Contributor Guide to the National Bibliographic Knowledgebase

August 2019

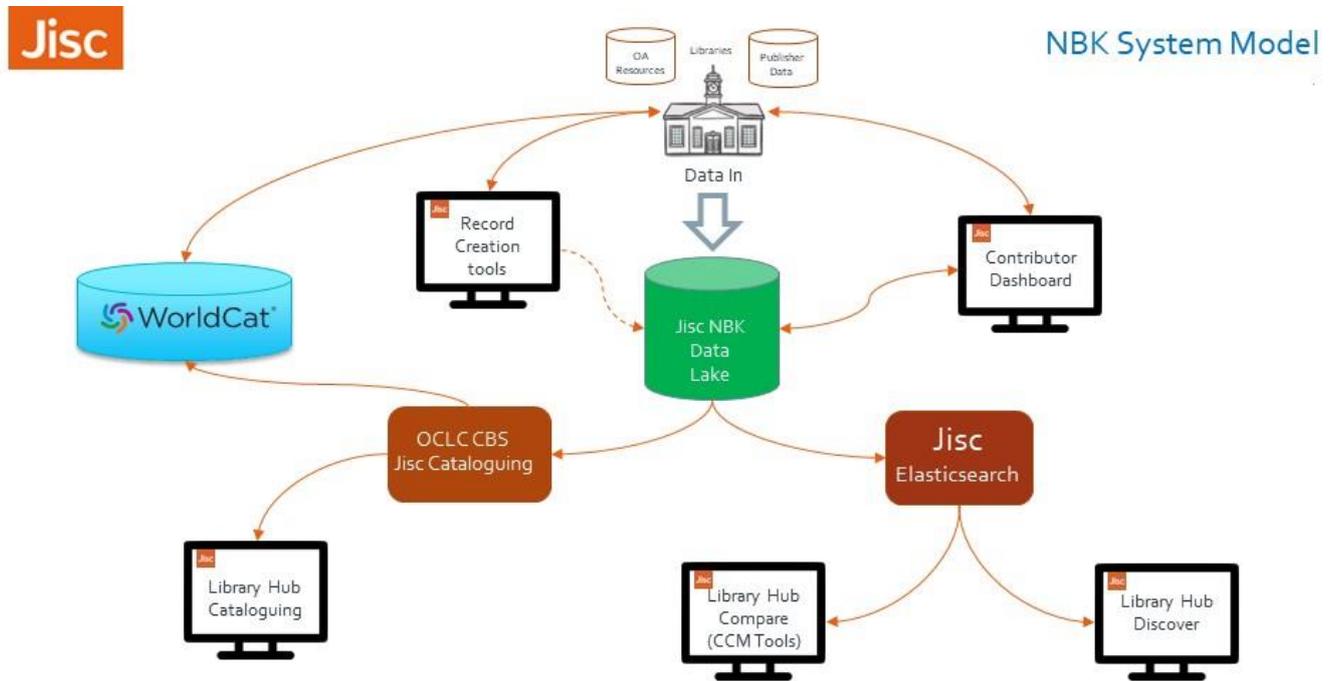# 1. What is the National Bibliographic Knowledgebase?

1.1

The Jisc National Bibliographic Knowledgebase (NBK) development is building on the long-term success of the Copac and SUNCAT services, expanded to include the catalogues of all university libraries wishing to be involved, along with a range of other data sources. The work has resulted in a shared database on a truly national scale supporting a range of Jisc services for the academic community.

The NBK is the aggregation of data which is being used to provide the Library hub discover, compare, and cataloguing services. This large data pool allows us to manage the flow of different subsets of data into each service and will support new service development in future.

These library hub services have now replaced Copac, SUNCAT, and CCM Tools.

1.2 Data model

The NBK system model relies on two databases: one for catalogue record download, run on the OCLC CBS system; and one for resource discovery and collection management, run on ElasticSearch.

The above figure represents how data flows into the various components and interfaces of the NBK system. All data that is sent to the OCLC CBS system for inclusion into the copy cataloguing system is of reasonable or good quality and excludes records with known licensing restrictions. All data that is submitted for inclusion into the NBK will be included in the Discovery instance of the system and will represent a comprehensive view onto the holdings of NBK libraries.

This divide means that we can make decisions about the data in both databases to meet the differing needs of cataloguers and resource discovery. It also makes it simpler to manage data that isn't licensed to be shared in MARC format. Where libraries tell us that there's data in your catalogue export that can't be shared in MARC format, that data won't be loaded onto the CBS database. If you've told us that you'd like your whole catalogue to be excluded from MARC sharing, or if your data is not natively in MARC format, then it won't be sent to OCLC and will only be used on the resource discovery and collection management database.

As shown in the diagram, only data from CBS will be loaded onto WorldCat.

1.4 Data ownership

The NBK is a Jisc database, but Jisc does not assert ownership over contributed data held within the NBK. Details about who is entitled to access each service, as well as options for controlling the re-use of MARC21 records, are included in the summary information about each service given below.

The cataloguing database aspect of the NBK platform is managed by OCLC on behalf of Jisc. Data supply and data management will continue to be through Jisc. OCLC does not assert ownership over contributed data held within the NBK.
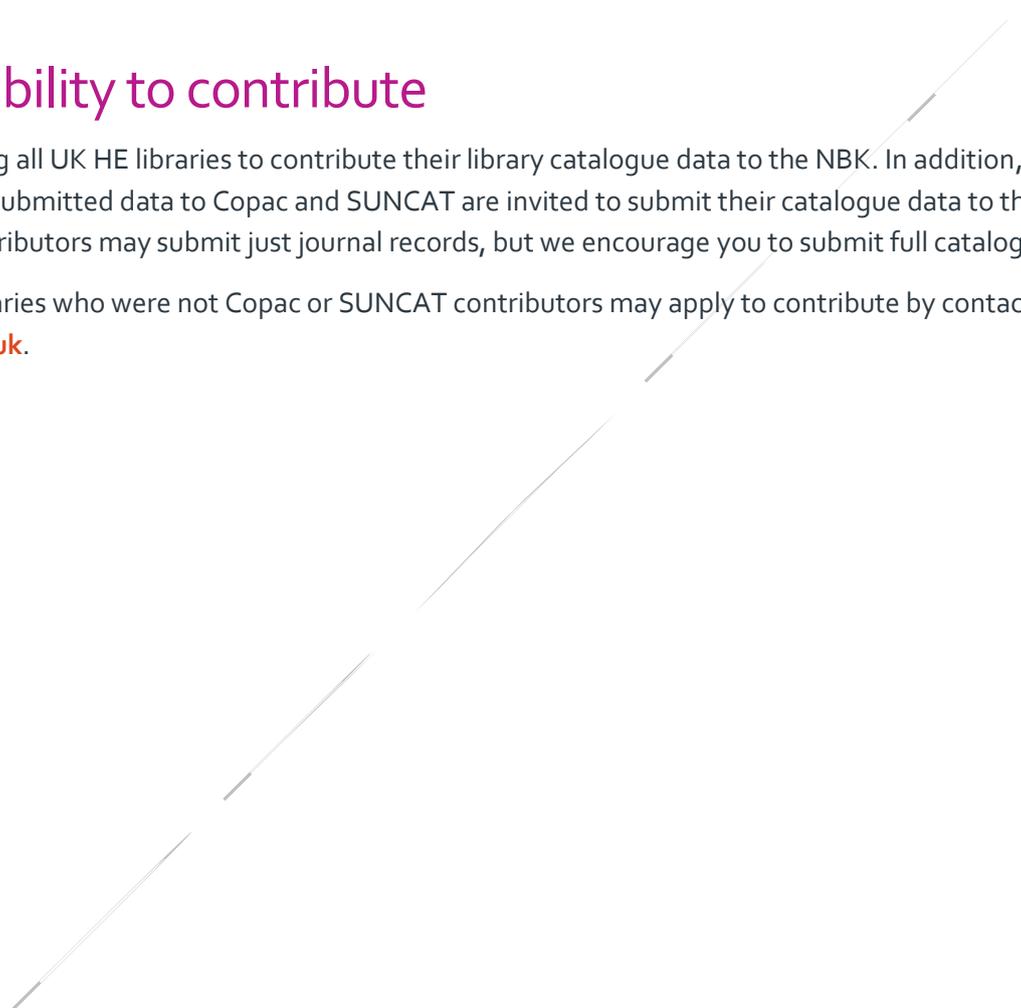
Records will be made available for re-use in a variety of formats through the different NBK services. Records that are distributed in MARC format will indicate the NBK as the immediate source of the data, including eg:

»   The Jisc (Copac) MARC code

»   The NBK record ID

1.5 Eligibility to contribute

We are inviting all UK HE libraries to contribute their library catalogue data to the NBK. In addition, non-HE libraries who submitted data to Copac and SUNCAT are invited to submit their catalogue data to the NBK. SUNCAT contributors may submit just journal records, but we encourage you to submit full catalogues.

Specialist libraries who were not Copac or SUNCAT contributors may apply to contribute by contacting **nbk@jisc.ac.uk**.

2. Data supply and processing

To include your catalogue on the NBK we need you to send us an export of your data. Data supply is usually via SFTP using public/private key authentication, however if this is not possible other options can be discussed.

Our aim is for library hub discover and compare to reflect your local public catalogue as closely as possible, ensuring consistency between your local catalogue and the NBK. This helps us provide the best possible overview of the national distributed collection, to support resource discovery and collection management activity. After the initial load we need regular updates to ensure data currency; depending on your library system we may be able to harvest updates.

For resource discovery and collection management, our minimum record quality requirement is that we require records to be processable, with a non-blank title field, and (optionally) holdings information. This inclusivity helps us to ensure comprehensive coverage. The cataloguing database has higher record quality requirements.

All incoming data is checked and, where necessary, standardised to MARC21 to support shared cataloguing and record deduplication. Contributors will receive reports from the cataloguing database workflow regarding any records dropped as well as potential data errors identified. These reports will start being generated in late 2019, once the process and requirements are finalised.

All the incoming records are retained within the NBK data pool unless/until a contributor sends a deletion for a record. This data is held in MARC21 format and this original data is updated each time a contributor sends an update file. From the NBK data pool there are two data flows of appropriate data to each of the Cataloguing database and the Discover database. In each case, incoming records are deduplicated against existing data and where a match is successful a consolidated record is formed, containing the holdings information for all the records within the set; if a match fails the incoming record is added as a single record. These merged data sets underpin the range of NBK services and are updated in line with the NBK data pool.

2.1 Contributor website and dashboard

Information for NBK contributors and potential contributors is held on our Contributor website, **https://contribute.copac.jisc.ac.uk**. This website is access restricted. Shibboleth and OpenAthens users may login with their usual credentials. If you do not use either of these identity providers, please contact **nbk@jisc.ac.uk** to arrange access.

Our Contributor Dashboard is at **https://contribute.copac.jisc.ac.uk/dashboard**. This may contain sensitive information, and as such requires an extra level of security to access. Please see details on the contributor website or contact us for more information.

On the Contributor Dashboard you will find the connection details to transfer data to us, and information about when we last received files from you. There are also four questionnaires we ask you to complete: data supply, circulation data, library information, and data management.

3. Your data on the Jisc library hub services

Once loaded we will make your data available through a range of library hub services to support your individual and collective activity across many areas. The services that have been released are detailed below, with other services being planned, particularly around eBooks.

We encourage contributors to supply a full copy of their local catalogue, however, we are aware that there can be concerns over record sharing, particularly in MARC format. So each contributing library can control the way their data is used within the NBK in two main ways:

❖ by requesting that their whole catalogue is suppressed from data sharing in MARC format,
❖ by specifying how to identify individual records for exclusion from data sharing in MARC format (eg. records from a specific eBook publisher).You can supply this this information in the Data Management questionnaire on the Contributor Dashboard

The excluded records will not be available for shared cataloguing and not be exposed to data aggregators (eg WorldCat). However, they will be visible on Discover and Compare in a non-MARC format (details below).

When a library supplies a data file we would expect that records not displayed locally would be excluded from the file during export. However, if this is not possible please provide this information in the Data Management questionnaire.

3.1 Library Hub Cataloguing service

A shared cataloguing service will support efficient data management within HE libraries. This is freely available to Jisc HE and FE member institutions, as well as other NBK contributors whose data is included in the cataloguing service.

❖ User access: A Library Hub Cataloguing service will be provided through **web** and Z39.50interfaces, with authentication restricting the service to authorised users (Jisc members and NBK contributors)
❖ Where contributor data is supplied in non-MARC format we will exclude this data from the cataloguing service as the automated data conversion is likely to create quite basic MARC, suitable for resource discovery but without human checks for data quality and accuracy.

In terms of data sharing in MARC format a contributor may request that:

❖ their whole catalogue is suppressed from the Library Hub Cataloguing service
❖ specific, identifiable, records are suppressed from the Library Hub Cataloguing service.

NB records that are supressed from Cataloguing will also not be synced to WorldCat.

3.2 Library Hub Discover service

Library Hub Discover Interface

❖ User access: the **Library Hub Discover** service is freely accessible to a UK and international audience through a dedicated web interface as well as a range of machine-to-machine interfaces. These will support resource discovery for the academic and library community, as well as providing opportunities to use the data for research (eg. establishing rarity; reviewing publishing patterns), with potential for new forms of data visualisation. Data will be available for export in non-MARC formats.

❖ Catalogue back-up: If you supply local holdings data you will be able to use Library Hub Discover as a back-up for your local catalogue during periods of planned (or unplanned) downtime.

❖ Live circulation data: Where possible we will use Z39.50 to retrieve live circulation data from your catalogue within the Resource Discovery service, ensuring availability information is always current. This will only be made available when a searcher asks to view a holdings display, avoiding unnecessary search requests to your local catalogue.

❖ We will promote library collections via twitter and encourage contributors to write Features on the NBK blog, e.g. focussing on a forthcoming exhibition or a specific collection. Our Library Information pages will give users information about your collections and links to your local services.

Note: We encourage contributors to include a link to **https://discover.libraryhub.jisc.ac.uk** within their own local catalogue search results. There will also be a range of APIs to allow machine-to-machine access of the data from your local systems.

NBK Data exposure to search engines

We will expose NBK data through the Discover service for harvesting by search engines, maximising collection exposure. This will guide those searching the web through to Library Hub Discover and thence to resources held by individual contributing libraries. We will also support linking from external sites/services into individual Library Hub Discover records, again bringing users through to individual contributor resources. Data exposure is in a non-MARC format. If, at some point, we do provide exposure of MARC data this would exclude suppressed records.

NBK data supply to aggregator services

We will supply data to aggregator services, again widening the visibility of each contributor's resources. In the first instance we will be synchronising a subset of the NBK cataloguing database with WorldCat, but we anticipate that other aggregators will also be interested in working with us. WorldCat sync is only available to Jisc member libraries who supply their data to us in MARC format. Only data which is non-restricted and has been loaded onto the Cataloguing database will be synced to WorldCat.

❖ A contributor may request that their whole catalogue is suppressed from the NBK Aggregator Data Supply service.

❖ A contributor may request that specific, identifiable, records are suppressed from the NBK Aggregator Data Supply service.

Automated synchronisation of contributor data with WorldCat will mean that existing WorldCat contributors can manage their data supply to WorldCat via the NBK, removing the need to send data to two separate services if they find that helpful. This automated synchronisation will be available for Jisc members.

Note that once a library's data is supplied to an aggregator database the data will then be subject to the data management practices of that database. So any library whose data is included in an aggregation will be agreeing to the requirements involved in participation in that database. You can indicate in the Data Management questionnaire if you are happy for your data to be aggregated.

3.3 Library Hub Compare: Collaborative Collection Management tools

The NBK data underpins the Library Hub Compare service (**https://compare.libraryhub.jisc.ac.uk**) , a development of the Collaborative Collection Management Tools (CCM Tools) service. It offers a range of facilities to support collection management activity, eg: identifying last copies, visualising collection strengths, developing new collections, assessing materials for digitisation or conservation.

User access: compare is available to Jisc HE and FE members, through UK Access Management Federation authentication (**http://www.ukfederation.org.uk/)** as well as to other NBK contributors through the Copac Identity Provider.

Library Hub Compare is focussed on aggregated data use and does not make individual records available in MARC21 format.

3.4 New services

The NBK offers new potential and we expect the range and nature of services will expand and evolve as the work develops. Initial areas for development include:

» Data enhancement: We will work with contributors to see how we might facilitate improvements to catalogue coverage and quality, making uncatalogued collections accessible and upgrading record quality where appropriate. Complete and accurate data is fundamental to the successful deduplication of records, which is not only important for resource discovery, but underpins an effective collection analysis and management service. So we will be looking at potential new facilities and services in this area. We also intend to work with the community to develop minimum data quality requirements for the Library Hub Cataloguing database.

» eBook management: This is an area that will develop in discussion with the library community. We will also be harvesting records for a wide range of open access electronic materials, increasing the range of full-text data that is readily accessible to the academic community. This will begin with harvesting Hathi Trust records for public domain materials.

» We will be looking at new opportunities for contributor and user support that exploit the clearer picture of the national distributed collection presented within the NBK, eg. looking at new opportunities for data analysis and visualisation.

4. More information

If you have questions about any aspect of contributing to the NBK please get in touch through the NBK Contributor Helpdesk: nbk@jisc.ac.uk