

Registering Researchers in Authority Files: Functional Requirements

These functional requirements were derived from the eighteen use case scenarios.

1. Related to reading/queries of multiple authority system

Ability to:

- Successfully link multiple identities a researcher might have to allow for collation of intellectual output.
- Associate metadata for a researcher's intellectual output with the researcher's identifier.
- Create metadata consistent and robust enough to allow for successful integration with other researcher's output and various types of sorting.
- Retrieve all the necessary metadata for the intellectual output related to a specific researcher.
- Dedupe or cluster results from searching multiple systems.
- Disambiguate similar results.
- Merge entities that represent the same researcher and split entities that represent different researchers.
- Determine that a researcher or a work related to a researcher is not already represented.
- Link the metadata for a researcher's intellectual output to grant funding body data.
- Link various name authority files and researcher registries, such as VIAF.
- Query aggregated authority files or researcher registries through the bibliographic utility used for name authority record creation (e.g.OCLC).
- Search and compare databases across institutions from which potential collaborators are sought.

2. Related to creating/updating/administering multiple authority systems

Ability to:

- Associate an identifier with an identity which resolves to information about the researcher.
- Freely use the identifier in other systems.
- Register a researcher who does not yet have a persistent identifier.
- Link a researcher's multiple identifiers.
- Affiliate an individual researcher with multiple departments or institutions and multiple disciplines.
- Track the history of a researcher's various affiliations over time.
- Record and disseminate a researcher's various roles or subject categories.
- Communicate updates from one system to another, and vice-versa.
- Tag each data element with its source/provenance.
- Retain **all** information received, even if it's not all displayed.
- Communicate information between and among different systems.
- Support batch searching.
- Support batch updating.
- Update information in the system or ask for it to be updated by a researcher.

- Track updates and to inform other systems about them.
- Exchange metadata among different identity management systems.
- Record identifiers within the metadata for the resource.
- Clearly identify an author/contributor from the information in the researcher identifier system.
- Use an existing record as a template to start building a name authority record.

Cross Cutting Requirements

- Search and retrieve information on researchers from a publicly-accessible database.
- Support Unicode to record researchers' names in their own language and writing system.

Authentication and authorization requirements

Applies to all use cases involving updates, and for all use cases involving queries across non-public information.

Ability to:

- Allow authentication of actor and authorization of query/update to name authority system.
- Authorize researchers against multiple independent systems, either using federated ID protocol such as OpenID or Shibboleth, or having client systems store and maintain multiple credentials and manage multiple protocols.
- Account for people who want to maintain multiple separate identities without linking those identities, i.e. the ability to “opt out” of how information is gathered and/or displayed

Interoperability Requirements

Applies to all use cases system-to-system queries or system-to-system updates

- Identify protocol for defining query/update, initiating query/update, and interpreting results (both content and error conditions resulting, if any)

Security requirements

Mix of specific functional requirements and system wide properties

- Use encryption for all network connections to service involving updates if warranted.
- Comply with enterprise security standards. (For example, Harvard has a well-constructed and systematic set of requirements at <http://security.harvard.edu>.)

Performance Requirements

System-wide properties

Document:

- The scalability that is required in terms of number of records for the given context.
- The load of queries/updates per institution the system must support.
- The availability of service that is required.
- The level of reliability that is required.

Licensing/legal Requirements

System-wide properties

Document:

- What restrictions on use of data are acceptable/permissions required for different types of uses? Reuse in catalogs? Sharing? E.g. MIT strongly favors open access but has not required it for bib data. Harvard has moved to OA for their bib data and identifiers.
- What restrictions on use of API's/code are acceptable for different types of uses?
- What generic licensing clauses are (un)acceptable? – e.g. MIT typically will not accept licenses that require MIT to indemnify other institutions
- What cost models are (un)acceptable? Per -query? Per institution subscription?