

CRIS 2014

Deceiving simplicity. Balancing the need for ready-to-use research information with the semantic and technical complexity of research data

Hannelore Vanhaverbeke^{a*}, Sabina Beullens^b, Lieve Timmermans^b, Klara Gijsbers^a,
Bieke Bras^b and Christelle Maeyaert^a

^aKU Leuven, Research Coordination Office, Schapenstraat 34, 3000 Leuven, Belgium

^bKU Leuven, Directorate ICTS, Competence Centre Management Information, Willem de Croylaan 54, 3001 Leuven, Belgium

Abstract

KU Leuven, Belgium's oldest university and member of the League of European Research Universities (LERU) generates a myriad of research-related data, managed in different technical environments. Faced with an increased demand for validated key performance indicators to support policy planning at the level of faculties and departments, the university's Research Coordination Office (RCO) set up an institutional central research information system (CRIS) in SAP Business Intelligence (BI) in close collaboration with the ICT office. The road to providing ready-to-use KPIs in such a way that users were not confronted with the underlying semantic and technical complexity of the data proved to be a long and winding one.

© 2014 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of euroCRIS.

Keywords: institutional CRIS; SAP; WebI; semantics; research policy; key performance indicators; KU Leuven

1. Introduction – The setting

Founded in 1425 KU Leuven is Belgium's oldest comprehensive university and home to more than 7.000 researchers (Full Time Equivalents or FTE), one third of which are of foreign nationality¹. A charter member of LERU, the League of European Research Universities, the university hosts its own technology transfer office,

* Corresponding author. Tel.: +32-16-324061

E-mail address: hannelore.vanhaverbeke@doc.kuleuven.be

Leuven Research & Development (LRD)². Currently KU Leuven ranks 6th in total FP7 funds awarded (cf. 5th interim FP7 report), highlighting its strong tradition in research. This high research intensity generates a myriad of research-related data. They range from macro-scale data such as the overall amount and allocation of research funding, or the number of people involved in research, to the micro-scale of individual scholarships and publications.

The registration and systematic collection of these information sources has been prone to different chronological and organizational trajectories. In 1998 a major project of administrative process mainstreaming was undertaken by the ICT office³. Instead of relying on mainframe systems, which were considered not well integrated, not sufficiently flexible or user friendly and not accessible to all stakeholders, six core domains were defined within the abundance of administrative actions: research, education, student administration, finance, human resources and logistics. All master data and processes within each domain were subsequently migrated to one software environment, the ERP package SAP (both ECC and CRM) from 1999 onwards. In 2003 nearly all central services were hosted by the same software platform. Research data in SAP were set up according to the CERIF model: project, person and organizational unit are defined as objects with relationships between the different objects. The relationships are defined in time and the relation type defines the role between the objects. Data from SAP were then pushed to SAP Business Warehouse (BW) - originally SAP BW 3.0, now SAP BW 7.0 Ehp2- for analysis and reporting by domain experts in the BI-tool Business Explorer Analyzer (BEx Analyzer). Simultaneously an administrative portal was developed offering a personalized access for KU Leuven staff to information from the six aforementioned domains through a web-browser with limited ‘self-service’ facilities (the so-called ‘KU Loket’). Upon logging into ‘KU Loket’ research related information is now presented to individual staff members (and their delegates). Depending on authorization rights, different applications within each domain are accessible (Table 1). The tool is an ongoing project with updates and additions being monitored by ICT services.

Table 1. Access to research-related data in KU Loket, dd. 04/2014

Info type	Staff	PhD Projects	Research Projects	Funding	Publications & Citations	Patents
Domain	HR	Student Administration & Research	Research	Research (internal funds) or Financial (external funds)	Research	Research
Managed by	Centrally by HR	Locally by faculties	Centrally by Research Coordination Office	Centrally by Financial Services	Individual Researchers	Centrally by Technology Transfer Office
Visible to	Heads of all units	Supervisors PhD students	Supervisors	Supervisors	All staff with publications	Inventor

Licenses and patents are one of the core tasks of KU Leuven’s Technology Transfer Office, Leuven Research & Development (LRD). LRD facilitates research collaboration with industry, manages intellectual property rights (patents and licenses) and supports the creation of spin-off companies. Although data relating to these fields are kept in an LRD-specific environment, information on patents and licenses has been included in KU Loket.

Since 2006 publications by KU Leuven researchers are stored in a DSpace-based institutional Open Access repository – the Leuven Institutional Repository and Information Archiving System (Lirias), hosting both metadata, full texts and datasets⁴. Metadata are publicly visible while the accessibility of any files that accompany the metadata is determined by the author. Although submission in the repository is not mandatory, many incentives are in place for researchers to submit timely and accurately: all data on publication behaviour accompanying peer reviews in promotion files, in applying for internal research funding, in internally allocating research money or in creating reports for external evaluation panels are based on the information stored in Lirias. Files uploaded in Lirias and publicly accessible are highly visible on the net, illustrated by Lirias’ 22nd place worldwide in institutional repositories⁵. For that reason many researchers are using the repository to enhance the visibility of their research. A direct link in KU Loket to the publications in Lirias is provided to enable researchers to keep a close eye on their registered output.

The information from these different strands is also coherently presented on the who-is-who pages of the university where, apart from contact data and affiliation information, staff functions and a CV are displayed. The latter offers an overview of the research topics, projects and expertise of the researcher, personal links, as well as an overview of publications, teaching assignments and council memberships. This information is either submitted by the researcher through KU Loket or obtained from SAP or the institutional repository.

The factor linking all these information types is each staff member's institutional identifier or staff ID. This is a unique and persistent identifier, used for all administrative processes, attributed to personnel by the HR department.

2. The need for a CRIS

By the end of the process outlined above, each staff member had his or her own portal, KU Loket, through which the consultation or management of research related data was channeled. KU Loket proved to be an adept answer to the needs of individual researchers. However, two external factors prompted the need for a new take on the presentation of research related data.

The first determinant was technical in nature. The SAP-support of BEx Analyzer is set to run out in 2016, necessitating the implementation of a new platform. ICT opted for SAP Business Objects (BO) that would enable the redaction of more or less standardized reports, including graphic representations, and different user levels: light (consulting), medium (minor adaptations) and expert (report building). To evaluate the possibilities of the new tool, a pilot project was required with which ICT could work.

The second element giving leverage to a new project was driven by the need for more integrated research related data on the occasion of a series of research evaluations, involving all research units of KU Leuven, organized by the Research Coordination Office (RCO). External reviewers were invited to Leuven to spend some time with the staff of the units to peer review the quality of research offered. To that purpose, stakeholders such as deans, heads of department and vice-deans for research made an appeal to the RCO for supportive data since most of the required information could not be provided automatically in an aggregated way to policy makers. Questions regarding data on research were until then handled by the RCO who either brought together the data from the different platforms (SAP, institutional repository – through BW and BEx Analyzer) or delegated the question to the relevant departments (e.g. Finances, HR).

Combining the need for a pilot project to evaluate the new technical platform BO with an urgent question for a new data platform gave rise to the start-up of the RCO and ICT borne BO project 'Research Statistics at KU Leuven' in 2012.

The demands of the data offered in this CRIS-platform were that they should be

- Considered relevant Key Performance Indicators by the stakeholders
- Validated by the office generating the data
- Uniformly calculated for all research units
- Uniformly presented for all research units
- Updated regularly
- Accessible by authorization only
- Presented in one platform
- Ready to use: clear presentation, intuitive interface, exportable reports

3. Setting up the CRIS

A consensus had to be reached as to the selection of relevant key performance indicators (KPIs) for research, the exact definition of the variables, and the way in which the data were presented both for an internal SWOT analysis

and a panel of external reviewers.

The first task was for the RCO to define the KPIs that should be provided in the new application. Feedback from the research units after the internal research evaluations turned out to be consistent and focused preliminary on five indicators: staff (as FTE), PhD projects, research projects, funding and publication output. However, the exact definition of these indicators proved to be deceptively difficult. Staff was defined as embracing five categories: professors (Assistant, Associated, Professors, Full Professors), research managers and experts, postdoctoral researchers, administrative and technical staff, and scientific collaborators (not involved in a PhD) and PhDs students. The latter group had to be presented as one category in spite of what was desired since the data with HR do not allow distinguishing between both groups. In this instance the basic data structure dictated the way in which the KPI could be presented.

In other instances, the way in which the KPIs were defined had to be adapted in the face business reality dynamics or upon request by the stakeholders. The allocation of KPIs to research units proved to be a challenge. What seemed to be a straightforward undertaking –assigning a branch of the university’s organizational chart to each staff identifier, the thread interlinking all KPIs- was fraught with complications. Changes in the chart are frequent and range from major reorganizations to the splitting or merging of units. Added to this are affiliation switches of individual researchers, although these do not occur frequently. Finally, the problem arose as to how to assign KPIs related to staff no longer employed at KU Leuven. The option of resolving these complexities manually was swiftly set aside for reasons of time management. Instead an automatic assignation was preferred controlled by a number of rules. For all staff currently employed at the university, the affiliation was set to the current situation, irrespective of individual historical trajectories. Former employees were assigned to the last research unit at which they were active. If that unit was no longer in existence, because of administrative reorganizations, they were assigned to a research unit of a hierarchically higher order, a level at which changes are exceptional (Fig. 1).

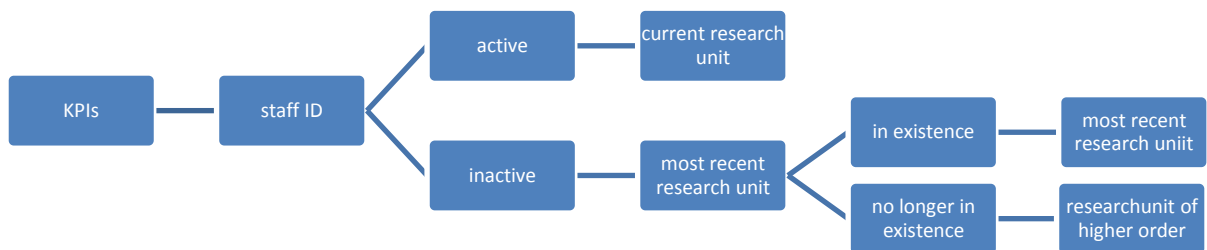


Fig. 1. Schematic representation of the process of assigning KPIs to a research unit.

When asked for feedback, stakeholders expressed the wish to not only see the publication output by senior staff (professors) as was first proposed, but to have an overview of all output generated by a research unit. Especially in the field of Social Sciences and Humanities (post)doctoral students publish without the co-authorship of their supervisors, in which case limiting the overview to senior staff co-authored publications only would not be a correct representation of the actual situation. The selection of staff identifiers for this KPI was thus broadened to include also postdoctoral researchers and PhD students.

Lastly, some policy decisions were made on the way KPIs were calculated and presented. In the case of research funding, for instance, the expenses budgeted on the grants that were obtained were listed, which is considered a more accurate proxy for research investment than the mere opening of the credit line.

Eventually a consensus was reached on the selection of KPIs, on the way they were calculated and the timing of releases and updates, all of which were identical for all research units. It was also decided that the project would start with providing data for the past five years only. The data were released after validation by the service managing the data: in the case of staff the HR department, for PhD projects the faculties, for other research projects the RCO, for funding the financial department and for publications the individual submitters to the institutional repository.

Once a year a new dataset covering one year is released, the timing of which depends on the internal organization of these services. The fiscal year for instance closes around April, after which the validated dataset for the preceding year is added. For some KPIs updates were frequent, others were updated once a year, at the time when a new dataset was released (Table 2).

Table 2. Definition, calculation, release and update timing of the KPIs offered in the CRIS ‘Research Statistics at KU Leuven’, dd. 04/2014

	Staff	PhD Projects	Research Projects	Funding	Publications & Citations
KPIs (reports)	professors, postdocs, PhD students & research collaborators, admin. & techn. staff	nr. of finished PhDs & nr. of current PhD projects	* nr. of intern. & extern. financed projects * nr. of extern. financed bursaries	intern. & extern. acquired funding	* papers in journals & proceedings * books & book chapters * citations * Impact factors (IF) * top journal publications
Calculation	FTE	nr of projects	nr of projects or bursaries	expenses on research grants	*nr. of publications *nr. of citations * IF: sum, max. & avg. IF in publication year & recent IF *nr. of papers in top journal (as defined by the Flemish Centre for Research & Development Monitoring) annually (January)
Release dataset	monthly	annually (January)	annually * January * May	annually (May)	
Update	weekly	weekly	* annually * weekly	weekly	weekly

Subsequently, the ICT team started to define what (further) information would need to be channeled from SAP or the institutional repository to BW for building the cubes. Recurrence could be made to existing cubes, for use by the HR or financial department, other cubes needed to be set up or rethought. With regard to the existing cubes, a match was made between the semantic definition of the KPIs and the corresponding fields in the cube. This process involved repeated consultation between information experts of the RCO and other departments and the ICT experts. As for the new cubes the trajectory from start to implantation was much longer, beginning with a listing of information fields available in the original database, for instance in the institutional repository, explicitly outlining the semantic background of these fields, and then selecting those fields that would constitute the building blocks of, in this case, a new ‘publication & citation’ cube in BW. Once all necessary basis data were available in BW, queries were built in BO to process the data in the required way – for instance, calculating the maximal impact factor per year. The results were compared to what was expected on the basis of either the BEx Analyzer tool, or manual analysis in the case of information derived from the institutional repository. Repeated rounds of testing and validating were needed to achieve the desired outcome (Fig. 2).

The tool was intended as an aid in policy making, supporting those in charge of setting down policy guidelines, and in following the outcome of these guidelines by using supportive data. A group of authorized users had to be defined. In short, all staff in charge of research had to have access: the rector of KU Leuven, the vice-rector for Research Policy, research coordinators, deans, vice-deans for research, heads of department and coordinators of research units. They could be identified in SAP since their role is registered. Together with their delegates, the group of users amounted to some 200 persons. The access to the tool was arranged in such a way that users could only see the KPIs for the research unit that they are in charge of. The explicit aim of the tool is not to benchmark between research units but to sustain policy plans per research unit, hence the limited view. In this way the head of a department can only consult data of his or her department and the underlying units, but cannot see the data of another department. The rector and vice-rector for Research Policy of the university, and all administrators involved at RCO or ICT, can see the data for the whole university, broken down by department or faculty and underlying units.

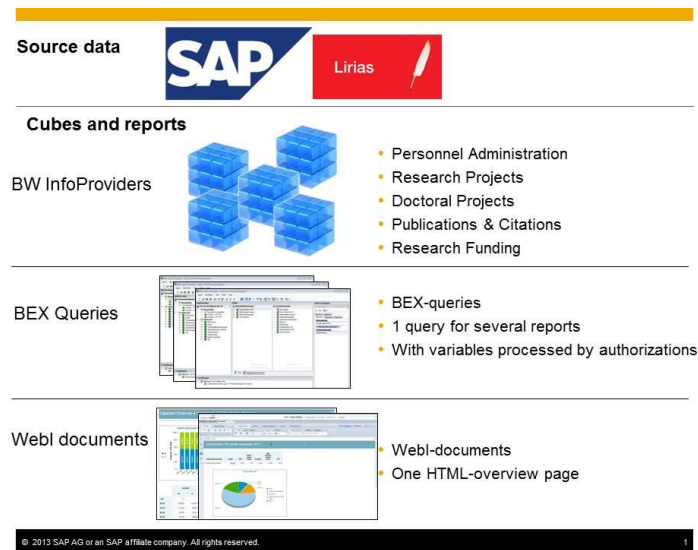


Fig. 2. Workflow preceding the launch of the BO application 'Research Statistics at KU Leuven'.

The procedural workflow being clear, the ICT team set to build the environment in which the KPIs reports could be consulted and downloaded. One of the basic requirements was that all the information should be accessible through one portal. To that purpose, the existing KU Loket application was adapted to provide a gateway to the BO reporting tool. For authorized end users this meant literally clicking a button and opening the application. The underlying technical complexities remain invisible.

The start page of this gateway provides the users with an overview of the available reports, arranged for each of the five information types. Most of these contain one report, 'Funding' and 'Publications & Citations' offer respectively two and five (cf. Table 2). Every report has a direct link to the manual explaining the source of the data, the way these are calculated and the timing of new releases and updates for the data set concerned. Clicking on the required report opens the tool and presents the user with statistics for the report chosen and for those research units for which he or she is an authorized user.

Each report consists of an overview of the data arranged per organizational unit – output being allocated to the unit as explained above. Most reports equally display the same data arranged per year, for the past five years. All reports, except for the citation report, are accompanied by a visual representation (pie, bar or line chart) of the chronological evolution of the KPI chosen. Behind all numbers in the reports detailed lists are stored which can be consulted (Fig. 3). The lists display the data upon which the number in the report is based - for instance the PhD student, the title, the supervisor(s) and the starting date in the case of the 'PhD projects in progress' report. These detailed lists are provided with predefined filters allowing the user to easily switch between views – e.g. only specific years within the five year overview. The report and the detailed lists can be exported together or separately in a number of formats (pdf, excel, text), depending on the end use. When exporting as a PDF the source of the data ('Research Statistics at KU Leuven'), the date of downloading and the staff ID of the downloader are automatically added to prevent reports being used without some context.

By January 2013 all was set to test the BO pilot project 'Research Statistics at KU Leuven'. The tool went live with a first package of reports (Staff, PhD projects, Research Projects excl. the externally financed bursaries, Publications & Citations, excl. IF report and top-publication report) in February 2013 and obtained its current form in July 2013 by the addition of the Funding, Externally Financed Bursaries, Top-Publications and Impact Factor reports.

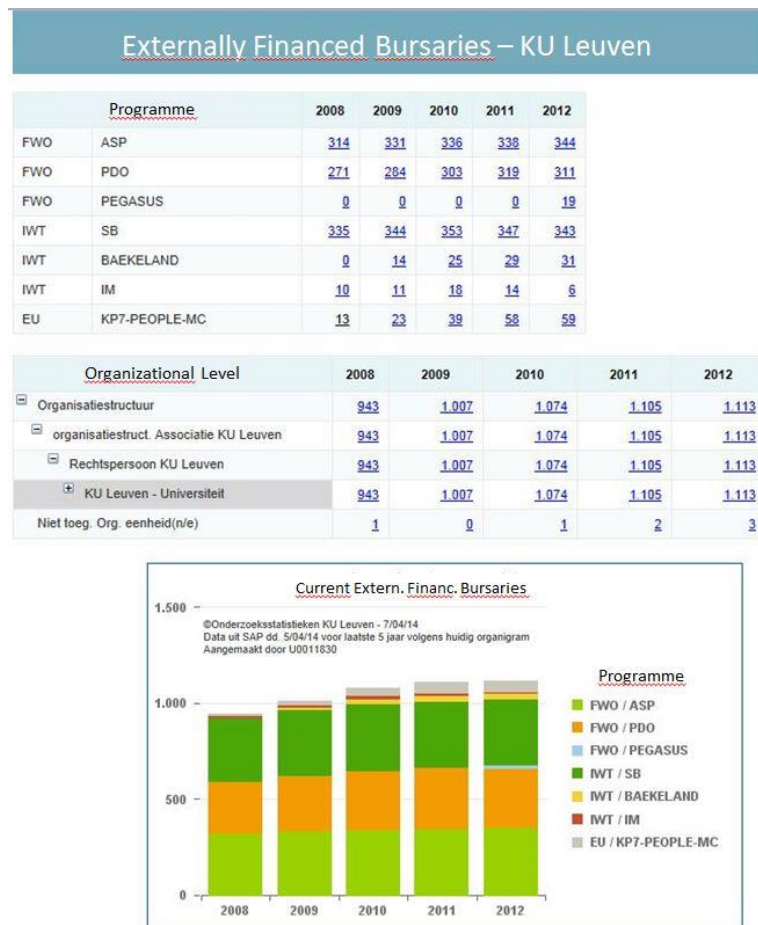


Fig. 3: Example of a 'Research Statistics' KU Leuven report: Externally Financed Bursaries

4. Evaluating the CRIS & its future

To date 'Research Statistics at KU Leuven' runs nearly a year. In general the user response is positive. The fact that all data are accessible through one application, the ease of use, the modern feel and look and the exporting options are all greatly appreciated.

From the beginning it was made clear that the tool is not meant to be static, but is open to additions and changes. In the course of the past months, changes in the organization of the university, as well as user feedback urged the revision of some aspects of the tool or an extension of the reports provided.

A major reorganization took place in October 2013, at the time when the academic university colleges, which were hitherto separate entities, were integrated in the structure of KU Leuven. This resulted in an inflow of new staff and students, new organizational units and new roles in research management. The organizational chart of the university was thoroughly reworked. These changes could easily be accommodated in the application 'Research Statistics at KU Leuven' due to the fact that affiliation and research output are not directly linked, but moderated through the staff ID. As long as these IDs remain unchanged, changes in the organizational chart are automatically processed. The effects of the reorganization on the CRIS were thus limited. The number and definition of authorized users had to be extended to incorporate staff from the university colleges and the KPIs 'Publications & Citations' and 'Staff' were augmented with categories specific to the university colleges.

Other adjustments were fuelled by end user requirements. The citation report, for instance, originally listed

citation counts for the past five years, but was extended to encompass the past ten years, providing better background data for research monitoring. Staff was counted as FTE but in the detailed breakdown of personnel lists, research managers expressed the wish to see the percentage of employment as well.

More modifications and add-ons are planned for the future. The university's technology transfer office, Leuven Research & Development (LRD), has expressed an interest in having an analogous tool at their disposal to manage data such as licenses & patents, spin offs, the development of which is planned for 2015. Once this is available, a future integration of both CRIS's could broaden the scope of reports on offer. Incoming feedback is monitored and discussed at regular time to determine whether it is appropriate for implementation, judged on the basis of technical opportunities and support by the user community.

Since 'Research Statistics at KU Leuven' is primarily a tool for internal use, displaying KPIs that are locally defined and relevant, there are at the moment no plans for an integration of the tool with applications outside the KU Leuven. However, recent developments at the Flemish level, where the re-launch of a regional CRIS with data from all five Flemish universities is planned, are closely followed to enable an easy exchange of data in the near future.

5. Conclusion

A research intensive university such as KU Leuven (Belgium) generates a multitude of research-related data. The registration and systematic collection of these information sources has been prone to different chronological and organizational trajectories. Most of this information was not provided in an aggregated way to policy makers. The need for an integrated research data portal gained momentum on the occasion of a series of internal research evaluations, involving all departments and research units of KU Leuven. This coincided with the news of the expected end of SAP-support of BEx Analyzer hitherto used for data analysis. Both developments urged the start-up of a joint CRIS project of the ICT, Research Coordination Office, HR & Finance department to provide aggregated data on research to authorized stakeholders for the purpose of research policy planning and monitoring. Thus was born the 'Research Statistics at KU Leuven' initiative. A consensus had to be reached as to the selection of relevant key performance indicators for research, the exact definition of the variables, and the way in which the data were presented. While the first challenge rapidly reached a satisfying conclusion, the preliminary reports struck bare the semantic complexity of the data. A sequence of consultation rounds with the data experts (ICT, HR, Finance and RCO) and the researchers followed which resulted in a clearly semantically defined set of relevant KPIs. Subsequently the multidimensional cubes in the BW Data Warehouse were optimized and reports were built with SAP BI-technology (Business Objects). KU Leuven's CRIS, 'Research Statistics at KU Leuven', live since mid-2013, now provides a tool to authorized users which is widely used. Feedback by users is actively solicited to be incorporated in future updates or expansions. The interface is extremely user-friendly and KPIs are numerically and graphically displayed, literally, by one click. It is deceptively simple – but that, in fact, is how it should be from the users' point of view.

References

1. <http://www.kuleuven.be/about/>
2. <http://lrd.kuleuven.be/en/index>
3. <https://admin.kuleuven.be/icts/services/anemoon/index> (Dutch only)
4. <https://lirias.kuleuven.be/>
5. dd. April 2014; cf. <http://repositories.webometrics.info>