Contributor Guide to the National Bibliographic Knowledgebase

August 2022

1. What is the National Bibliographic Knowledgebase?

The Jisc National Bibliographic Knowledgebase (NBK) builds on the long-term success of the earlier Copac, CCM Tools and SUNCAT services, expanded to include the catalogues of all university libraries wishing to be involved, along with a wide range of other libraries and data sources. This is a collaboration that has produced a shared database on a truly national scale, supporting a range of Jisc services for the benefit of the academic community.
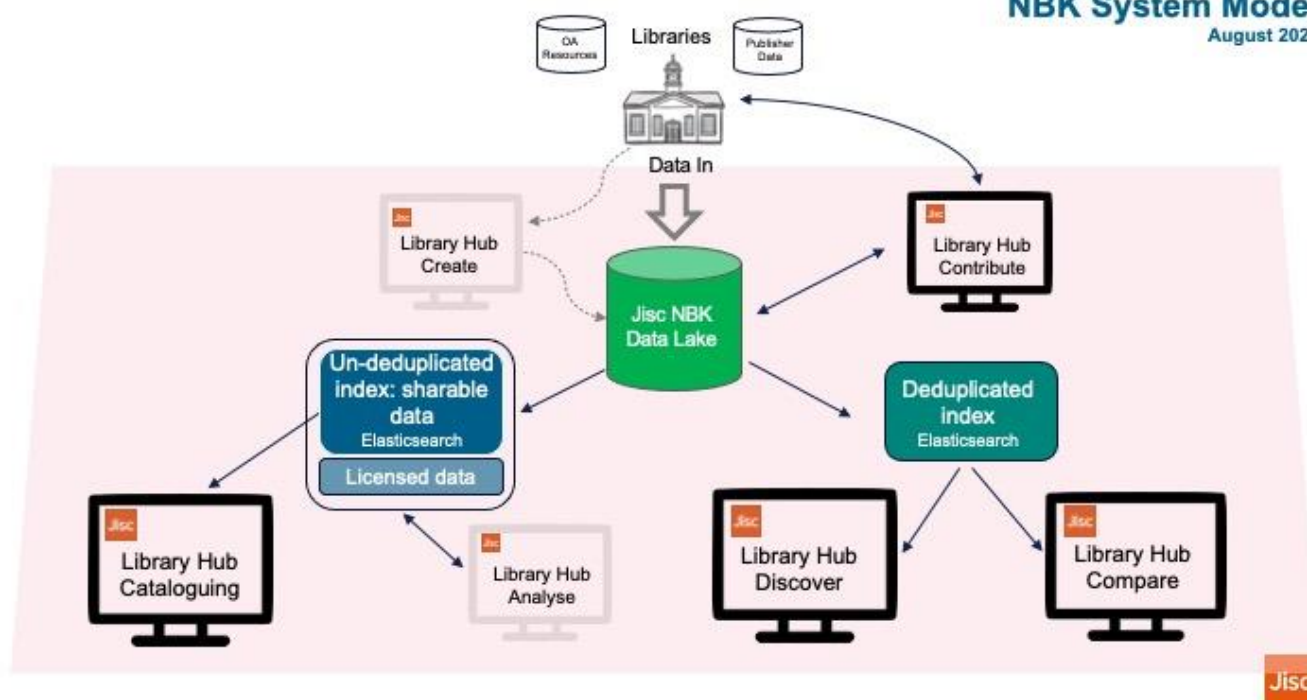
The NBK is the aggregation of data which is being used to provide the Library Hub Discover, Library Hub Compare and Library Hub Cataloguing services. The NBK data lake allows us to manage the flow of different subsets of data into each service and will support new service development in future.

1.1 Overview

The NBK system model is illustrated below, showing how data flows through the various components of the NBK system. To create the NBK, data is supplied by, or harvested from, a wide range of contributors and incorporated into the NBK data lake. This data is then selectively used to generate two other databases:

- An undeduplicated database: this supports the Library Hub Cataloguing service. This gives access to MARC format data, excluding those records with known licensing restrictions. As the data is loaded, MARC format issues are noted in each record and a 'quality' score assigned. In addition to library contributed data, the undeduplicated database includes BNB records and we expect to add publisher sourced records in future.

- A deduplicated database: this gives access to data in non-MARC format. It includes all the library catalogue data that is submitted for inclusion into the NBK, providing a comprehensive view onto the holdings of NBK contributing libraries within Library Hub Discover and Library Hub Compare.

This separation into multiple databases means that we can make decisions about the data to meet the differing needs of cataloguers, searchers and collection managers. It also makes it easier to manage data that isn't licensed to be shared in MARC format.

NBK System Model
August 2022

The other main element of the NBK is the Library Hub Contribute service, which supports data supply activity, acting as a shared information resource for NBK staff and contributing libraries. Library Hub Analyse and Library Hub Create are potential services that we have been considering for the future. Analyse would support library review and upgrade of data, whilst Create would offer a route to contribute to the NBK for small specialist libraries with no local catalogue system.

1.2 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.

Records are made available for re-use in a variety of formats through the Library Hub services. Records that are distributed in MARC format indicate the NBK as the immediate source of the data, including:

- The Jisc (NBK) MARC code
- The NBK record ID

1.3 Eligibility to contribute

We encourage all UK HE libraries to contribute their library catalogue data to the NBK. In addition, we are keen to hear from specialist libraries, with collections of interest to the national or international research community, who may apply to contribute.

1.4 Contact

If you would like to become involved by becoming an NBK contributor, or you have any questions about the process, please contact us at: **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, or data harvest via OAI-PMH, however if this is not possible other options can be discussed.

Our aim is for Library Hub Discover and Compare to reflect your own public catalogue as closely as possible, ensuring consistency between your local catalogue and the Library Hub services. This helps us provide the best possible overview of the national distributed collection to support resource discovery and collection management activity. It also ensures clarity for service users regarding the materials covered by the Library Hub services. After the initial load we need regular updates to ensure data currency, the frequency can be dependent on the amount of change in your catalogue.

All incoming data is checked and, where it is not native MARC, standardised to MARC21 to support record deduplication. We are looking at the best way of supplying data feedback on update files in future.

For resource discovery and collection management, our minimum record quality requirement is that records are processable, with a non-blank title field; and ideally include holdings information. This inclusivity helps us to ensure comprehensive coverage. In addition, for the cataloguing database, any records that are not natively MARC21 are excluded, as are records that a library indicates they would like to exclude on quality or licensing grounds.

All the incoming records are retained within the NBK data pool unless/until a contributor sends a deletion for a record. The data is held in MARC21 format and this original data is updated each time a contributor sends an update file. Incoming records are deduplicated against existing data and where a match is successful a consolidation 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. From the NBK data pool there are two data flows of appropriate data to each of the undeduplicated Cataloguing database and the deduplicated Discover/Compare database. These 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 Library Hub Contribute, **https://contribute.libraryhub.jisc.ac.uk**. The basic website is access restricted; Shibboleth and OpenAthens users may login with their usual credentials. As a contributor, if you do not use either of these identity providers, you can contact **nbk@jisc.ac.uk** to arrange access.

Library Hub Contribute has our Contributor 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, as well as information about when we last received files from you. There are also four questionnaires we ask you to complete: Data supply, Data management, Circulation data and Library information. For small libraries that are not supplying MARC data we have an alternative, shorter, questionnaire. Thre are plans to incorporate more information into the Dashboard regarding the way your data is being managed, as well as other service and data feedback.

3. Your data on the NBK

Once your catalogue is loaded into the NBK we will make your data available through Library Hub services to support your individual and collective activity across many areas. The current services are detailed below.

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

- by requesting that your 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 will not be available to data aggregators. However, they will be visible on Library Hub Discover and Compare in a non-MARC format ensuring full visibility of your collection.

When you supply a data file we expect that records not displayed locally are excluded from the file during export. However, if this is not possible you can provide information about exclusions in the Data Management questionnaire.

3.1 Library Hub Cataloguing service

The Library Hub Cataloguing shared cataloguing service (**https://cataloguing.libraryhub.jisc.ac.uk**) supports 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: Library Hub Cataloguing has web and Z39.50 interfaces, with authentication restricting the service to authorised users (Jisc members and NBK contributors)

- Where contributor data is supplied in non-MARC format we 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.

- Contributor records are stored individually allowing for local preferences in record source. We check for basic MARC errors and include a note in each record along with 'quality' measures to help with record selection.

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.

Records that are supressed from Cataloguing will also be excluded from any data supply to other another aggregator in future. Participation in any such aggregator data supply would always be optional..

3.2 Library Hub Discover service

The Library Hub Discover service (**https://discover.libraryhub.jisc.ac.uk**) is freely accessible to a UK and international audience. It facilitates 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):

- User access: There is an easy to use and flexible web interface as well as APIs, with future potential for new forms of data visualisation. Data is available for export in non-MARC formats.

- Catalogue back-up: If you supply local holdings data, during periods of planned (or unplanned) downtime you can point your local catalogue users to Library Hub Discover, giving you an informal back-up for your local catalogue.

- Live circulation data: Where possible we use Z39.50 to retrieve live circulation data from your catalogue within Discover, ensuring document availability information is always current. This live information only appears in the holdings display for your data, avoiding unnecessary search requests to your local catalogue. We are considering access to availability data through proprietary routes in future.

- We promote library collections via twitter, mailing list, and on our blog. We encourage contributors to write Features on the NBK blog, e.g. focussing on a forthcoming exhibition or a specific collection, see: **https://libraryservices.jiscinvolve.org/wp/tag/features/.**

- Our Library Information pages give users information about your collections and links to your local services (**https://discover.libraryhub.jisc.ac.uk/about/libraries/).**

Note: As a contributor we ask you to include a link to **https://discover.libraryhub.jisc.ac.uk** within your own local catalogue search results where your system supports this. There are a range of APIs to allow machine-to-machine access of the data from your local systems; details at: **https://discover.libraryhub.jisc.ac.uk/support/api/**

NBK Data exposure to search engines

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

NBK data supply to aggregator services

We are interested in the potential supply of data to aggregator services, again widening the visibility of each contributor's resources. If this takes place, as a contributor you can request:

- that your whole catalogue is suppressed from the NBK aggregator data supply.

- or that specific, identifiable, records are suppressed from the NBK aggregator data supply.

Note that once a library's data is supplied to an aggregator database the data would 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 supplied to another aggregator.

3.3 Library Hub Compare: Collaborative Collection Management tools

The NBK data underpins the Library Hub Compare service (**https://compare.libraryhub.jisc.ac.uk**). Compare uses the deduplicated NBK data and 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. Data can be downloaded and used with locally generated data to improve your understanding of aspects of your collection, assist in reporting and support work with academic staff on collection related activity.

User access: Library Hub Compare is available to Jisc HE and FE members, as well as to other NBK contributors. Shibboleth and OpenAthens users may login with their usual credentials. As a contributor, if you do not use either of these identity providers, you can contact **nbk@jisc.ac.uk** to arrange access.

Library Hub Compare is focussed on aggregated data 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 over time. Initial areas for possible development include:

- Data reporting and enhancement: We will work with contributors to see how we might facilitate improvements to catalogue coverage, reporting on data issues and supporting data upgrade 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.

- eBook management: This is an area that will develop in discussion with the library community. We are harvesting records for a wide range of open access electronic materials, increasing the range of full-text data that is available for shared cataloguing and readily accessible to the academic community.

- Metadata research: 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, e.g. 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**