the NBK as the UK data infrastructure for bibliographic record availability will require agreements to be reached with a number of other creators/suppliers. The principle for Plan M will be to funnel those sources into the NBK so that the data is conveniently in one place and is consistently licensed for the whole community to use, edit, share and re-use as it wishes.
- For this sharing model to be effective, Jisc will need to work with all parties to reduce friction and increase prospects for automation, both in the exchange of data and for the purpose of matching the descriptive record with the described item. ISBN has an obvious role to play here for books but there is certainly scope for encouraging more integration of persistent identifiers across the whole ecosystem and to join publication up with other entities such as author, organisation, funder, grant, etc.
- Whilst there is wide support for the NBK to act as a national data broker, there is concern about the ability of libraries to continue to seamlessly receive shelf-ready records along with items they have purchased from suppliers. One model suggested was to not break the link between library and supplier but to ensure that the record was deposited in the NBK at the same time as it was delivered to the library.
- According to Plan M, however, libraries may expect the NBK to be the primary source for new records. This will require rapid delivery and high availability after creation.

Collaboration

- One of the key drivers for success and one of the most beneficial outcomes of Plan M will be that libraries will have to work together to coordinate their activities and understand how best to exploit the opportunities that Plan M may provide.
- One of the early collaborative activities will be to share data on the costs of maintaining the data ecosystem as it currently stands (see '*Economics*' section below)
- Another activity will be to participate in the data requirements mapping exercise mentioned above (see '*Records*' section above).

- Advocacy with publishers and suppliers may be required to ensure that all actors across the ecosystem/marketplace are aware of the scope and scale of Plan M and its potential benefits for all participants
- Given the size and scale of their collections, all the legal deposit libraries will have a key role to play in maintaining the comprehensiveness and currency of the national data aggregation
- As the Plan M coordinator, Jisc obviously has a key role in orchestrating collaboration but other entities, especially SCONUL and RLUK will be key facilitators for the sector. The Jisc convened Task & Finish groups should provide key insights which will need to be joined up with the guidance coming from governance groups such as the Library Strategy Advisory Group (LSAG); and the Library Hub Community Advisory Board (LHCAB – formerly CMCAB). Initiatives such as the RLUK Collection Strategy Network should be an advocate for change, along with others such as the National Acquisitions Group (NAG), the CILIP Cataloguing and Indexing Group (CIG), and consortia (e.g. SHEDL, WHELF, White Rose, Northern Collaboration, etc.).
- There is an international dimension to the Plan M work in as far as lessons may be learnt and models might usefully be borrowed from other countries.

Economics

- There is acknowledgement that there must be a ‘pay to share’ model underpinning Plan M, both for record creation and for the provision of infrastructure to provide sharing mechanisms.
- There is also a determination to: only pay once (for the data to be created to a reasonable standard and for it to be made widely available); and for the overall future financial burden on libraries to remain roughly comparable to its current level.
- There is a need to better understand and quantify current library expenditure as result of acquiring data from the current data marketplace.
- Any cost modelling outcomes and preparations for negotiations or procurement will need to be carefully overseen and validated by senior library representative groups, particularly if is going to be done at a national level – which there is a strong appetite for.
- Procurement processes should be used more actively to pressurise suppliers to deliver fit-for-purpose data
- Ordinarily in the library domain, there is never one solution that either solves the entire problem or provides all the data. This will inevitably be true of Plan M (and the NBK) but it should nevertheless remain a success criterion that the transformed data ecosystem will provide significant long-term savings and efficiencies. This transformation should have the effect, over time, of demonstrating which parties are delivering most benefit into the marketplace.

The 10 Year Vision

As part of the Focus Groups discussion, library directors and senior managers were invited to articulate their thoughts on how the overall direction of libraries - and the strategies to take them in such a direction – might influence the formulation and implementation of Plan M. The conversation

was based on the idea that the NBK is anticipated to be a critical piece of national data infrastructure for the next few decades and it was therefore appropriate to imagine at least 10 years ahead to try and avoid designing any unhelpful path dependencies.

The following is a selection and a synthesis of some of the main points that were raised.

Anticipate & Prioritise

Whilst it is difficult to imagine precisely what today's 8-year olds will expect from their university library when they are studying for their undergraduate degrees in 2030, it is certain that the creation, communication, consumption and evaluation of knowledge will evolve. Some ideas include:

- Publishing as a service, with even more of a focus on dynamic changes to book editions, more chapterisation and atomisation of the text to draw out significant research, possibly using techniques similar to [nano-publishing](#).
- A concomitant increase in using text mining and text analysis techniques to take the place of close reading and full consumption of texts.
- Much more agnostic behaviour about the relative benefits of flipping between electronic and print texts according to the convenience afforded by either format for the specific task required.

It was felt that the current focus was very much on decisions required within the library about collections, space, resources, skills and budgets. Planning for the long term, the focus needs to shift decisively onto user needs and the future requirements and opportunities for research.

Implications for Plan M

The data that we need to bring into the NBK aggregation must extend as comprehensively as possible into the eBook domain and must be usefully descriptive at a granular level. Metadata standards may need to evolve or adapt to describe non-traditional formats of research.

Digital Scholarship

This terminology has gained ground recently and is a more broadly encompassing term than 'digital humanities' – which has discipline-related connotations. The expectation is that researchers will be looking to use collection-related descriptive data alongside primary data sources to create novel mash-ups and visualisations with a suite of emerging and evolving analytic and processing tools. It seems likely that an enhanced institutional and sector-wide focus on promoting digital scholarship will require libraries to pay particular attention in the following ways:

- 3-dimensional metadata – bringing together all of the metadata that libraries can pull together to enhance and complement their bibliographic collections
 - Library data
 - Archival data
 - Museum data
 - Repository data
 - Research data

- Linked data with a particular emphasis on the use of
 - Persistent identifiers
 - Authority controls
 - Linked data formats to enable connections

These ideas further encouraged the idea that at some point MARC data may become redundant (see 'Records' section above); but a note of caution was also sounded about how expensive it may be to build in a critical reliance on authority-controlled data.

Implications for Plan M

Resources will need to be identified and opportunities sought to work with researchers to explore how the data in the NBK aggregation can be used for novel types of research. Exports in appropriate formats will be required and licenses negotiated with third-parties that allow the data to be integrated and visualised into research outcomes. Linked data instances of the NBK will need to be created and maintained requiring convincing business-cases around the impact this could have on research.

Data Science

There is broad scope for libraries to work more intensively with data scientists to better understand how researchers and learners most effectively discover, consume and exploit information resources. Over the next ten years, there may be a decreasing reliance on catalogues if there is an increase in full-text indexing in concert with machine learning to more accurately predict (or even possibly dictate) the research choices that individuals make. With access to vast stores of textual information on every conceivable research topic and enormous amounts of computational power, significant changes in the way that research is conducted may be in the offing. What role libraries can play might depend on how invested they become in owning the skill-sets required to participate in this kind of work

Implications for Plan M

These scenarios feel like they may be significantly disruptive towards the end of a 10-year period (starting now). In the meantime - and by way of facilitating the change described - libraries will still need to provide inventories of the resources that are available to research.

National Coordination

The discussion returned at various times to the need for bold and ambitious thinking and for Jisc to be helping to coordinate activity at the national level wherever possible. The following issues were discussed:

- Efficient shared print storage
- Coordinated data-driven digitisation programmes based on rarity, usage and value
- Coordinated digital preservation activities

- Advocacy and policy towards more open data and more openly available resources
- Better alignment and continuity of data through the book supply chain and into the library and discovery environment
- An increased expectation on centralised and routine cataloguing being part of data infrastructure provision rather than a visible role in the library
- An increased focus in the library on the cataloguing of special collections and surfacing of hidden collections to offer new value to research and teaching
- For the scale and complexity of managing monographs and monograph data to become as routine and achievable as it now appears to be with serials

Implications for Plan M

The appetite for coordinated national action has very fundamental implications for the NBK programme and Plan M in particular. The types of activities noted above: digitisation, digital preservation, collection management, routine and special cataloguing, will all be achieved more efficiently and effectively if the data in the NBK is comprehensive, current, fit-for-purpose and as open as possible. These are principles that we have been working towards with the NBK programme but it is clear that they can usefully be transposed onto any negotiations and/or procurement processes that will be undertaken as part of Plan M.