Quality Assurance in the Research Documentation System Frida

Grete Christina Lingjærde, Andora Sjørgren
Centre for Information Technology (USIT), University of Oslo, Norway.

Summary

Frida is an integrated research environment for the documentation and presentation of research activities, research results and scientific competence. Data from Frida is used to generate statistics for research activities at Norwegian universities. Information provided by this system plays a major role in the annual funding of universities by the Norwegian Ministry of Education and Research. Therefore, data quality has been a major issue in the development of the system.

1 System overview

Frida, a web-based application that accesses an Oracle database, has been in use since 2003 and was developed by the Centre for Information Technology at the University of Oslo. It is the result of a collaboration between four universities in Norway (University of Oslo, University of Bergen, The Norwegian University of Science and Technology and University of Tromsr) and is currently used by these universities as well as by some other Norwegian institutions (Oslo University College, The Norwegian Institute of Public Health, and The Norwegian Knowledge Centre for Health Services).

The main purpose of Frida is to facilitate the registration of research activities in order to meet a number of internal and external needs: Cataloging of scientific activities, reporting to the Ministry of Education on scientific performances, distribution of performance-based funding and information-gathering for the development of institutional strategies for research activities, profiling of researchers and lastly, the presentation of the research activities, projects etc to the public and other stakeholders. Frida consists of four modules:

- *Module I (Research results)* provides an overview of scientific activities.
- *Module II (Catalogue of Expertise)* provides an overview of researchers' and research units' scientific competence.
- *Module III (Annual reporting)* provides additional information used for the annual reporting of research activities, such as annual work load, research visits abroad, visitors, awards, prizes, honors, etc.
- *Module IV (Research projects)* provides an overview of ongoing and concluded research projects, including a brief description of each project and lists of associated researchers and all project-related publications.

Frida also facilitates the retrieval of research results from bibliographic data sources. Suppliers of bibliographic data to Frida are the Institute of Scientific Information (ISI), the Nor-

wegian and Nordic index to periodical articles (NORART) and the national library database for academic book publishing (BIBSYS). The Frida project can be visited at http://Frida.usit.uio.no/.

2 Background

Data entered in Frida by researchers provide the foundation for an overview of their own research activities and as well as a basis for annual reporting to the Ministry of Education and Research.

3 A new financial model

The Norwegian documentation system for research funding was approved by the Ministry of Education and Research in 2005 and the model was applied for the first time during budget allocations in 2006, the system is designed to facilitate a performance-based distribution of research funding to institutions based on factors including academic publishing activity.

The Ministry of Education and Research took initiative to improve the quality of publication data when the new financial model was implemented. This resulted in: (1) The creation of a national register of publication channels (periodicals, series, publishers) and institutions (organizations); and (2) An information pool of bibliographic data to be distributed to local research documentation systems. In addition, a system called ITAR (Import Service and Authority Registers) was developed in order to organize information from authoritative registers and bibliographic data. These data are made available to Frida via an export service in ITAR.

4 Data registrations in Frida

Data regarding research activities and projects are mainly registered by the academic staff themselves. Publication data can also be imported via the ITAR system by the academic staff as well as by the administrative staff.

5 Data model and the authoritative registers

The Frida system is developed around a highly structured data model. The system contains registers/separate tables of periodicals, series, publishers, organizations (institutions) and common code tables. Such standardized structures increase the quality of data and prevent inconsistencies in the data, such as as might arise if the same journal name were spelled differently during registration.The Frida institutions share these registers. The common use and maintenance of these registers is an important quality measure in Frida.

6 Institution register

Frida includes an authoritative register of national and international institutions. The register is imported into Frida from ITAR every night. This register plays a significant role in producing statistics regarding collaboration between institutions, both national and interna-

tional. It is also used as a tool in quality assurance procedures and the reporting of data. Reports listing publications co-authored by researchers from different Frida institutions are generated in order to find inconsistencies that may arise during registration locally.

7 Publication channel registers

Frida has developed its own register of periodicals and series in which researchers from Frida institutions publish. Data concerning these channels is imported from ITAR into Frida on a regular basis. Close collaboration among the Frida institutions ensures the maintenance of the register. The publication channel register in Frida is larger than the corresponding register in ITAR, as it contains not only academic periodicals/series but also other publication channels. This is due to the fact that Frida allows academic staff to register a new periodical directly in the system, after which it is approved by the central authority. Again, the use of shared, national registers prevents inconsistencies that might occur if a journal were to be identified differently in different information systems.

8 Data from external bibliographical data sources

Frida facilitates the import of external bibliographic data from sources such as ISI, NORART and BIBSYS through ITAR. An import component has been developed in the Frida-application which allows academic staff to import their own publications as well as allowing administrative staff to import all publications for their institution.

The import component in Frida has been designed to handle the different statuses a publication may have:

- The import publication has already been manually registered
- The import publication has already been imported but lacks additional data
- The import publication is new (has not been previously registered in Frida)

During the import phase, a selection of ITAR-data is defined as authoritative and will override manually registered data. This is particularly relevant for data later submitted when applying for funding from the Ministry of Education and Research, including publication channel, the number of authors and the publication type (article, letter etc.). These data can not be changed by the user. Other data such as title and volume can be changed.

9 User administration system

For a user to be able to register data in Frida, a personal record must be imported into Frida from the institution's user administrative system. One example is the BAS/Cerebrum system developed at the University of Oslo and used by several other institutions. Data in such systems are based on data from the institution's personnel system. In other words, a user must be employed at the institution in order to register data in Frida.

10 Personal data register

Each Frida institution has its own personal data register. All personnel employed by the institution are imported into Frida from the institution's local personnel system (such as SAP) via the user administrative system. Each person employed by the institution or associated with the institution is uniquely identified by a social security number. This has presented a challenge for all institutions as guests/associated persons such as visiting researchers, professors emeritus, etc, are not always registered in the local personnel system. Some of the institutions have resolved this problem by registering non-employees in the personnel system as guests, after which their data are imported into Frida. Using the local personnel system as an authoritative source increases the quality of data as data are registered and maintained in one place only.

11 Frida and adjoining systems

As described above, Frida is connected with many other systems (see Figure 1). The exchange format between the systems is in XML and is based on predefined common XML schemas.

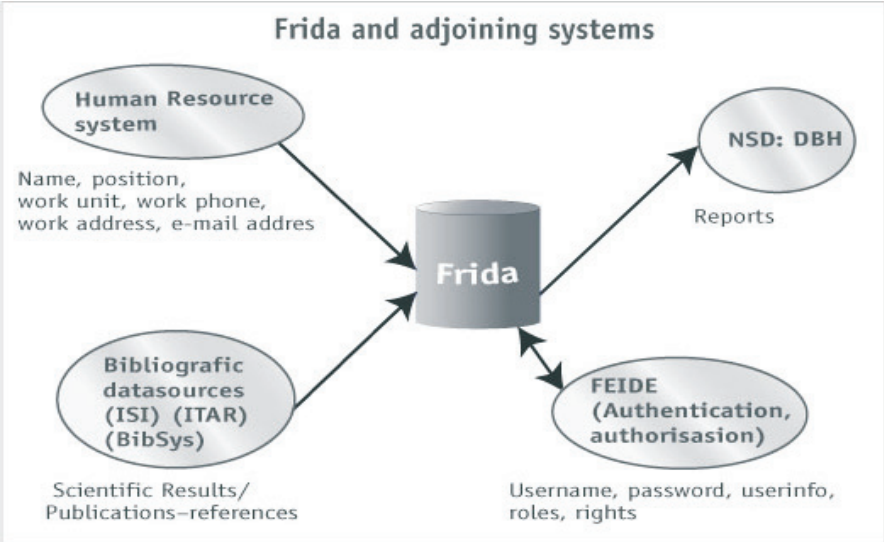

Figure 1. Frida and adjoining systems

12 Control of data in the database

Much of the information in Frida is subject to automatic validation in the database. Tables with code values are implemented in many areas. For example, there is a central code table for publication type. All legal values that the user may register in a field must be represented

in the corresponding code tables. In addition, various business rules are implemented in the database. These rules prevent the direct alteration of data in the database (or via the web interface) that might introduce inconsistencies or otherwise lead to poor quality. For example, the user cannot register a journal that does not have an ISSN. Furthermore, only authors registered with a social security number will be recognized by the system as affiliated with the institution (confirmation of author address).

13 Duplicate control

Standard duplicate controls used in bibliographic data systems have been implemented in Frida. Due to the strict and normalized database structure, other approaches and controls have also been implemented. The common use of authoritative registers makes duplicate control functions easier to implement.

14 Quality reports

Several reports have been developed as part of the quality assurance process. As some publications include authors from different Frida institutions, these publications are often registered in several institutional instances of Frida. In such cases, the control reports ensure data quality across the different institutional databases. One such report controls that the sum of publication points for a single publication authored by researchers from different institutions does not exceed the publication's total point value. The reports that handle data from several institutions are very simple to develop in Frida, since all institutions have their data in a common physical database based on the VPD (Virtual Private Database) functionality in Oracle. In addition, Frida generates reports based on data collected at different levels (institutional, faculty or department/institute levels) of the organization.

15 Approval regime

All publications that are candidates for reporting have to be approved by a super user or other administrative staff before they are reported to the Ministry of Education and Research. These types of checks and approvals improve the data quality as they ensure that only data that meets the specifications for performance-based funding is reported. Several functionalities are implemented in Frida in connection with the final approval phase.

16 Quality routines

The Frida Secretariat has provided a common routine description to its member institutions. However, as it is important that the institutions themselves are responsible for their own routines, this description represents suggested guidelines. It describes all of the phases from data entry, data control and approval to the annual reporting of data. This document provides a basis for an institution's own description of routines and procedures after they have made minor adjustments in order to adapt to local conditions. The different institutions could also have different responsibility charts.

17 Open information system

Data in Frida are made available through an open web interface. This is an important quality measure, because publication on the web creates an incentive for researchers to correct errors involving their own data. There has also been an increased interest in the presentation and profiling features of Frida. Recently, functionality of Frida was enhanced to allow researchers and project participants to edit presentation layouts in order to add more information and to add pictures. In the future, Frida may function as the Yellow Pages to scientific activities at an institution. All researchers using Frida are registered in the catalogue and can be retrieved. The user can add search for researchers, projects, publications, etc. Frida also uses the Norwegian Science Index for searching, indexing and statistical purposes.

18 Reuse of Frida data

We allow reuse of Frida data in different contexts, e.g. by allowing dynamic links to be added from local web pages to Frida, thus encouraging researchers to register more and to care about the quality of the data. It is also possible to save the results of a publication query to an XML tagged file. Frida has created its own XML scheme for Academic activities (where all publications are registered). These data can be used in e.g. CVs or funding applications.

An important issue, both for the general acceptance of the system and for the prevention of inconsistencies in the registered data, is the need to safeguard against the registration of the same information multiple times. There is also an increasing demand for the use of data from Frida in other contexts, both locally and publicly. The number of stakeholders are increasing, such as students, the researchers themselves, research managers, administrators, journalists, sponsors, foundations, companies and the general public.

Frida has developed a submit service which allows the submission of journal papers and other contents to Open Access Systems. The metadata in Frida are reused in Open Access systems. At present, only the University of Oslo makes use of this facility, but the service is implemented in a manner that will allow data to be delivered to any OAI compliant Open Access system. It is also important that researchers are not required to register in many different systems. For researchers to report data to their institution, they should only have to relate to one system: Frida.

19 Data registration and alternation of data

Only authorized users can log on and enter or change data in the system. Only authors of a publication can alter the data concerning this publication. Super users within the same line of organization can also make changes to the data. For example, the super user at the Department of Mathematics may perform alterations to publication records involving authors working at this particular department. It is also possible for the institution to activate an email-messaging system that notifies all the authors involved whenever a change has been made to a publication record or a new publication is being registered.

20 Ethical declaration

Functionality concerning ethical issues related to both publications and projects has also been implemented in Frida. This functionality is currently being tested in a pilot project at the University of Oslo, where the idea was initiated and the functionality was specified by. Using this functionality, researchers may choose to confirm that they vow to follow discipline-specific ethical guidelines (such as the Helsinki declaration). They may also choose to declare that a publication adheres to the Vancouver guidelines for co-authorship. In addition, researchers must confirm that they have acquired the required approvals and contracts for a project. The goal is to increase general awareness regarding sound scientific practices and the correct use of and accurate registration of data. So far researchers appear to be eager to register compliance with ethical standards and the Vancouver guidelines, and it seems that Frida may function as a communication tool with and for the researcher.

21 Conclusion

The quality of Frida is always on the agenda, and the system is constantly being improved. The total number of publication points reported to the Ministry of Education and Research over the last two years has stabilized. We interpret this as an indication that the level of quality is acceptable. There is great interest from researchers to register data. Researchers are only required to register data that is reported to the Ministry, but most researchers register data about all their activities. Researchers demand that the data registered in Frida can be used for profiling them as researchers. Introducing VPD resulted in an opportunity to develop national control routines across all participating Frida institutions. Frida is an excellent starting point for a national system for registering research activities, especially since the data belonging to large research institutions is already represented in an identical structure in a common VPD-database. It takes time to introduce a new system to an organization.

Contact Information

Grete Christina Lingjærde
Usit /Center for Information technology
University of Oslo
Forskningsveien 3B,
0313 Oslo
Norway
E-mail: g.c.lingjarde@usit.uio.no