



13th International Conference on Current Research Information Systems, CRIS2016, 9-11 June 2016, Scotland, UK

Even minor integrations can deliver great value – A case study

Mona Hide Klausen^{a,*}

^a*CRISStin, Fridtjof Nansens vei 17/19, 0369 Oslo, NORWAY*

Abstract

When CRISStin, Current research information system in Norway, was established the system mainly served as a tool for reporting academic publications to the ministries. The implementation of a new API service has facilitated integrations with other systems, and is now changing the CRISStin system into the research information system it was originally meant to be. A recent integration with another national system demonstrates a CRIS system's potential for not only increasing visibility of research, but to substantially reduce research administrative costs.

© 2016 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the Organizing Committee of CRIS2016.

Keywords: system integration; research administration; interoperability; standardisation; integration benefits

1. Introduction

CRISStin (Current research information system in Norway) was established in 2011 under the ownership of the Ministry of Education and Research in cooperation with the Ministry of Health and Care services. CRISStin refers to both an organisation and the Norwegian national research information system. The organisation's main task is to support and promote Norwegian research on a national and international level. The CRISStin system plays an important part in reaching the organisation's goals. For the remainder of this paper, the name "CRISStin" will refer to the system.

* Corresponding author. Tel.: +47 228 45 611

E-mail address: m.h.klausen@cristin.no

The main purpose of CRISStin at the time it was established was to give Norwegian research institutions a single tool for reporting their scientific publications to the Ministries. In Norway, most research institutions receive part of their funding based on their academic publication activity. Before 2011, the institutions used a variety of systems at the institutional level to manage the required annual reporting.

CRISStin's 158 member institutions come from three different research sectors: higher education institutions, research hospitals and independent research institutes. They register information about their employees, and their research, projects, presentations and art work in CRISStin's central database (see Appendix A. for a list of the contents).

CRISStin was always meant to be more than just a tool for reporting publications, but limitations in technology prevented the system from reaching its full potential. The implementation of a new service for sharing data in CRISStin with other systems is now changing CRISStin into the research information system it was supposed to be.

In this paper the potential benefits of making data in CRISStin interoperable and more easily accessible will be described. Section 2 describes CRISStin's transformation from primarily being a reporting tool to becoming a research information hub, and the benefits this has resulted in. Section 3 briefly describes the potentials of an interoperable CRIS system. In section 4 a national project for making Norwegian research information interoperable will be introduced. Finally, section 5 outlines the steps believed necessary for CRISStin to become part of a larger, international research information network.

2. Becoming a research information hub

One of CRISStin's aims has always been to reduce the time researchers and research administrators spend on manual, administrative tasks. In order to do that, information in CRISStin would have to be made available for reuse in other systems. A web service has been in place since the beginning. This has mainly been used by the institutions to retrieve information about their employees and their research results, in order to present it on the institutions' web site. It has reduced the institutions' need for manually updating information, but it was clear that to have an impact on the daily lives of researchers and research administrators, data in CRISStin would have to be reused in other systems.

In 2013 CRISStin integrated with the project reporting system of the Research Council of Norway (RCN). When reporting the results of RCN-funded projects, project leaders may now harvest data about the publications from CRISStin, rather than entering the information manually.

Additional integrations with other systems have been hindered by major limitations in the scalability of the old web service. Larger institutions cannot retrieve all the data they are interested in, or as often as they would like, without the service shutting down. It also puts a massive strain on the entire system. Thus, in April this year we launched our new service for retrieving data about projects, persons and institutions based on REST API. This service is now being used in a new integration with another national system.

The technical aspect of the new REST API is outside the scope of this paper[†], which will primarily focus on the impact of making data in CRISStin more easily accessible, in terms of reducing administrative costs.

2.1. Ethically approved projects in CRISStin

This spring CRISStin integrated with the project application system of the Regional Committees for Medical and Health Research Ethics, called SPREK. Research projects involving the study of humans or human biological material have to apply for ethical approval before start up. The approved projects are automatically transferred to and registered in CRISStin.

[†] A technical description of CRISStin's new REST API service can be found in Vallipuram, Y. & Sachse, D. (2016): "Managing and sharing research data using RESTful web services", presented on the CRIS2016 conference.

The main goal of the integration has been to provide the institutions with an overview of their clinical trials, which they are legally required to have. Gathering this information in CRIStin means the institutions no longer need to maintain their own local systems, or using other manual tools, for this purpose.

SPREK, where the project applications are registered, is basically a document repository. In addition to storing project information in pdf-files, SPREK uses a lot of free text fields. The main purpose of the system is to create documents for use in meetings where the Regional Committee members evaluate the ethical aspects of the projects. This makes it ill-suited for retrieving data.

In order to enable integration with the CRIStin system, which contains highly structured data, SPREK had to be substantially altered. This included structuring key information, which was solved by SPREK implementing CRIStin's person, project and institution ID's. When registering a project application in SPREK, they use the new REST API to get updated information about persons and institutions from CRIStin. The ID's are then registered in SPREK. When the project is transferred to CRIStin, it receives a unique project ID upon registration which is returned to SPREK, and stored on the case file for the project application (see left side of Figure 1).

The implementation of the CRIStin project ID in SPREK was necessary to allow for transfer of updated information about a project, without creating duplicates in CRIStin. The person and institution ID's enable CRIStin to link the project to the identified researchers' profiles and the participating institutions' project portfolios. An added bonus of the integration is better data quality in SPREK.

SPREK contains project information only. By transferring project data to a CRIS system like CRIStin, this information can be connected to other types of data. The project leaders can register any activity or result related to the project, such as publications, conference presentations, interviews, newspaper articles, or commercial products. Related projects can be linked together and also presented on a research group's profile. This creates a much richer dataset, and enhances visibility and discoverability of the project.

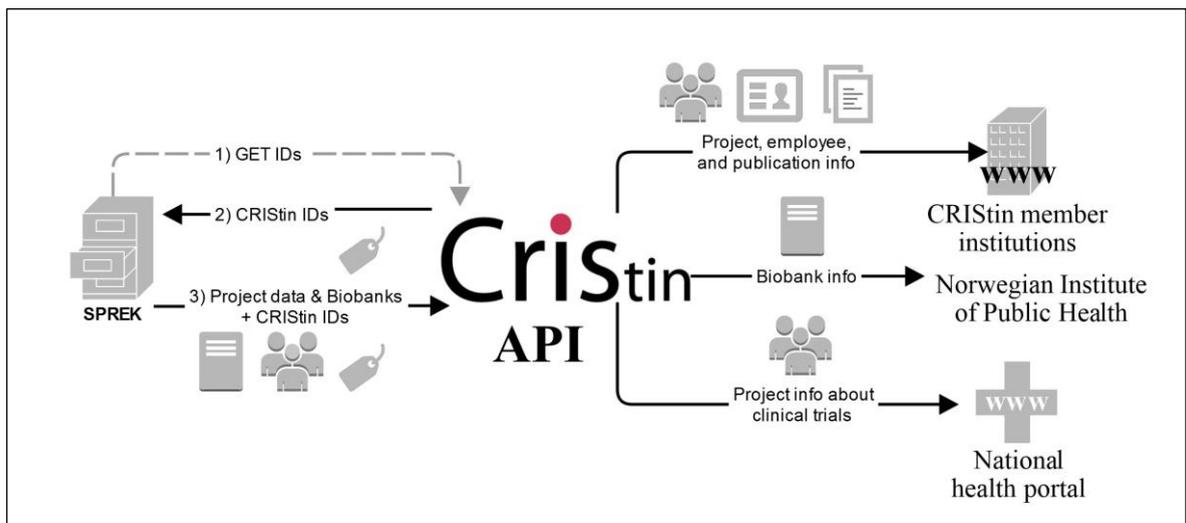


Figure 1 The data flow of the integration between CRIStin and SPREK, and recipients of the data made available through the REST API. (Icons from www.glify.com)

2.2. Added benefits

In addition to giving the institutions an overview of clinical trials in CRISStin, they will now be able to harvest their ethically approved projects from CRISStin to their web site, without having to register the information a second time. Gathering information about all of their projects in one system will also provide a more comprehensive overview of their project portfolio.

Shortly after we started talking to SPREK about integration of our systems, other national services expressed an interest in making use of the data we would be receiving from SPREK. Data about clinical trials are also of interest to the national web portal for providing the Norwegian public with information about health related issues, www.helseportalen.no (right hand side of Figure 1). The portal is used to inform patients about on-going clinical trials they might want to participate in. Today the hospitals have to manually register these projects online. The API will enable them to harvest relevant information from CRISStin and automatically present it in the web portal.

SPREK is also used for applications to establish biobanks, either as a part of a specific project or as a separate biobank. This information is transferred to CRISStin as part of the project profile. The biobank information is now made available to the Norwegian Institute of Public Health (NIPH), in charge of the national biobank registry (right hand side of Figure 1). Since the implementation of SPREK in 2009 the NIPH has not been able to properly update the biobank registry, due to the aforementioned difficulty in retrieving and collating information from SPREK. Rather than developing a comprehensive API service of their own, to harvest the data from SPREK directly, they will use CRISStin's new REST API to obtain the information they need.

The CRISStin API was deliberately designed to be generic, rather than being tailored to the SPREK integration. Anyone interested in information about registered projects, persons or institutions may use the same API services to extract the data they need. We are currently in a dialog with several other organisations interested in data from CRISStin, and we are looking into new integrations to further reduce the need for manual, multiple registrations of the same information.

2.3. Estimated time savings

We have not carried out a formalized study to measure the actual benefits of the integration in terms of time and money saved, but we may illustrate with an example. Without the integration and the API service, the researchers or administrators would have to register the relevant project information in a local system, or in a spread sheet, for the institutions to get the legally required overview of the clinical trials. If three institutions collaborated on the project, the same information would have to be registered in three separate local systems. Additionally, someone would have to type relevant information on the national health portal, in order to inform patients about the study. If the institutions wanted to present the project on their local web site, the same data would have to be registered online, manually. To get the complete overview of all of the institutions' research, the project would also have to be registered in CRISStin.

The information we receive from SPREK would take just shy of 10 minutes on average to register in CRISStin. If we anticipate about the same amount of time for the other systems and web sites, our example would require approximately one work hour in order to get the information into all the relevant systems and online. We anticipate about 800 projects from SPREK each year, which is a fairly low number. Even so, the time saved on administrative tasks may be at least 800 work hours in total, nationally, as a conservative estimate. In Norway, a work year is considered to be about 1700 hours.

This is estimated time saved on the registration process alone. It does not take into account the time administrators spend on finding the information that needs to be registered, or on the development and maintenance of the local systems. When the information is shared with even more systems, the benefits will increase.

2.4. Challenges in the integration process

The process of integrating the two systems has been a long and arduous task. The main challenge was changing SPREK from a predominantly free text based archive system, to a system one might extract data from. Implementing structured data for key information was vital to enable CRISStin to understand the data it would receive.

Another challenge was combining the different development methodologies used in the two development teams. CRISStin's developers use agile methodology; the SPREK developers used a type of "water fall" method. When integrating two systems, continuous testing of "the goodness of fit" between the systems is crucial. Waiting long periods of time to get a version of a system to test, and then waiting another lengthy time period for any errors to be fixed, causes unnecessary delays. Eventually the SPREK-developers started to work in a more agile fashion, to collaborate more efficiently.

Securing data integrity between the systems has also been a topic of much discussion. If data were edited in CRISStin, but not in SPREK, the data integrity would suffer. To make sure data would remain identical in both systems, it was decided that key information had to be edited in SPREK and then transferred to CRISStin as an update of the already registered project. These fields are locked for editing in CRISStin.

Choosing the authoritative source of information has not been a major challenge in this integration. When we start to integrate with other systems containing the same type of information as SPREK, it will be necessary to create a national, authoritative chart of the data flow between the various systems.

3. The potentials of an interoperable CRIS system

The recent integration with SPREK has not only given CRISStin access to quality assured information of clinical trials – it demonstrates an interoperable CRIS system's potential for reducing administrative workload when integrating with other systems:

- Institutions do not have to develop and maintain their own systems to maintain an overview of the relevant data.
- Researchers and administrators will not have to register the same information in multiple systems.
- Information can be harvested for presentation online, removing the need for manual registration.
- Information about research infrastructures used in a project can be shared with relevant registries, reducing the need for manual updates.
- The CRIS's ability to link various types of research information will give research institutions, and relevant Ministries, better statistical data as foundation for political and strategic decisions.

4. Structuring data for national systems integrations

Even in a small country like Norway, a minor integration may generate substantial savings. Representatives from organisations delivering IT-systems and services to the Norwegian higher education sector (within the areas of research, education and admission) have thus begun to work towards structuring key information, which may be shared between two or more systems. Typical examples are persons in different roles (researchers, students, authors) and institutions. National or international standards, e.g. the national Register of Business Enterprises and ORCID, will be used if they exist. Otherwise, one of the national systems will be chosen as the authority. Agreement to use the same codes in all systems will greatly facilitate integrations, and also open up for the generation of new types of statistics.

5. Becoming a part of the international research network

CRISStin has been the central storage for Norwegian research data since the implementation in 2011. The recent integrations are another step towards becoming the research information system it was meant to be. By retrieving, collating and redistributing information, the system is able to reduce administrative costs, in all instances that are

able to make use of the information we are making openly accessible. The redistribution of information further increases the discoverability of research data.

We have started to look into the CASRAI dictionary, in order to identify the necessary steps to making our services ready for integration with international systems. In the not too distant future, there will also be a need to implement a more standardised format for data sharing. Making our schemas CERIF compatible will be a necessary first step. At the time being, we are using an in house developed solution for the REST API.

The more we are able to make our data interoperable, the more time we can save researchers and institutions from having to spend time and money on bureaucratic tasks – hopefully giving them more time to do research.

Perhaps the most important effect of the recent integration, from the CRISin organisation's point of view, is the positive reactions we get from researchers, who are now starting to realize what the system may be able to offer them, in terms of reduced administrative workload, increased visibility of their research and a comprehensive overview of their work.

Appendix A. Contents of the CRISTin database

A.1. Core entities

Entities	Number
Persons, uniquely identified with a Norwegian national security number	143 500
Projects	11 000
Results	1.2 million
- Quality assured academic publications (since 2011)	107 000
Research groups	800
System administrators/ super users	700

A.2. Sources of structured data

To minimize the need for manual registrations, we import structured data from the following sources:

- Publication data from Scopus (Elsevier) and the National Library of Norway. About 10 000 publications, half of what is reported to the Ministries, is imported and checked by CRISTin each year.
- The Norwegian Register for Scientific Journals, Series and Publishers (owned by The Norwegian Centre for Research Data), which contains 27 000 quality assessed journals and publishers.
- Information about the employees of our member institutions from the institutions' personnel system. At the moment 15 of our member institutions use this automatic employee update service.
- An institution registry with the names of 21 500 other Norwegian and international institutions from FSAT – a national service center that develops student administration and admission systems for higher education institutions.