

## **Accuracy, Matching & Curation Activities**

Accurate data lies at the heart of ISNI's mission to correctly recognize public identities and to assign an ISNI identifier to each. It deploys a number of approaches to detecting data inaccuracies and to rectifying incorrect assignments that may have been made.

### **Data inaccuracy**

Inaccurate data is difficult to detect by program but some systematic checks have been put in place. The ISNI system includes for example a date anomaly check program that looks for creators who are recorded as having created before their 10<sup>th</sup> birthday. This program is run regularly and the results are reviewed manually: this can highlight, for example, a typographical error or a situation with two or more persons in the same record. Naturally, due attention is paid to records for child prodigies such as Wolfgang Amadeus Mozart that may appear to be anomalous when first checked in this way!

Other checks are made by searching for unusual source combinations. The ISNI Quality Team regularly samples the database and incoming test data from new or established sources. Data submitted for batch loading is rejected if the error rate for duplicates and mixed identities is above 5%. Results of database samples may indicate corrections that can be made programatically or areas in the database that require manual review.

### **Two or more persons or organizations in the same record**

A situation where two or more persons or organizations feature in the same record may occur because of errors in the incoming data or because of matching errors. Where records needing splitting because two or more persons or organizations are found in the same record (same name but different identities), the matching program is reviewed and may be refined.

To date, program modifications have been successful for all reported errors. A delicate balance needs to be struck to ensure the maximum amount of matching without erroneous matching creeping in. As a compromise, a class of "possible match" has been devised and data contributors are invited to help resolve these instances by checking them manually. Where it is determined that an ISNI record includes two or more separate identities, the ISNI Quality Team is notified and performs the splitting procedure that automatically splits off good data to new ISNI records and notifies all sources concerned. The old ISNI record is then deprecated.

### **Two or more identities of the same person in the same record**

Some sources regard pseudonyms as name variants not separate identities. The ISNI system includes a suite of programs to detect these cases and convert them appropriately. ISNI's policy is to assign different identifiers to separate public identities of the same person. The relationship between the identities may be designated as public or private. Where the relationship is public, a link is made

between the two records in the ISNI database IF the data is also coming from a separate source.

### **Duplicate assignment**

The ISNI system deliberately errs on the side of not making a false association. Bearing this in mind, duplicate assignment may from time to time occur. When a duplicate is detected, the two ISNI records concerned are merged and the former (now deprecated) ISNI is retained in the record, such that a search or link using the deprecated ISNI will always result in the record being retrieved or resolved. All sources concerned are notified of merges in a dated notification message that remains on the database record. The public are also invited to report duplicates via the ISNI web search facility and all ISNI member organizations are able to resolve duplicates.

All cases of seemingly inexplicable failure to merge are examined with a view to finding a remedy. For example once ISNI found a record that was not matching because of a zero in the publication date of one of the associated titles.

Where records are merged, the ISNI that is regarded as the “retained ISNI” will be determined according to the following decision sequence:

- A record with assigned status trumps a record with provisional or other status (the ISNI has already been diffused)
- The record contains a “preferred source”, currently either JISC Names (JNAM) or Ringgold (RING)
- The ISNI with the most sources (therefore where the ISNI has been more widely diffused)
- If still no ISNI can be chosen, the one that was first minted is selected and retained.

In this context, preferred sources are recommended by the ISNI Technical Advisory Committee and approved by the ISNI-IA Board as being base sources in their own areas with high quality curated data that has been disambiguated.

### **Requests for data removal or deletion**

ISNI data sources (including Registration Agencies, Members and other data contributors) may if necessary request that their data be removed at any time. If the source concerned is the last remaining source for that record, the core metadata, including at a minimum name and ISNI number, is retained and the source of the data is re-designated as "ISNI". For example, a source may no longer need a record for an organization that has ceased to exist. In this case, the source is changed to "ISNI" and the ceased date may also be recorded.

Via the web interface on the ISNI website, a data subject (whether an individual or an organization) is able to request deletion of data elements or complete removal of the ISNI data. Such requests are reviewed, verified and actioned by the ISNI Quality Team.