

JISC Name Identifier Survey

- Dutch Response -

Reviewers: Maurice Vanderfeesten (SURF), Magchiel Bijsterbosch (SURF), Martin van Muyen (OCLC)

Last update: 2011-09-26T15:25

Status: FINAL

1. *What was the motivation for developing the identifier system?*

The DAI, Digital Author Identifier, in the Netherlands was developed to prevent ambiguity in retrieving the work of one author, and to uniquely identify an author by a numeric identifier in order to overcome the problems imposed by name variants.

2. *Which organisation(s) is (are) responsible for the identifier system?*

SURF is responsible for the governance and strategic developments of the overall identifier system.

The system consists of a central component - the National Thesaurus for Author names (NTA), part of the Shared Cataloguing System (GGC) - and decentral components - the local Current Research Information System (CRIS) - located at each university. Currently, all universities have implemented the METIS system as their CRIS solution.

OCLC is responsible for technical application management ([BiSL](#): Application management) of the GGC and by that, the NTA. The partner who is responsible for the functional management (BiSL: Business information management) is still under discussion at the UKB meetings (UKB is the collaboration between the research libraries and the National library). The National Library is a candidate. The CRIS systems are under the responsibility of the universities and research institutes.

3. *What is the scope of your identifier system, in terms of the type of people it covers? (For example, does it include: book authors, active current researchers, formerly active researchers, doctoral students, masters students etc.)*

The GGC is part of a generic bibliographic cataloguing database, thus the NTA contains a broad spectrum of authors from novelists to scientific authors.

By design a DAI is a promoted Pica Production Number (PPN) resulting from a positive match between a person identified by a record in the NTA and an appointment registered in the local CRIS system. This makes that a DAI can only be assigned to researchers appointed at a university.

The match against the NTA theoretically also implies that the researcher must have previously published to be included in the NTA. However when an NTA record has not been found when trying to create match with a local CRIS record, an NTA record may be created from the CRIS record in order to still be able to mint a DAI.

We are gradually widening the scope to researchers appointed to non-university research institutes and universities of applied sciences, that traditionally do not have a local CRIS.

4. *How is your system populated with data? (by researchers themselves/their institutions/funding bodies)*

The NTA contains authors who are registered by Public Libraries and Research Libraries who are member of OCLC (Previously OCLC-PICA).

Population of the data is done by qualified cataloguers at libraries or specialist departments, and

employees who administer the CRIS currently used by all Dutch universities, METIS.

5. *Who is authorised to make changes to the information in the system?*

Changes in the central system may be made by cataloguers at university libraries and research information departments with access to WinIBW or WebGGC and METIS administrators, through a special version of the WebGGC. Note that each research institute is only allowed to make changes to the record-part associated with their own individual institution.

6. *How are identifiers assigned?*

As said, the identifier is minted by a positive match between a person record in the NTA and an appointment record in the local CRIS.

From within the METIS application, the administrator uses a special version of the WebGGC for the NTA that preserves the context of the person selected in METIS. The administrator then searches the NTA for a positive match. If a match cannot be found, the administrator may create a new NTA record. The administrator then registers some metadata for the person according to the local METIS (e.g. name, date of birth, employee id, appointment date). Upon completion of this process, the resulting DAI is sent back to the METIS.

Institutions that do not use the METIS application can use either the WinIBW application (a generic application used by libraries that are already cataloguing in the GGC) or the WebGGC webinterface.

7. *What form does the identifier take?*

<http://wiki.surffoundation.nl/display/standards/DAI#DAI-Syntax>

The DAI (or PPN) contains 9 to 10 characters. The first 8 to 9 characters are numbers. The last (9th or 10th) character is a control character. The control character is a modulus 11 check digit, like in the ISBN. More on the MOD11 algorithm can be found on for example [wikipedia|http://en.wikipedia.org/wiki/MSI_Barcode#Mod_11_Check_Digit].

A DAI is for example: 123456785

A URI-fied DAI looks like this: `info:eu-repo/dai/nl/123456785`

The DAI is the number after the string `info:eu-repo/dai/nl/`. A DAI is a number like 123456785. The last character is a MOD11 check-digit.

The string: `info:eu-repo/dai/nl/` is just an authority namespace, telling the user or machine that the number is a DAI originating from the Netherlands.

At the moment the INFO-URI namespace is used as an authority namespace. The DAI is URI-fied under the EU-REPO sub-namespace. This namespace defines components for compound objects in the Institutional Repositories.

Click here to read more about the [namespace specifications info:eu-repo/dai/nl/](#)

In practice the DAI can be used to unambiguously retrieve publications from one author across different systems. [See the for an example in NARCIS.](#)

8. *What information is maintained in the system? (e.g. names, alternative forms of names,*

email addresses, dates of birth, institutional affiliation(s), details of publications, details of grants received/applied for) Are any standard metadata schemes supported?

According to the registry of the Dutch Data Protection Authority (Dutch DPA), the NTA data that is allowed to be shared among catalogues for dis-ambiguifying author names. The sharable NTA fields are: name data (including name variances and pseudonyms, etc.), maiden name, date of birth, date of death, sex, title, work relation.

See for the register: <http://www.cbweb.nl/asp/ORDetail.asp?moid=808d858982>

The GGC is based on CBS, the Cataloging system of OCLC. The metadata scheme includes various field including names, alternative forms of names, email addresses, dates of birth, institutional affiliation(s), dates of the institutional affiliation(s) and much more. more about CBS can be found here: <http://www.oclc.org/cbs/>

CBS uses a multitude of metadata format standards such as MARC21 and UNIMARC. The Dutch cataloguers are constrained to extra rules for inserting names (especially for Dutch name separations). These can be found here: <http://www.oclc.org/nl/nl/support/documentation/ggc/persoonsnamenformaat.pdf> (Dutch only, version from 2006)

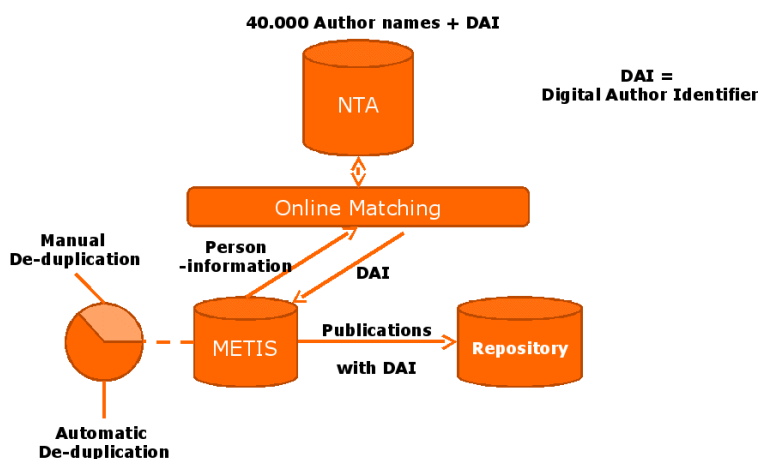
The DAI system XML Export format can be found in the Appendix of the workflow document http://wiki.surffoundation.nl/download/attachments/3473692/2006-06-27_DAI_werkprocessen.pdf

The METIS system is compliant to the CERIF datamodel. more: <http://en.wikipedia.org/wiki/CERIF>

The high level METIS datamodel can be found here: <http://aptest.uci.kun.nl/metis/service/Metisguide/Inleiding-alg.htm#diagram>

9. *With which other systems (if any) does your identifier system interact?*

The NTA interacts with METIS which in turn interacts with institutional repositories that are harvested by the NARCIS metadata aggregator.



See Appendix B for a more wider and complete overview.

10. *Is the information in the system made available to other services?*

At the moment of writing, there are no direct interfaces between the central NTA system and third-party services. However through exposure by repositories, NARCIS and other systems linked to METIS, such as employee pages and reporting systems on data warehouses the number of systems actually consuming may potentially be numerous. These are by no means however part of the governed infrastructure.

REMARK!

The data in the NTA contains personal data. This data is protected by Dutch law. Users of the NTA have permission to use this data only for bibliographic purposes. This permission has been granted by the Dutch Data Protection Authority (Dutch DPA). The permission has been registered. The registration can be found here: <http://www.cbpweb.nl/asp/ORDetail.asp?moid=808d858982>

Only services that meet the restrictions are able to access the data behind the identifier. The Identifier itself can be used in other systems that are related to bibliographic usage.

11. *Is there a licence on the data? If so, what is the licence?*

Yes there is a licence on the data. The use of the data is restricted to OCLC and its Dutch members of the formerly OCLC-PICA organisation. Organisations who are allowed to use the data are identified by the Dutch Data Protection Authority (Dutch DPA) <http://www.cbpweb.nl/asp/ORDetail.asp?moid=808d858982>

12. *If yes, how is this achieved (what interfaces/protocols are used) and is the system free to access?*

The system is not free to access. Existing users of the GGC can use the NTA and DAI functionality free of additional charges. For research institutions that do not have access to the GGC, a separate 'DAI contract' is available for the use of light version of the WebGGC for a limited fixed fee.

The workflow of the DAI proces is descibed in this document: http://wiki.surffoundation.nl/download/attachments/3473692/2006-06-27_DAI_werkprocessen.pdf (DUTCH)

13. *How is the system funded?*

The NTA is funded through license fees for the use of GGC.
METIS is funded through license fees for the use of METIS.

14. *Is the system still under active development? If so, what are your priorities for future enhancements?*

Currently the system is in 'maintainance mode'. That is: the underlying software (CBS) for the GGC is actively maintained by OCLC and METIS software by UCI. The WebGGC was recently completed by OCLC.

We are currently orienting towards opportunities to open up the system with standard APIs such as SRU and SPARQL but these plans are by no means certain nor final.

Finally, a future development could include integration of author identifier within the federated authentication infrastructure.

15. *Do you have any plans for integrating your system with external initiatives/services such as ORCID, ISNI, Mendeley, Zotero, Academia.edu?*

There are some careful thoughts about the possibility connecting the NTA via GGC and

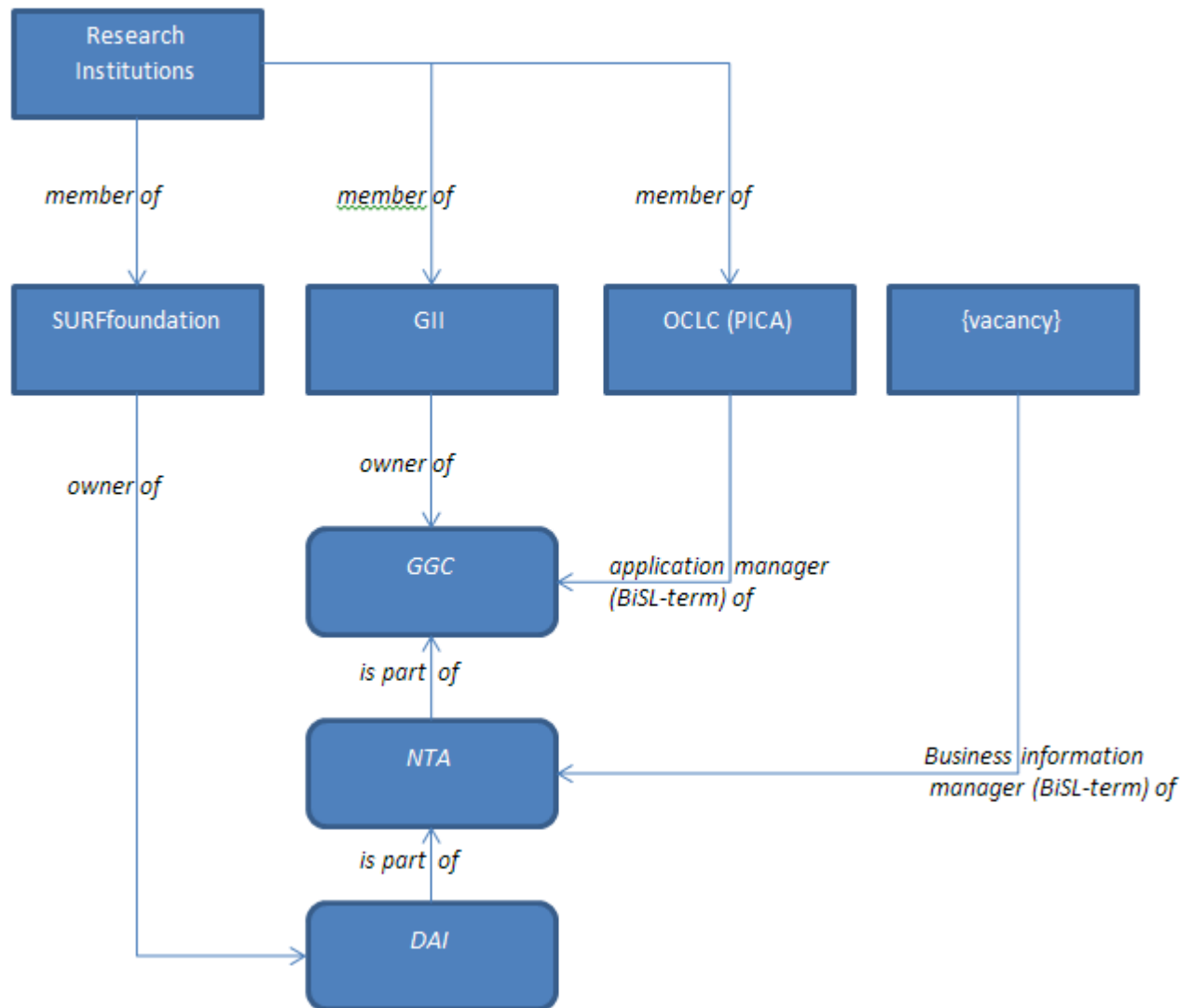
WorldCat to the VIAF. Currently we are awaiting the outcomes of discussions within the Knowledge Exchange initiative (DEFF, DFG, JISC and SURF) on digital author identification, as well as filling in the role of business information management for the NTA before initiating any net development.

Currently the ISNI database is placed at OCLC in Leiden, the Netherlands. This database is populated using the CBS application. Mind: this application is also used to populate the GGC including the NTA. For the initial population of the ISNI, the VIAF database has been used, added with other databases containing composers and performers, etc. Tests are currently performed with the BL (British Library) en BnF (Bibliothèque national de France).

In short, the NTA is in theory migratable with the VIAF and ISNI.

Appendix A - DAI Organisational Structure

DAI – Digital Author Identifier – Organisational Structure



Appendix B - DAI Infrastructure

