The "Research Core Dataset" as a standard for research information: towards quality-assured and comparable data on research processes and outputs

Sophie Biesenbender [1]

OECD Blue Sky Forum on Science and Innovation Indicators

Informing science and innovation policies: towards the next generation of data and indicators

Ghent, 19-21 September 2016

Abstract

This paper presents an in-depth case study of the policy process behind the project to specify a Research Core Dataset (RCD) for the German science system. It discusses an approach that lays the foundation for generating quality-assured, consistent and reliable research information (RI) in a heterogeneous science system with multi-level governance. The analysis puts emphasis on the development process, involved actors and stakeholders, the results of the project as well as the principles for the development of the RCD standard. The paper also discusses the RCD project from a policy analytical perspective by comparing the project with other similar current initiatives on the national level. The paper concludes with a discussion of the adequacy of the RCD standard to develop useful science and technology indicators.

[1]   German Centre for Higher Education Research and Science Studies (DZHW), Berlin;
         E-Mail: biesenbender@dzhw.eu

Introduction

The development of science and technology indicators is heavily reliant on data generated and collected at decentral level in research institutions. Such information about the research process and its outputs (hereinafter research information or RI) is often based on e.g. institutional staff or funding information or combines data from institutional repositories with commercial databases (on e.g. publications by their staff). Research institutions in different science systems are usually faced with a number of external as well as internal reporting requirements, which they address with the help of these internal and external data sources. In some science systems, this is facilitated with the use of institutional or national current research information systems (CRIS).

Reporting purposes and requirements vary over contexts, recipients and bindingness, which is why few general standards exist. This is especially manifest in the German science system with fragmented regulation and responsibilities of both federal and state governments (Biesenbender and Hornbostel 2016a, 2016b). Research institutions gather different data for e.g. official statistics, reports to funding organizations, annual reports, rankings or internal controlling. There are only few cross-institutional standards or best practices for the decentralized collection and processing of research information. For this reason, the data are often not comparable over either research institutions or reporting purposes. This has negative implications for the quality of data that are currently used for institutional reporting and official statistics as well as in (comparative) science and innovation studies.

Thus, the German science system faces considerable pressure to optimize and streamline the processing and use of research information through the development of standards as well as the harmonization of reporting requirements, definitions of concepts and data formats. Consistent definitional standards for research information are likely to improve the quality and robustness of data used for institutional reporting, science and technology indicators as well as policy decisions.

Given the multi-level structure of the German science system with decentral competences for the regulation of higher education, harmonization is only possible through voluntary standards developed in bottom-up rather than linear top-down processes. Such an inclusive approach not only potentially increases the acceptance of the standardization endeavour as such. It is also suitable to develop standards that are adequate and feasible for implementation at decentral level (at research-performing institutions).

The recently completed project to develop a "Research Core Dataset" (RCD) for the German science system (2013-2015) was based on this assumption. The goal of the project was to propose a set of standard definitions and concepts for research information, which both research-performing institutions and report-requesting organizations (such as e.g. funding agencies) might flexibly use in the future to structure and harmonize their reporting processes.

This paper presents an in-depth case study of the policy process behind the project to specify a Research Core Dataset for the German science system. In so doing, it puts emphasis on the development process, involved actors and stakeholders as well as the principles for the development of the RCD standard. The paper is structured thusly. The next section explicates the focus and scope of the paper, followed by a description of the project and its results. This comprises an account of the processes and principles that were applied

in the RCD project. The section also illustrates the strategies for generating acceptance with future stakeholders and the larger public. The subsequent section discusses the project from a policy analytical perspective by comparing the RCD with other similar current initiatives on the national level. The paper concludes with a discussion of the adequacy of the RCD standard to generate useful science and technology indicators.

Focus and scope of the paper

The aim of this paper is twofold. On the one hand, it provides an introduction to the recently completed project to develop a Research Core Dataset for the German science system and illustrates the context of this particular case of standardizing definitions for research information (Riechert et al. 2016a: 304). On the other hand, it compares the characteristics of the RCD project to other national initiatives for the definition of (relevant) research information. This is done both in conceptual terms from an analytical point of view and with respect to the lessons learned as well as best practices.

The RCD project, its processes, principles and results

The following subsections summarize the standardization process, give an account of involved actors and stakeholders, depict the principles of the RCD project and present an overview of the results (i.e. the parameters of the RCD standard).

Development process

The standardization of research information has been debated in the German science system for a long time (Glänzel et al. 2016, Biesenbender and Hornbostel 2016a). It has constantly been on the political agendas since the implementation of the New Public Management in the 1980s in a number of higher education institutions (HEI).

There are different approaches and avenues to realize the standardization of research information across research performing and report-requesting organizations in Germany. One way is the adaptation and expansion of the legal basis for reporting research information (i.e. statistical laws for higher education institutions and non-university research institutions). However, this would target exclusively reporting for single purposes, such as official statistics. Yet, institutional reporting purposes comprise many other scenarios: e.g. reports to (public or private) research funding organizations, public authorities, to different agencies (e.g. the German Council of Science and Humanities), annual reports, private rankings or internal processes (such as controlling or performance-based funding).

These reporting processes involve a number of actors and can hardly be standardized through external intervention or regulation alone. Therefore early 2013, the German Council of Science and Humanities (henceforth, science council) started an initiative to establish a voluntary standard for research information – the so-called Research Core Dataset (RCD) (German Council of Science and Humanities 2013; see Biesenbender and Hornbostel 2016a, Biesenbender and Hornbostel 2016b). The subsequent project to develop a

Research Core Dataset for the German science system was carried out between 2013 and 2015.

Standardization in the context of the RCD project was thought to address two different dimensions. On the one hand, it should explicate, which research information is to be considered relevant and should be subject of future institutional reporting (conceptual standards). On the other hand, it had to deliver the definitions of the concepts (definitional standards for different parameters). Thus, the RCD is not dataset in a strict sense but implies a set of rules and recommendations for the standardized and decentralized collection and processing of research information in research institutions of different kind.

The RCD project was based on a paper by the science council, in which *focus areas* of relevant research information were already defined along with preliminary lists of *parameters* to be addressed and defined or specified in the RCD project (German Council of Science and Humanities 2013).

The focus areas of research information for which the project eventually derived standard definitions include (a) researchers employed by the institutions, (b) young researchers, (c) third-party funded projects, (d) patents and spin-offs, (e) publications and (f) research infrastructures (see Biesenbender and Hornbostel 2016b for details on the structure and logic of the RCD).

The project for the development of a RCD was co-ordinated by the Institute for Research Information and Quality Assurance (which merged with the German Centre for Higher Education Research and Science Studies in January 2016). The project was planned and carried out in close co-operation with the science council; it was funded by the Federal Ministry of Education and Research. In addition, a working group composed of members of the science council acted as the advisory board and supervisory committee throughout the project. The RCD project comprised three project stages. In the course of the project, discussion and decision making became more and more transparent and open to the public in order to include all perspectives and stakeholders of the German science system.

During the first stage, close expert group discussions dominated the decision-making process. During that stage, experts screened and evaluated existing definitions for the concepts of interest (as specified by e.g. official statistics, the Frascati manual, CERIF,[2] CASRAI,[3] see below). After first rounds of discussions the expert groups developed an initial draft specification for the RCD. This initial draft was then evaluated by a range of external co-operation partners, including a number of academic associations (71 out of 259), pilot institutions, software providers (for research information systems), funding organizations and the four umbrella organizations for non-university research institutions in Germany. The feedback was then incorporated in the final rounds of discussion and revision of the RCD standard in the expert groups. The first stage of the project concluded with the presentation of the *first* version of the RCD standard as developed and agreed on by the expert groups – the so-called alpha release.

---

[2]   Short for Common European Research Information Format.

[3]   Short for Consortia Advancing Standards in Research Administration Information.

During the second stage of the project, the advisory board assessed and adjusted the alpha release. In this context, the board invited further selected actors of the German science system to support the critical examination and evaluation. The second step concluded with the development of the so-called beta release, a modified version of the alpha release of the RCD standard.

The third stage included the on-line publication of the standard (i.e. the so-called open-beta phase) along with a public request for comments to interested stakeholders of the German science system. Representatives of different organizations (research-performing institutions as well as different types of report-requesting organizations) made use of this opportunity, which allowed commenting on the single parameters of the RCD standard. Through a standardized commenting tool participants were able to express support or criticism and to evaluate the respective definitions. The commenting tool was also intended to examine the usefulness and feasibility of the standard, its elements and definitions. For this purpose, participants were asked to evaluate the costs associated with collecting and processing the respective data and information for the different parameters of the RCD.

The third and final stage led to the final revision of the RCD standard. The comments and suggestions collected in the course of the open-beta phase were systematically analysed by the co-ordinating institution – the Institute for Research Information and Quality Assurance. Based on a description and analysis of the public feedback, the advisory board decided on changes to the beta release. Subsequently the project concluded with the specification of the *first official release* of the RCD standard. In January 2016, the German Council of Science and Humanities recommended implementation and use of the RCD standard specification to both research-performing institutions and report-requesting organizations (German Council of Science and Humanities 2016).

Actors and stakeholders

The decision-making process of the RCD project involved a number of different actors and stakeholders from the German science system. Overall, the debate was characterised by increasing transparency from closed expert consultations to an open-beta phase involving the interested public. Gradually increasing transparency did however not imply fundamental breaks in the formal decision-making rules. The decision-making process was based on the provisions of the German Council of Science and Humanities (German Council of Science and Humanities 2013). In the first stage, it included expert consultations, the subsequent development of a draft specification, a feedback procedure with external co-operation partners, the modification of the specification by the groups of experts based on external feedback and the final authorization of the results by the advisory board (a working group of the science council) that had strong supervisory powers. The decision-making rules within the expert groups were consensus based where feasible. Controversial issues were occasionally settled through majority decisions.

Experts to the project groups were appointed by the science council with the goal to represent all major actors and organizations from the science system with an interest or stakes in research reporting (including representatives from research-performing institutions of different kind, funding agencies, government, official statistics, science studies etc.; see German Council of Science and Humanities 2016). Given the complexity of the issue and the

objective to deliver exhaustive conceptual standards and definitions for as many as six focus areas, the project implemented innovative visualization tools that enable deliberation and decision making in complex situations (Riechert et al. 2016a, Riechert et al. 2016b). These visualization tools allowed for structured discussions and decision making by the experts. The tools recorded and illustrated the lists of pros and cons of the single decisions and respective alternatives considered or rejected during the discussion and decision-making processes throughout the project.

A major step towards enhanced transparency and openness as well as 'democratic legitimacy' of the results and the standard was made through the publication of the beta release of the standard and the public request for comments. The publication of the specification not only contained the concepts, their definitions and rejected alternatives but also the pros and cons of the respective decisions as they were brought up in the expert-group discussions. The open-beta phase covered a period of two months. Interested stakeholders were invited to submit comments with regard to the single parameters of the RCD in a structured format. More than a 100 hundred representatives of different institutions and organizations from the German science system made use of this possibility. In total, 1,820 specific comments (concerning single parameters and definitions of the specification) were gathered through a standardized on-line commenting tool. The number of negative comments slightly exceeded the positive comments. It is worth noticing that participants to the open-beta phase solely had the possibility to comment on parameters of the RCD and the respective definitions as specified in the beta release. They were not asked to suggest additional parameters or items or to propose far-reaching or fundamental changes to the specification. In other words, they were able to comment on the definitional standards but not the conceptual aspects of the RCD, i.e. which research information is to be considered relevant and should be subject of future reporting by research-performing institutions.

The advisory board went through the suggested changes with a focus on those parameters for each of the focus areas of the RCD that caused the most controversy in the public (in terms of the number of negative comments). Subsequently, the advisory board decided on the changes to the beta release with regard to the different focus areas.

In sum, the development processes in the project is based on both discussions in closed expert groups and public commenting of the draft specification. Increasing transparency and openness encouraged many stakeholders to actively participate in the standardization project (see Biesenbender and Hornbostel 2016a).

Principles of the RCD project

The principles and decision-making rules of the RCD project were chosen to enhance the legitimacy of decision making and the standard itself (cf. Biesenbender and Hornbostel 2016a).

First, the goal was to create acceptance through *process legitimacy* (Riechert et al. 2016b; Werle and Iversen (2006) refer to the concept as "input legitimacy"). Such a strategy is suitable to not only address the "legitimacy deficit" (Werle and Iversen 2006: 20) inherent to many standardization projects, but more importantly to create the acceptance, knowledge and dissemination needed for the RCD standard to be successful in the science system in

the future. Success of the project will eventually depend on the wide use and diffusion of the RCD standard in both research-performing institutions and report-requesting organizations. Complex and costly decision-making processes and the use of tools for computer supported argument visualization (Riechert et al. 2016a, Riechert et al. 2016b) in the RCD project illustrate the goal to reach and involve as many stakeholders in the development process as possible. This is true not only for the open-beta phase but also for the expert groups that considered feedback from a high number of external co-operation partners (see above).

Second and related to the issue of process legitimacy, the RCD project sought to enhance *output legitimacy* (Werle and Iversen 2006) by developing good and adequate standards that are indeed apt to reduce the costs of managing and processing of RI data in the mid or long term. Transparency, openness and the publication of the pros and cons of different solutions were intended to constantly advance and improve project discussions, opinion-forming and decision-making processes.

Third, the standard was designed to interfere as little as possible with the *decentralized management* of research information and institutional reporting processes. In other words, the standard should respect institutional autonomy and existing legal requirements (with regard to e.g. data protection issues or intellectual property rights). For this reason, the RCD is a standard specification for existing decentralized processes of RI management but not the establishment of a central database or new reporting requirements. "Research institutions remain the owners of their data, which they however hold available in an aggregated and standardized format (i.e. the RCD as a white book instead of a database). As opposed to other approaches, the RCD does not entail the establishment of a national current research information system" (Biesenbender and Hornbostel 2016a: 845).

Results of the RCD project

The current release of the RCD standard covers six focus areas with different elements or parameters. Table one summarizes the specification for the six focus areas. The recommendations by the science council (German Council of Science and Humanities 2016) report the full standard: that is the parameters of the RCD standard along with their definitions and rejected alternatives.

Table 1:   Focus areas and parameters of the RCD

Focus area	Parameters	Relation with other parameters of the RCD	Classification
Employed researchers	- gender - citizenship - date of birth - personnel category (for each employment contract) - salary, denomination, information on joint appointment if apl., name of the second institution (for "professors" only) - percentage of full-time position (for each employment contract) - start and end of contract (for each employment contract) - type of funding - highest academic degree - type of work	- supervisor to a doctoral student (see " Young researchers") - participant in a doctoral program (see "Structured doctoral programs") - speaker of a doctoral program (see "Structured doctoral programs") - inventor (see "Patents") - author (see "Publications")	- organizational unit - discipline - research field
Young researchers			
a) Young researchers	- start of the dissertation - supervisor at the research institution - participation in structured doctoral program - co-operation with other research institutions - country where the entitlement to proceed with a doctorate has been awarded - date of graduation (doctorate) - date of completing the habilitation procedure	- participant in a doctoral program (see "Structured doctoral programs") - inventor (see "Patents") - author (see "Publications")	- organizational unit - discipline - research field
b) Structured doctoral programs	- title - participating institutions and countries - funding of the program	- speaker (see "Employed researchers") - doctoral student or graduate (see " Young researchers")	- organizational unit - discipline - research field
Third-party funded projects	- title of the project - name of the project co-ordinator (institution) - start / end of the project - amount of funding - project family if apl.	- funding organization (see "Publications") - grant agreement number (see "Publications")	- organizational unit - discipline - research field
Patents and spin-offs			
a) Patents	- title of the patent - date of initial registration - patent ID - patent family - date of registration	- inventor (see "Employed researchers")	- organizational unit - discipline - research field

Focus area	Parameters	Relation with other parameters of the RCD	Classification
b) Spin-offs	- title of the spin-off - date of the spin-off		- organizational unit - discipline - research field
Publications	- type of author - title - year - publisher - place - source - Identifier - language code - access rights - format (volume, issue, name of the conference, page range if apl.) - peer-review - thesis - document type - publication type - resource	- author (see "Employed researchers") - funding organization (see "Third-party funded projects") - grant agreement number (see "Third-party funded projects")	- organizational unit - discipline - research field
Research infrastructures	- title - description - type - type of access - use intensity - operator - co-ordinator - staff (category)	- publications (see "Publications")	- organizational unit - discipline - research field

Source: Own elaboration based on German Council of Science and Humanities (2016) and adapted from Biesenbender and Hornbostel 2016b: 407.

The Research Core Dataset as a standard and policy instrument for the German science system

The aim of the RCD project was to derive and establish standards for research information. Standards play a role in both the political science literature (e.g. Van Vught 1995) and the standardization literature (e.g. Borraz 2007). The political science literature, on the one hand, often focuses on the nature of the policy instrument, i.e. whether it is a soft or hard instrument (see e.g. Van Vught (1995) who distinguishes between two categories of policy instruments: (hard) directives and (soft) incentives). Scholars of policy science often take a narrow perspective on *standards as obligatory policy instruments* with regulatory focus, for example by analysing binding technical requirements for industrial plants in order to mitigate emissions or the definition of water quality standards (see e.g. Jordan et al. 2005; cf. Werle and Iversen 2006). An example for mandatory standards in the field of science studies and higher education research is the standardization of national degrees as initiated by the Bologna process (Van Vught 1995).

In contrast to the policy science perspective, the standardization literature conceives of *standards as co-coordinative instruments* (Werle and Iversen 2006, Jordana and Levi-Faur 2004, Majone 1999). This perspective typically categorises standards according to their

development and generally distinguishes between two types of standards: de-jure and de-facto standards. While the development of de-jure standards is officially endorsed by government authorities and usually commissioned to a (national) standards institute, de-facto standards are established as the result of dominant use or convention and "emerge[s] through the mediation of market processes" (Borraz 2007: 58). De-jure standards are either voluntary or obligatory, whereas de-facto standards are usually non-binding.

The RCD is a voluntary standard. Research-performing institutions and report-requesting organizations do not have to adhere to the specification. Though being a voluntary standard, the initiative and process to develop the RCD differs from conventional de-jure or de-facto standardization processes for different reasons. A closer look at the history of research information standardization in Germany shows that self-sustained "market mechanisms" were not sufficient drivers for the development of RI standards in the science system. A number of past initiatives to standardize research information or different aspects of research information did not have any systematic effect on the institutional processing and exchange of information and data on research processes and their outputs (Biesenbender and Hornbostel 2016b). These initiatives date back to the first edition of the Frascati Manual by the Organization for Economic Co-operation and Development (OECD) in 1963. This and further editions of the manual contain mainly conceptual standards for the collection of research information and institutional classification. The CERIF (short for Common European Research Information Format) initiative, which was launched in 1987, is another example. CERIF offers standards for data formats and data models. In this sense, it provides a common syntax and semantics for processing research information in research information systems. The CERIF standard has been established in other science systems. It was adjusted and further developed in 1997. Finally, CASRAI (Consortia Advancing Standards in Research Administration Information) is a non-profit organization founded in 2006. CASRAI offers a shared standard vocabulary (both semantics and classification systems) for research information.

All three initiatives provide standards for different facets of research information. Nevertheless, they have so far not contributed to the harmonization of RI management in German research-performing institutions. The lack of a standard specification for research information is further aggravated by the heterogeneity of the systems and technical solutions in place to manage and process research information.

In general, the German science system is characterized by considerable heterogeneity. There are currently 239 state-regulated public higher education institutions and 255 non-university research institutions that are jointly financed by federal and state governments (through the Joint Science Conference) (Biesenbender and Hornbostel 2016b). While universities and higher education institutions fall under sub-national state jurisdiction (there is a total of 16 states with different legal provisions), non-university research institutions are within the scope of federal regulation. Regulatory fragmentation and institutional heterogeneity of the German science system has led to different strategies of research-performing institutions to deal with and manage reporting requirements. A recent survey of higher education institutions in Germany (Sticht 2015) revealed the degree of institutional professionalization in the management of research information. According to the survey, only 16 percent of German universities have sophisticated and integrated research information systems in place. A share of 24 percent uses rather simple information systems. Still, the survey also indicates a trend towards the rationalization of data management and reporting

processes. The survey reveals that 45 percent of German universities plan to streamline their processes by implementing research information systems in the near future (see Sticht 2015, DINI-AG Research Information Systems 2015).

Structural heterogeneity and limited professionalization in the management of research information in research-performing institutions might challenge the eventual success of the standard and the RCD project. In this way the German case arguably qualifies as an extreme case for RI standardization (Riechert et al. 2016a, Seawright and Gerring 2008).

The complexity and regulatory heterogeneity of the German science system are the main reasons for the decision to develop a *voluntary* standard for research information whose future success will ultimately depend on acceptance in the science system. The RCD standardization process has features of both de-jure and de-facto standards. Its development was originally initiated by an organization with close government ties, which makes the RCD similar to de-jure standards. It was recommended by the German Council for Science and Humanities (an organization advising German federal and state governments) and subsequently financed by the Federal Ministry of Education and Research. The development process itself, however, was then co-ordinated by a research institution (Institute for Research Information and Quality Assurance) and not by an official standards institute. The co-ordinator does not have the same authority as a (national) standards institute in the development of de-jure standards.

While the initiative was in a way top-down, the actual implementation process – and hence the eventual success of the standard – resembles a de-facto standardization process. In fact, the idea to include as many research-performing institutions and report-requesting organizations as possible in the discussion and feedback phases demonstrates the objective of the project to establish something similar to a 'market' for research information that will enhance the success of the RCD standard. The idea behind the project was to develop and propose a standard specification for research information that is more convincing to all stakeholders in the science system than existing alternatives or a multitude of different parallel solutions. Having to report research information according to different competing definitions is supposed to be costly and detrimental to the quality of data and reports. The degree of implementation of the RCD in both research-performing institutions and report-requesting organizations will eventually show whether the RCD qualifies as a satisfactory standard that is more favourable to the stakeholders than the previous situation of multiple report-specific RI definitions for one and the same concept. In sum, the goal of the RCD project was to develop definitions in co-operation with all stakeholders of the science system. The assumption was to eventually establish the RCD as the result of dominant use or convention by the targets of standardization (i.e. research-performing institutions and report-requesting organizations). For this reason, the development of the RCD standard was a bottom-up process

However, the RCD is not the only policy instrument in the German science system to define a specification for research information. There are currently other parallel initiatives that focus on the definition of suitable research information, for example in the focus area of young researchers (doctoral students, see above). Doctoral education and research in Germany is an area that particularly lacks systematic evidence and research. Official data on the total number of doctoral candidates at German universities is missing. Contrary to other focus areas (like e.g. publications or third-party funded projects) reporting on young

researchers is not yet very established in the German science system. In other word, the goal is not to standardize existing and competing definitions but to develop 'new' suitable and feasible definitions also for future reporting purposes. To address the lack of missing definitions and concepts, other actors simultaneously started similar standardization initiatives (beyond the RCD).

First, a recent reform of the federal statistics law on higher education institutions (HEI) makes the decentral registration of doctoral students obligatory for German universities. In the future, they will have to report on doctoral students and their characteristics (gender, age, citizenship, start and end of doctoral program, academic subject etc.) to official statistics on an annual basis (BMJV 2016).

A second initiative to achieve consistent information on the situation of doctoral students in Germany was started by representatives of HEIs themselves. A network comprising 49 [4] German universities – the German University Association of Advanced Graduate Training (UniWiND/GUAT) – was founded in 2009 with the objective to improve (post-)doctoral education and training in German HEIs. In this context, a working group of the network developed definitions for the decentral registration of amongst others doctoral students in HEIs. The goal was to ease institutional reporting requirements not only towards official statistics (see above) but also for other purposes. The working group developed an extensive specification that goes well beyond the RCD and the current version of the federal statistics law on HEIs (UniWiND 2015).

The three initiatives co-ordinated their activities as far as possible during the development processes. Figure one illustrates the overlaps between the contents of the three initiatives. As the figure shows, the RCD standard is less comprehensive than the provisions of the statistics law and the UniWiND initiative. The latter is most detailed by containing as many as 68 parameters that go beyond the other two standard specifications. Altogether, there are strong overlaps between the three initiatives. The overlaps indicate a considerable consensus between the participants of the different projects about the relevant concepts of the specification. In other words, there is little disagreement across the initiative about the question what information on young researchers is relevant or key.

---

[4]   Out of a total of 111 universities that have the right to award doctoral degrees.

Figure 1:   Overlaps between different standardization initiatives (focus area: young researchers)
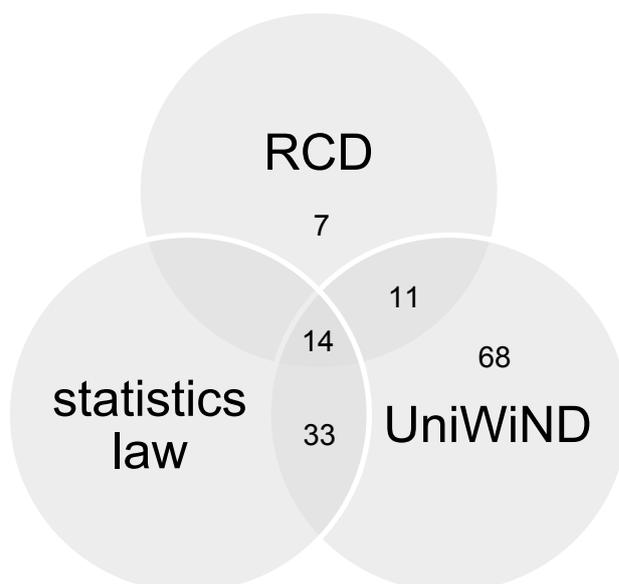
Table two summarizes the characteristics of the three initiatives to standardize research information (RI) on the focus area of young researchers (doctoral students).

Table 2:   Different initiatives of RI standardization in the German science system (focus area: young researchers)

	Type of initiative / legitimacy	Scope / targets	Involved actors	Degree of binding-ness	Development process
RCD	RI definition by deliberation	Research-performing institutions, report-requesting organizations	Stakeholders of the science system	Weak, voluntary	Expert groups, open/public consultations
UniWiND	RI definition by deliberation	Research-performing institutions (only HEIs)	Representatives of HEIs	Weak, voluntary	Working group of the UniWiND network (49 universities)
Statistics law (for HEIs)	RI definition by law	Research-performing institutions (only HEIs)	Federal government (legislative and executive branch)	Strong, mandatory	Federal legislation

All three initiatives propose a standard specification for the processing of information and data on young researchers. Yet, there are different logics and strategies inherent to the different standardization processes.

The RCD was initiated by government (as advised by the German Council of Science and Humanities) and delegated to a research institute as an objective co-ordinator of the development process. Given complex regulatory structures and responsibilities (multi-level governance) in the German science system the RCD has to convince by its contents and the transparency and (process) legitimacy of its development process. While the RCD was initiated from above (top-down) and developed in a bottom-up process of increasing transparency and openness (bottom-up), the reform of the statistics law and the UniWiND project show more coherent approaches (see table three). On the one hand, the reform of the statistics law on HEIs was initiated from above and carried out through regular legislative processes of the federal government (involving both the executive and legislative branch). On the other hand, the UniWiND proposal is based on a bottom-up initiative by the German University Association – that is representatives from higher education institutions themselves. The specification was subsequently developed by an internal project group (self-organization) composed of representatives of HEIs that are members of the association.

However, in contrast to the RCD project, the both the reform of the statistics law and UniWiND have a more limited scope than the RCD. While the RCD standard potentially targets all research-performing institutions and report-requesting organizations in the German science system, the UniWiND project and the reformed statistics law apply to higher education institutions (a subset of research-performing institutions) only (see table two).

Table 3:   Comparing processes of RI standardization in the German science system (focus area: young researchers)

| | | Nature of the development process | |
		Top-down	Bottom-up
Origin of initiative	**Top-down (by the state)**	Statistics law on HEIs	RCD
	Bottom-up (by the targets)		UniWiND

Source: Own elaboration.

The RCD project was guided by the idea to enhance the success of the standard by focusing on process legitimacy of the standardization process (through bottom-up development). This approach is not new but has hardly been applied in actuality to other standardization projects (Werle and Iversen 2006). Whether such a strategy is worth the investment cannot be determined at this time, because the RCD standard still lacks broad implementation and usage in the German science system. Nevertheless, the current trend towards institutional professionalization in the processing of RI data (through increased implementation of current research information systems; see above) indicates the demand for standards and best-practices for the management of research information.

Expected advantages for research institutions to provide research information in the RCD standard include reduced efforts and more efficient processes for the collection, management and processing of research information. In addition, broad involvement of stakeholders along with openness and transparency in the development of the specification enhance the process legitimacy of the standard.

In the future, information or report-requesting organizations might also rely on the standard because they expect reporting organizations to be able to provide the research information in this format. In the mid-term, such a standard will likely increase the comparability of research information. Thus, reliance on the RCD standard has the potential to not only create reliable and quality-assured research information but also to indirectly improve science and technology indicators derived from these data.

The RCD project might serve a blueprint to other (national or international) initiatives of RI standardization in complex systems with multiple levels and stakeholders. The design of the RCD project could for instance be scaled up and implemented in other (national or supranational) cases of RI standardization (such as e.g. CERIF or CASRAI) that seek to enhance process legitimacy. This would however require sufficient financial and personnel investment in order to ensure the adaptation and implementation of complex tools for computer supported argument visualization (Riechert et al. 2016a, Riechert et al. 2016b). Experience from the RCD project shows that the use of innovative tools for moderating decision-making processes is necessary for streamlining complex discussions and problem solving with multiple participants and different forms of involvement. The involvement of a high number of stakeholders in the RCD project (see above) was only possible with the help of real-time on-screen modelling of the discussions and their results (i.e. the definitions, alternatives and related pros and cons) (Riechert et al. 2016a). In addition, these tools laid the foundation for transparency, open project documentation and subsequent publication. Such an approach in turn requires the use and customization of software and considerable personnel resources to be implemented. In other words, focusing on process legitimacy (Werle and Iversen 2006) is an expensive and time consuming endeavour (Riechert et al. 2016b: 7). Nevertheless, it might be the only way to reach sustainable standards in complex multi-level governance systems.

The RCD standard has not yet been fully implemented in German research-performing institutions and report-requesting organizations, which currently makes an evaluation of its (mid- or long-term) impact and success impossible.

Conclusion

The paper provided an introduction to the project to develop a Research Core Dataset for the German science system. In so doing, it presented an approach that lays the foundation for generating quality-assured, consistent and reliable research information in a heterogeneous science system with multi-level governance. The paper contained an in-depth analysis and systematic account of the RCD project and discussed the conversion and applicability of the approach to other use cases.

Being a voluntary standard for internal and external reporting by research institutions, the success of the RCD standard ultimately depends on acceptance with potential users. This

assumption has had implications for the design and decision-making processes of the RCD project (Riechert et al. 2016a): given complex responsibilities and multi-level regulation in the German science system, standardization can only be achieved on the basis of voluntary agreement and implementation by all stakeholders – organizations requesting research information and organizations providing research information.

Comparable and reliable data that can subsequently be used for deriving science and technology indicators requires consistent research information. The trend towards professionalization in institutional data processing in German research institutions is a first necessary step towards more efficient and consistent management of research information. Even though German HEIs lag behind in terms of professionalization of RI management, there is a trend towards more efficient systems and processes at the decentral level in research institutions.

The factual quality of indicators on science and technology processes – and hence their adequacy for subsequent policy decisions – crucially depends on the underlying data. Poor data quality not only impacts negatively on indicators and their interpretability but also diminishes the acceptance of the use of metrics to inform policy decisions in general. In this regard, standards may enhance the quality of data and indicators for two main reasons.

On the one hand, use and interpretation of indicators are often limited due to incomplete or incompatible data. This in turn may be a consequence of limited professionalization in the management of data collection and processing at the decentral level in research institutions. Gathering data on research processes and outputs typically involves different levels of management. Complete recordings of an institution's publications for instance require input from both libraries and researchers and may be complemented with external (commercial) publication databases. Records of publications are necessary not only for bibliometric analyses, reporting to funding organizations or internal purposes like controlling or performance-based funding. They are also relevant for public relations or a researcher's personal website. These different scenarios (cf. Biesenbender and Hornbostel 2016b: 402f. for further details) illustrate the usefulness of a standard specification for keeping the efforts of data collection and processing as low as possible for all involved actors and for enhancing the acceptance of continued data collection in general. Thus, rationalized data management processes and the implementation of data standards are likely to improve the quality of the raw data and their consistency across different institutions. This, in turn, eases the use and traceability of indicators across different units of analysis (e.g. research-performing institutions or individual researchers).

On the other hand, standards for the decentral collection and processing of data as recommended through the RCD (see German Council of Science and Humanities 2016) allow research-performing institutions to flexibly process and aggregate the 'base data' for different reporting purposes or indicator construction (see Biesenbender and Hornbostel 2016b: 404ff. for details; see above table one). In this way, they might create the basis for a more systematic development and comparison of different variants of indicators and alternative metrics.

Altogether, the RCD standard potentially creates the empirical basis not only for the use of science and technology indicators but also for the development, evaluation and comparison of indicators.

In sum, the main objective of the RCD project was to develop standards that reduce efforts and costs for the institutional reporting of research information by improving the quality and comparability of RI data processed in different research institutions. In the long term, such a standard has the potential to improve the quality of reports on scientific performance and to also indirectly enhance the data used for other purposes such as the use and construction of indicators. However, whether the RCD standard will actually facilitate the development or improvement of indicators for science and technology studies or policy making eventually hinges on the comprehensive implementation of the RCD specification, its acceptance and success in the German science system.

References

Biesenbender, S., and Hornbostel, S. (2016a). The Research Core Dataset for the German science system: challenges, processes and principles of a contested standardization project. *Scientometrics*, 106(2), 837–847. doi: 10.1007/s11192-015-1816-y.

Biesenbender, S., and Hornbostel, S. (2016b). The Research Core Dataset for the German science system: developing standards for an integrated management of research information. *Scientometrics*, 108(1), 401–412. doi: 10.1007/s11192-016-1909-2.

Borraz, O. (2007). Governing Standards: The Rise of Standardization Processes in France and in the EU. *Governance: An International Journal of Policy, Administration, and Institutions*, 20(1), 57–84. doi: 10.1111/j.1468-0491.2007.00344.x.

Bundesministerium der Justiz und für Verbraucherschutz – BMJV (2016). Gesetz über die Statistik für das Hochschulwesen sowie für die Berufsakademien (Hochschulstatistikgesetz – HStatG). Änderung durch Artikel 1 des Gesetzes vom 2. März 2016 (BGBl. I S. 342. Available: http://www.gesetze-im-internet.de/bundesrecht/hstatg_1990/gesamt.pdf [Accessed 25 July 2016].

Glänzel, W., Beck, R., Milzow, K., Slipersæter, S. Tóth, G. Kołodziejski, M., and Chi, P. (2016). Data collection and use in research funding and performing organizations: General outlines and first results of a project launched by Science Europe. *Scientometrics* 106(2), 825–835. doi: 10.1007/s11192-015-1815-z.

DINI-AG Research Information Systems. (2015). Research information systems at universities and research institutions: Position paper. *Zenodo*. doi: 10.5281/zenodo.17491.

German Council of Science and Humanities (2013). Empfehlungen zu einem Kerndatensatz Forschung (Drs. 2855-13), Berlin, Germany. Available: http://www.wissenschaftsrat.de/download/archiv/2855-13.pdf [Accessed 25 July 2016].

German Council of Science and Humanities (2016), Empfehlungen zur Spezifikation des Kerndatensatz Forschung (Drs. 5066-16), Berlin, Germany. Available: http://www.wissenschaftsrat.de/download/archiv/5066-16.pdf [Accessed 25 July 2016].

Jordana, A., Wurzel, R., and Zito, A. (2005). The Rise of 'New' Policy Instruments in Comparative Perspective: Has Governance Eclipsed Government? Political Studies, 53(3), 477–496. doi: 10.1111/j.1467-9248.2005.00540.x.

Jordana, J. and Levi-Faur, D. (2004). The Politics of Regulation in the Age of Governance. In J. Jordana and D. Levi-Faur (eds.), The Politics of Regulation: Institutions and Regulatory Reforms for the Governance Age. Manchester: Edward Elgar.

Majone, G. (1999). The Regulatory State and Its Legitimacy Problem. *West European Politics*, 22(1), 1–24. doi: 10.1080/01402389908425284.

Riechert, M., Biesenbender, S., Dees, W., and Sirtes, D. (2016a). Developing definitions of research information metadata as a wicked problem? Characterisation and solution by Argumentation visualisation. *Program*, 50(3), 303-324. doi: 10.1108/PROG-01-2015-0009.

Riechert, M., Biesenbender, S., and Quix, C. (2016b). Developing and Standardising Definitions for Research Information: Framework and Methods of Successful Process Documentation. *Procedia Computer Science*, online first. uri: http://hdl.handle.net/11366/504.

Seawright, J. and Gerring, J. (2008). Case selection techniques in case study research a menu of qualitative and quantitative options. *Political Research Quarterly*, 61(2), 294–308. doi: 10.1177/1065912907313077.

Sticht, K. (2015). Einsatz von Forschungsinformationssystemen an Universitäten und Hochschulen mit Promotionsrecht in Deutschland: Ergebnisbericht. *Zenodo*. doi: 10.5281/zenodo.13841.

Universitätsverband zur Qualifizierung des wissenschaftlichen Nachwuchses in Deutschland – UniWiND (2015). Handreichung der UniWiND-AG Promovierendenerfassung. Arbeitspapier vom 2.11.2015. Available: http://www.unikon.uniwind.org/fileadmin/media/documents/UniKoN/151105_Arbeit spapier_Promovierendenerfassung_Final.pdf [Accessed 25 July 2016].

Van Vught, F. A. (1995). Policy Models and Policy Instruments in Higher Education: The Effects of Governmental Policy-Making on the Innovative Behaviour of Higher Education Institutions. *Institute for Advanced Studies, Vienna, Political Science Series No. 26*. Available: http://www.ssoar.info/ssoar/handle/document/26695 [Accessed 25 July 2016].

Werle, R. and Iversen, E.J. (2006). Promoting legitimacy in technical standardization. *Science, Technology & Innovation Studies*, 2(1), 19–39.