Information Services & Use 40 (2020) 87–113
DOI 10.3233/ISU-200076
IOS Press

Mapping and semantic interoperability of the German RCD data model with the Europe-wide accepted CERIF

Otmane Azeroual[a,*] and Nico Herbig[b]
[a]*German Center for Higher Education Research and Science Studies (DZHW), Berlin, Germany*
[b]*University of Applied Sciences Zwickau, Zwickau, Germany*

Abstract. The provision, processing and distribution of research information are increasingly supported by the use of research information systems (RIS) at higher education institutions. National and international exchange formats or standards can support the validation and use of research information and increase their informative value and comparability through consistent semantics. The formats are very overlapping and represent different approaches to modeling. This paper presents the data model of the Research Core Dataset (RCD) and discusses its impact on data quality in RIS. Subsequently compares it with the Europe-wide accepted Common European Research Information Format (CERIF) standard to support the implementation of the RCD with CERIF compatibility in the RIS and so that institutions integrate their research information from internal and external heterogeneous data sources to ultimately provide valuable information with high levels of data quality. As these are fundamental to decision-making and knowledge generation as well as the presentation of research.

Keywords: Research core dataset (RCD), CERIF, research information, institutional information systems, research information systems (RIS), data models, standardization, data quality, mapping, interoperability, semantics, knowledge representation, science system

1. Introduction

Standardization of research information helps universities and non-university research organizations to aggregate, reuse and shares their research information. The demand for quality-assured and comparable research information has increased with the introduction of control mechanisms in accordance with New Public Management in the German higher education system. As a result of numerous and diverse reporting obligations, universities and non-university research institutions have begun to introduce research information systems (RIS) in recent years. An RIS is understood to mean a *specialized database* or *federated information system* that can collect, manage and provide information about research activities and their results [1]. Nowadays, the provision and exchange of research information is done via a RIS. National and international standards exist to support RIS and to allow compatibility and interoperability between different systems as well as to represent the research area. The Europe-wide accepted Common

*Corresponding author: Otmane Azeroual. Email: azeroual@dzhw.eu.

European Research Information Format (CERIF) standard is founded and maintained by the European organization euroCRIS[1] and recommended to European member states for the administration and exchange of research information. It describes relevant object types from a wide range of research and development areas. In addition to the introduction of RIS in Germany, the German Council of Science and Humanities (in German "Wissenschaftsrat") initiated a process for the specification of a RCD[2] (in German "Kerndatensatz Forschung (KDSF)" in 2013 [6]. The offer of RCD is a voluntary standard for German universities and non-university research institutions and is recommended by the German Council of Science and Humanities in 2016. This RCD in its version 1.0 completed in 2015, provides a basis for providing and disseminating information about research activities [3,4]. The two technical standards essentially include data model specifications, Extensible Markup Language (XML) and semantics, and are publicly available on the euroCRIS and RCD websites. With their help, the data maintenance and data provision processes as well as the data quality in the context of data queries and reporting processes can be improved. Both the internal use and the distribution of comparable information on research activities can be facilitated. At the same time, this will reduce the workload for researchers and administrations in the medium to long term. Clear and standardized definitions increase the validity of the data and make it easier to use [7]. In order to support and facilitate the implementation of the core definitions for research information and their easy exchange within the framework of the German science system, the aim of this paper is to present the technical data model of the RCD and its implementing impact on data quality in RIS. Afterwards to compare it with the European CERIF standard in order to implement the RCD with compatibility of CERIF into the RIS and to enable the international connectivity of the RCD.

2. Description of the RCD and CERIF data model

This chapter first introduces the RCD data model and its impact as an application case to the quality of RIS. Finally, the international CERIF data model will be presented.

2.1. RCD data model

The recommendation for the development and implementation of a RCD has the goal of both the standardized recording and updating of the performance data on research activities of universities and non-university research institutions in the context of decentralized data management [7] and the best practice for a better data quality of the RCD to reach research information. In 2016, the German Council of Science and Humanities published the recommendations for the specification of the RCD. Since February 2017, a central helpdesk of the German Center for Higher Education and Science Research (DZHW) supports the interpretation of the RCD specification. RCD defines six different areas of research reporting (*employees*, *promotion of young talent*, *third-party funded projects*, *patents and spin-offs*, *publications* and *research infrastructures*) and these are divided into so-called core data and their characteristics and aggregation measures based on existing definitions and standardization (such as CERIF, FRASCATI, CASRAI). To support the interoperability and longevity of research information, a technical RCD data model will be presented compatible with the CERIF data model for in-house data provision. Therefore, the RCD extends the CERIF data model and adds further entities and attributes. The RCD data model is further divided

[1] https://www.eurocris.org/

[2] https://www.kerndatensatz-forschung.de

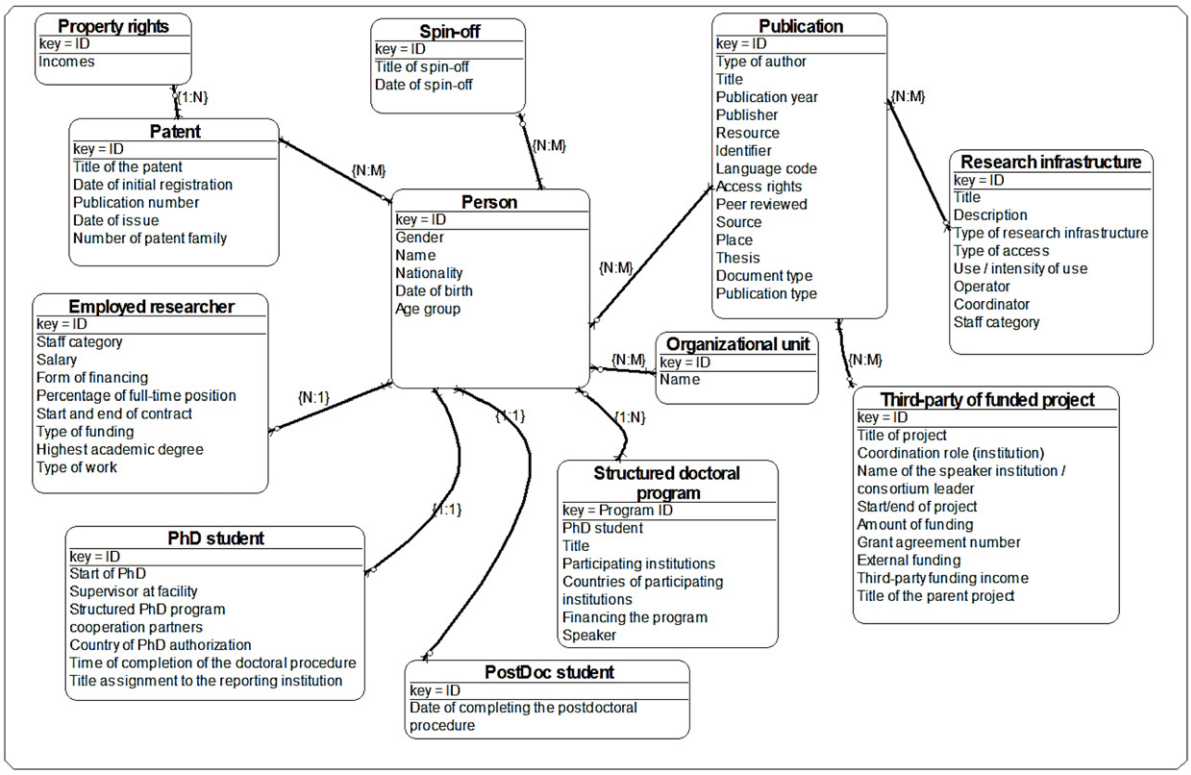

Fig. 1. RCD Entities and their relationships from the basic data model.

into basic data and aggregate data. The basic data model corresponds to the objects, the description of the objects with the relationships and attributes. The aggregate data model defines only the core data, without characteristics or specializations. However, the basic data model provides person-related information, whereas the aggregation model does not. The RCD data model was created at baseline and at aggregation level using an XML Schema and in the Web Ontology Language (OWL) modeling language. Further details about the XML Schema of the RCD can be found in [5,11].

Figure 1 shows the Entity Relationship Model (ERM) of the RCD. This contains the underlying objects of the specification, their attributes and the relationships between them.

2.2. *Impact of implementing RCD on data quality in RIS*

The German approach to standardization of research information reflects the heterogeneous research landscape and federal governance structure of Germany [4]. RCD serves as orientation for institutions intending to represent the RCD in their technical systems. Implementation can feasibly take place at both institutional and RIS provider level; both instances can be observed in the German science system. The RCD's XML Schema can be utilized as a data source before importing into RIS and/or as an export format to facilitate report creation.

While the introduction of the RCD has likely numerous effects on research information management processes and research information quality, we focus here on effects we perceive to most immediately impact the data quality dimensions addressed in this paper. First, the standard provides the basis for a

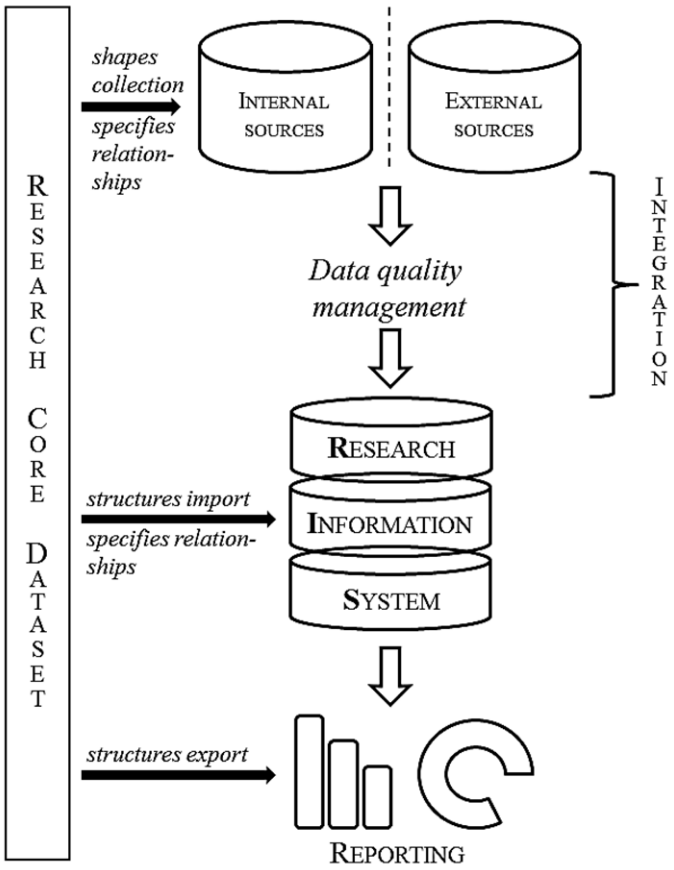

Fig. 2. Research information quality management framework.

common understanding and interpretation of research information through its semantic specifications, thus likely improving consistency of the data over time and across institutions, as well as correctness and completeness. Second, it structures the data acquisition process at institutional level and, especially if incorporated in RIS software, potentially reduces the need to harmonize previously heterogeneous data sources and formats. Impact on correctness and completeness of the data is expected here as well. In addition, it specifies relationships between research information entities, which in combination with RIS capabilities facilitate data integration. We expect this aspect of the RCD to impact correctness, consistency as well as timeliness of the relationships described. All the impacts described here will be mediated by existing data quality assurance procedures present in Higher Education and research institutions. Figure 2 provides an overview about the research information management process and the RCD's impacts.

With the increasing integration of research information from various sources in RIS and their growing importance for institutional management, data quality is becoming a growing area of interest for Higher Education and research institutions. Incorrect, inconsistent, inaccurate and missing data will lead to erroneous research information and interfere with decisions within an institution. In order to avoid the most costs in the academic institutions, a holistic data quality management process is required in RIS. The framework presented in this paper provides institutions with the means to improve the quality of research

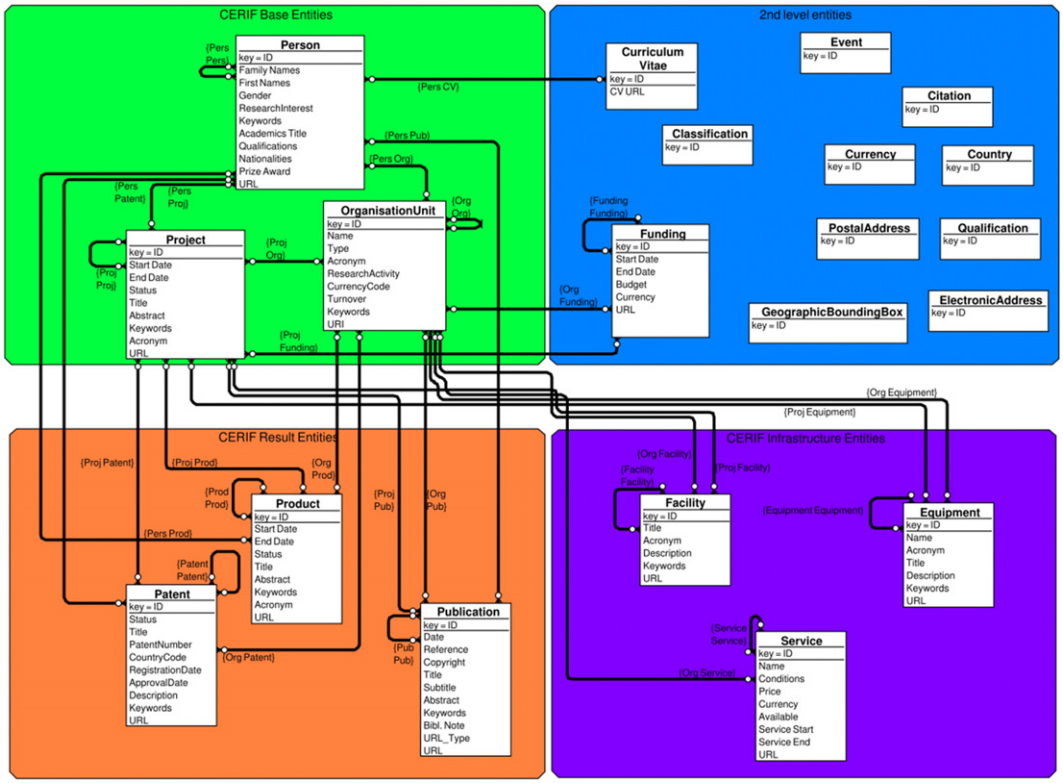

Fig. 3. CERIF Entities and their relationship.

CERIF				
Entities	Attributes	Relationship	Language Entities	Classification schemes
293	1814	665	76	68

Fig. 4. CERIF model metrics.

RCD				
Entities	Attributes	Relationship	Specialization	Classification schemes
7	58	48	77	19

Fig. 5. RCD model metrics.

information before integration into RIS. We report positive results of the application of our framework for sample publication data (detailed information can be found in the work of [2]).

The framework further sketches the impact of the German research information standard RCD on data quality. Our results show that data quality is to some extent contingent on standard adoption and that data

Fig. 6. Comparison conditions.

quality will likely improve as a result. A standardized data model, such as RCD, is an essential prerequisite for achieving data governance in terms of monitoring and strengthening data management in institutions. This makes it possible to introduce and permanently guarantee quality in institutions as an overall target for research information.

2.3. CERIF data model

Using the CERIF data model or a CERIF compliant IT solution for current research information systems (CRIS) is a European Union recommendation to the member states [13]. The organization euroCRIS is committed to the development and distribution of the CERIF standard on data formats for research information. The uniform European format CERIF represents information about the entire research process (such as *person*, *organizations*, *projects*, *publications*, *patents*, *service*, *facility* and *equipment*, etc.). CERIF is a relational database model available as SQL scripts based on a common Entity Relationship Model (ERM) [13]. The ERM of the CERIF 1.6 release contains objects where attributes are linked by relationships. The CERIF data model differs in *base*, *result*, *link*, *infrastructure*, and 2nd *level* of entities. Further details on the CERIF data model can be found in [8–10,12]. Therefore, the CERIF model is conceptualized with its conceptual structure of colors as shown in Figure 3 below.

3. Mapping RCD and CERIF

This section is intended to provide a meaningful mapping recommendation for the elements of the RCD data model and CERIF data model to simplify use of the RCD in existing CERIF-compliant systems. RCD and CERIF essentially include XML Schema, data model, and semantics specifications for the exchange of research information. Figure 4 and 5 below list and explain the metrics of RCD and CERIF.

RCD and CERIF are translated into classes and relationships in ontology and in elements of an XML schema. To make the implementation understandable, it is therefore necessary to record and manage the links between the content definitions and the various data models. The mapping of RCD base data to CERIF is straightforward and much of the elements mentioned in the RCD basic data model are also present in CERIF. This means that RCD extends the existing CERIF elements by further attributes but also adds missing, e.g. the aspect of promoting young talent and spin-offs. CERIF data model captures the data in full detail; the RCD aggregate data model instead focuses on an aggregated presentation of research information for reporting. Linking the RCD with the already defined concepts in CERIF appears to make sense through the investigation. These results were agreed with experts in this field at the workshop on "Using the RCD Data Model as the Standard for Processing Research Information and Comparison with CERIF" organized by RCD team.

For the conditions of the comparison for each area or objects of the RCD or tables of the CERIF we have selected two different colors to better understand them. This is illustrated as follows in Fig. 6.

Table 1
Mapping (basis data) between RCD and CERIF

RCD			CERIF
Area	**Attribute**	**Entity**	**Mapping**
(Beschäftigte) (Doktorand) **Person**	(hat Organisationseinheit) has organizational unit	Person	cfPers cfPers_OrgUnit cfOrgUnit (The CERIF classification scheme "Person Organization Roles" should be used.)
	(Geburtsdatum) birthdate	Person	cfPers.cfBirthdate
	(Geschlecht) gender	Person	cfPers.cfGender
	(Name) name	Person	cfPers cfPersName_Pers cfPersName (The CERIF classification scheme "Person Names" should be used.)
	(Staatsangehörigkeit) nationality	Person	cfPers cfPers_Country cfCountry
	(hat Promotionsberechtigung aus) has doctoral eligibility from	Person	cfPers cfPers_Country cfCountry
	(Qualifikation) qualification	Person	cfPers cfPers_Qual cfQual cfQualTitle
	(Altersgruppe) age group (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, the attribute should be removed from the RCD base model because it depends on the birthday and survey date.)	Person	not modeled
	(hat Qualifikationsverfahren) has qualification procedure (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdPers_QualProc.)	Person	not modeled

Table 1 (Continued).

	(hat Beschäftigung) has employment (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdPers_Employ.)	Employee (In RCD version 1.0, this entity semantically extends the RCD entity "Person".)	not modeled
	(hat Beschäftigten) has employee (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdPers_Employ.)	Employment	not modeled
	(hat Organisationseinheit) has organizational unit (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdOrgUnit_Employ.)	Employment	not modeled
	(hat Finanzierungsform) has form of financing (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdEmploy_Class.)	Employment	not modeled
	(hat Befristung) has limited term contract (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdEmploy_Class.)	Employment	not modeled
	(hat Personalkategorie) has personnel category (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdEmploy_Class.)	Employment	not modeled
	(hat Tätigkeitsart) has type of work (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdEmploy_Class.)	Employment	not modeled

Table 1 (Continued).

	(hat Fach) has subject (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdEmploy_Class.)	Employment	not modeled
	(hat Forschungsfeld) has research area (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdEmploy_Class.)	Employment	not modeled
	(hat Besoldung) has salary (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdEmploy_Class.)	Professorial Employment (In RCD version 1.0, this entity semantically extends the RCD entity "Employment".)	not modeled
	(hat Professurbezeichnung) has professorial title (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdEmploy_Class.)	Professorial Employment (In RCD version 1.0, this entity semantically extends the RCD entity "Employment".)	not modeled
	(gemeinsame Berufung) shared vocation (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdOrgUnit_Employ.)	Professorial Employment (In RCD version 1.0, this entity semantically extends the RCD entity "Employment".)	not modeled
	(hat Person) has person (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdPers_QualProc.)	Qualification Procedure	not modeled
	(Abschlusszeitpunkt) date of completion	Qualification Procedure	not modeled
	(Altersgruppe) age group at completion (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, the attribute should be removed from the RCD base model because it depends on the birthday and the date of completion.)	Qualification Procedure	not modeled

Table 1 (Continued).

	(Start der Promotion nach Annahme als Doktorakt am Fachbereich/an der Fakultät der titelvergebenden Einrichtungen) start of the doctorate after acceptance as a doctoral act at the faculty/department of the institution awarding the title	Doctorate (In RCD version 1.0, this entity semantically extends the RCD entity "Qualification Procedure".)	not modeled
	(Start der Promotion nach der Betreuungsvereinbarung/ des Arbeitsvertrags an der nicht-titelvergebende Einrichtungen) start of the doctorate after the supervision agreement/ employment contract at the non-title awarding institution	Doctorate (In RCD version 1.0, this entity semantically extends the RCD entity "Qualification Procedure".)	not modeled
	(hat Erstbetreuer) has supervisor (In RCD version 1.0, this is a separate attribute within the documentation but not modeled within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdPers_QualProc.)	Doctorate (In RCD version 1.0, this entity semantically extends the RCD entity "Qualification Procedure".)	not modeled
	hat Kooperationspartner has cooperation partner (In RCD version 1.0, this is a separate attribute within the documentation but not modeled within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdQualProc_Class.)	Doctorate (In RCD version 1.0, this entity semantically extends the RCD entity "Qualification Procedure".)	not modeled
	Kooperationspartner bei kooperativer Promotion cooperation partner in cooperative doctorate (In RCD version 1.0, this is a separate attribute within the documentation but not modeled within the XML Schema.)	Doctorate (In RCD version 1.0, this entity semantically extends the RCD entity "Qualification Procedure".)	not modeled

Table 1 (Continued).

	(hat Strukturiertes Promotionsprogramm) has structured doctoral program (In RCD version 1.0, this is a separate attribute within the documentation but not modeled within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdQualProc_DocProg.)	Doctorate (In RCD version 1.0, this entity semantically extends the RCD entity "Qualification Procedure".)	not modeled
	(In RCD version 1.0 no further attributes are specified for this entity; indicated for completeness only.)	Habilitation (In RCD version 1.0, this entity semantically extends the RCD entity "Qualification Procedure".)	not modeled
(Drittmittelprojekt) **Third-party funded project**	(hat Organisationseinheit) has organizational unit	Third-party funded project	cfProj cfProj_OrgUnit cfOrgUnit cfOrgUnitName (The CERIF classification scheme "Organization Project Engagements" should be used.)
	(Titel des Projekts) title of project	Third-party funded project	cfProj cfProjTitle
	(Projektbeginn) start of project	Third-party funded project	cfProj.cfStartDate (In CERIF version 1.6, this mapping is deprecated; no viable alternative specified by euroCRIS.)
	(Projektende) end of project	Third-party funded project	cfProj.cfEndDate (In CERIF version 1.6, this mapping is deprecated; no viable alternative specified by euroCRIS.)
	(hat übergeordnetes Projekt) has parent project	Third-party funded project	cfProj_Proj (The classification term "Part" of the CERIF classification scheme "Activity Structure" should be used.)
	(Förderkennzeichen) funding number (In RCD version 1.0, this is a separate attribute within the documentation but not modeled within the XML Schema; in later versions, the CERIF concept of federated identifier should be used.)	Third-party funded project	not modeled
	(Bewilligungssumme) amount of funding	Third-party funded project	cfProj_Fund::cfAmount
	(Drittmitteleinnahmen) third party funding cameralistic	Third-party funded project	not modeled

Table 1 (Continued).

	(Drittmittelerträge) third party funding double entry	Third-party funded project	not modeled
	(Koordinationsrolle) coordinating role (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, the attribute should be removed from the RCD base model because it is calculable via the "coordinating institution" of the RCD entity "Third-party funded project".)	Third-party funded project	not modeled
	(KoordinatorEinrichtung) coordinating institution	Third-party funded project	cfProj_OrgUnit (The classification term "Coordinator" of the CERIF classification scheme "Organization Project Engagements" should be used.)
	(hat Mittelgeber) has founder	Third-party funded project	cfProj_Class (In CERIF version 1.6, this linking entity is used to store different classification schemes and terms.)
	(hat Fach) has subject	Third-party funded project	cfProj_Class (In CERIF version 1.6, this linking entity is used to store different classification schemes and terms.)
	(hat Forschungsfeld) has research area	Third-party funded project	cfProj_Class (In CERIF version 1.6, this linking entity is used to store different classification schemes and terms.)
(Strukturiertes Promotionsprogramm) **Structured doctoral program**	(hat Organisationseinheit) has organizational unit (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent linking data structure like in CERIF should be used, e. g. rcdOrgUnit_DocProg.)	Structured doctoral program	not modeled
	(Titel des Promotionsprogramms) title of the doctoral program (In RCD version 1.0, this attribute is not translatable; in later versions an equivalent localization data structure like in CERIF should be used, e. g. rcdDocProgTitle.)	Structured doctoral program	not modeled
	(hat Finanzierungsform) has form of financing (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdDocProg _Class.)	Structured doctoral program	not modeled

Table 1 (Continued).

	(hat Sprecher) has speaker (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent linking data structure like in CERIF should be used, e. g. rcdPers_DocProg.)	Structured doctoral program	not modeled
	(hat laufende Promotionen) has ongoing doctorate (In RCD version 1.0, this is a separate attribute within the documentation but not modeled within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdQualProc_DocProg.)	Structured doctoral program	not modeled
	(beteiligte Institutionen) participating institution (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent linking data structure like in CERIF should be used, e. g. rcdOrgUnit_DocProg.)	Structured doctoral program	not modeled
	(Länder der beteiligten Institutionen) country of participating institution (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, the attribute should be removed from the RCD base model because it is calculable via the "participating institution" of the RCD entity "Structured doctoral program".)	Structured doctoral program	not modeled
	(hat Fach) has subject (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdDocProg_Class.)	Structured doctoral program	not modeled
	(hat Forschungsfeld) has research area (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdDocProg_Class.)	Structured doctoral program	not modeled

Table 1 (Continued).

(Patente und Ausgründungen) **Patents and Spin-offs**	(hat Organisationseinheit) has organizational unit	Patent	cfResPat cfOrgUnit_ResPat cfOrgUnit cfOrgUnitName (The CERIF classification schemes "Organization Output Contributions" and "Organization Output Roles" should be used.)
	(Titel des Patents) title of patent	Patent	cfResPat cfResPatTitle
	(Veröffentlichungsnummer) publication number	Patent	cfResPat.cfPatentNum
	(Datum der prioritätsbegründenden Erstanmeldung) date of initial registration	Patent	cfResPat cfResPat.cfRegistrDate
	(Datum der Erteilung) date of approvement	Patent	cfResPat.cfApprovDate
	(Anzahl der Patentfamilien) number of patent family	Patent	not modeled
	(hat Erfinder) has inventor	Patent	cfPers_ResPat (The classification term "Patentee" of the CERIF classification scheme "Person Output Contributions" should be used.)
	(hat Fach) has subject	Patent	cfResPat_Class (In CERIF version 1.6, this linking entity is used to store different classification schemes and terms.)
	(hat Forschungsfeld) has research area	Patent	cfResPat_Class (In CERIF version 1.6, this linking entity is used to store different classification schemes and terms.)
	(hat Organisationseinheit) has organizational unit (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent linking data structure like in CERIF should be used, e. g. rcdOrgUnit_SpinOff.)	Spin-off	not modeled
	(Name der Ausgründung) name of spin-off (In RCD version 1.0, this attribute is not translatable; in later versions an equivalent localization data structure like in CERIF should be used, e. g. rcdSpinOffName.)	Spin-off	not modeled
	(Datum der Ausgründung) date of spin-off	Spin-off	not modeled

Table 1 (Continued).

	(hat Fach) has subject (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdSpinOff _Class.)	Spin-off	not modeled
	(hat Forschungsfeld) has research area (In RCD version 1.0, this is a separate attribute within the XML Schema; in later versions, an equivalent classification data structure like in CERIF should be used, e. g. rcdSpinOff _Class.)	Spin-off	not modeled
(Publikation) **Publication**	(hat Organisationseinheit) has organizational unit	Publication	cfResPubl cfOrgUnit_ResPubl cfOrgUnit cfOrgUnitName (The CERIF classification schemes "Organization Output Contributions" and "Organization Output Roles" should be used.)
	(Titel der Publikation) title of publication	Publication	cfResPubl cfResPublTitle
	(Veröffentlichungsjahr) publication year	Publication	cfResPubl.cfResPublDate
	(Förderkennzeichen) funding number (In RCD version 1.0, this is a separate attribute within the documentation but not modeled within the XML Schema; in later versions, the CERIF concept of federated identifier should be used.)	Publication	not modeled
	(Identifier) identifier	Publication	cfFedId cfFedId_Class (The CERIF classification scheme "Identifier Types" should be used.) cfDCResourceIdentifier (In CERIF version 1.6, the complete Dublin Core part is deprecated; the concept of federated identifier should be used instead.)
	(Ressource) resource	Publication	cfDCResourceType (In CERIF version 1.6, the complete Dublin Core part is deprecated; no viable alternative specified by euroCRIS.)
	(Sprachcode) language code	Publication	cfDCLanguage (In CERIF version 1.6, the complete Dublin Core part is deprecated; no viable alternative specified by euroCRIS.)
	(Zugangsrechte) access rights	Publication	cfDCRightsMMAccessRights (In CERIF version 1.6, the complete Dublin Core part is deprecated; no viable alternative specified by euroCRIS.)
	(hat Verlag) has publisher	Publication	cfOrgUnit_ResPubl (The classification term "Publisher" of the CERIF classification scheme "Organization Output Roles" should be used.)

Table 1 (Continued).

	(hat Schöpfer) has creator	Publication	cfPers_ResPubl (The classification term "Creator" or "Author" of the CERIF classification scheme "Person Output Contributions" should be used.)
	(hat Format) has format	Publication	cfResPubl cfResPubl_Event cfEvent_Class (The classification term "Conference" of the CERIF classification scheme "Event Types " should be used.) cfResPubl.cfVol cfResPubl.cfIssue cfResPubl.cfStartPage cfResPubl.cfEndPage
	(hat Quelle) has source item	Publication	not modeled
	(hat Qualifikationsschrift) has thesis	Publication	cfResPubl_Class (In CERIF version 1.6, this linking entity is used to store different classification schemes and terms.)
	(hat Unterstützung durch) has support by	Publication	not modeled
	(hat Forschungsinfrastruktur) has research infrastructure	Publication	cfResPubl_Equip (In CERIF version 1.6, this linking entity is used to store different classification schemes and terms.)
	(hat Dokumenttyp) has document type	Publication	cfResPubl_Class (In CERIF version 1.6, this linking entity is used to store different classification schemes and terms.)
	(hat Publikationstyp) has publication type	Publication	cfResPubl_Class (In CERIF version 1.6, this linking entity is used to store different classification schemes and terms.)
	(ist Peer-reviewed) is peer-reviewed	Publication	cfResPubl_Class (The CERIF classification scheme “Peer Reviewes” should be used.)
	(hat Fach) has subject	Publication	cfResPubl_Class (In CERIF version 1.6, this linking entity is used to store different classification schemes and terms.)
	(hat Forschungsfeld) has research area	Publication	cfResPubl_Class (In CERIF version 1.6, this linking entity is used to store different classification schemes and terms.)
(Forschungsinfrastruktur) **Research infrastructure**	(hat Organisationseinheit) has organizational unit	Research infrastructure	cfEquip cfOrgUnit_Equip cfOrgUnit cfOrgUnitName (The CERIF classification schemes “Organization Research Infrastructure Roles” should be used.)

Table 1 (Continued).

	(Bezeichnung der Forschungsinfrastruktur) name of research infrastructure	Research infrastructure	cfEquip cfEquipName
	(Beschreibung der Forschungsinfrastruktur) description of research infrastructure	Research infrastructure	cfEquip cfEquipDescr
	(hat Publikation) has publication	Research infrastructure	cfResPubl_Equip
	(hat Typ der Forschungsinfrastruktur) has type of research infrastructure	Research infrastructure	cfEquip_Class (In CERIF version 1.6, this linking entity is used to store different classification schemes and terms.)
	(hat Art der Forschungsinfrastruktur) has kind of research infrastructure	Research infrastructure	cfEquip_Class (In CERIF version 1.6, this linking entity is used to store different classification schemes and terms.)
	(hat Zugangsart) has type of access	Research infrastructure	cfEquip_Class (In CERIF version 1.6, this linking entity is used to store different classification schemes and terms.)
	(hat Nutzungsintensität) has use intensity	Research infrastructure	not modeled
	(hat Betreiber) has operator	Research infrastructure	cfOrgUnit_Equip (The classification term "User" of the CERIF classification scheme "Organization Research Infrastructure Roles" should be used.)
	(hat Koordinator) has coordinator	Research infrastructure	cfOrgUnit_Equip (In CERIF version 1.6, this linking entity is used to store different classification schemes and terms.)
	(hat Betriebspersonal) has staff	Research infrastructure	cfPers_Equip (The classification term "Staff" of the CERIF classification scheme "Person Research Infrastructure Roles" should be used.)

Table 2
Extract of Mappings (semantics) between RCD and CERIF.

RCD			CERIF	
Area	**Entity/Attribute**	**Examples of Classification Terms**	**Entity**	**Examples of Classification Terms**
Organisationseinheit **Organizational Unit**	not re-modeled (In RCD version 1.0, the CERIF linking entity is reused.)	not re-specified (In RCD version 1.0, the CERIF classification schemes and terms are reused.)	**cfOrgUnit_Class**	Academic Institute; Company; Research Institute; University; etc.
(Beschäftigte) (Doktorand) **Person**	Person/ has organizational unit (In RCD version 1.0, the CERIF linking entity is reused.)	not re-specified (In RCD version 1.0, the CERIF classification schemes and terms are reused.)	cfPers_OrgUnit	Consultant; Research Fellow; Lecturer; etc.
	Person/ name (In RCD version 1.0, the CERIF linking entity is reused.)	not re-specified (In RCD version 1.0, the CERIF classification schemes and terms are reused.)	cfPersName_Pers	Initials; Passport Name; etc.
	Person/ nationality (In RCD version 1.0, the CERIF linking entity is reused.)	Nationality (In RCD version 1.0, a separate classification scheme with a new classification term is specified.)	cfPers_Country	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Person/ has doctoral eligibility from (In RCD version 1.0, the CERIF linking entity is reused.)	Doctoral Eligibility (In RCD version 1.0, a separate classification scheme with a new classification term is specified.)	cfPers_Country	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Person/ qualification (In RCD version 1.0, the CERIF linking entity is reused.)	Doctorate; Habilitation (In RCD version 1.0, a separate classification scheme with new classification terms is specified.)	cfPers_Qual	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Person/ has qualification procedure	Doctorate; Habilitation (In RCD version 1.0, a separate classification scheme with new classification terms is specified.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)

Table 2 (Continued).

	Employee/ has employment	not specified (In RCD version 1.0, the specification of a uniform classification scheme is not available and therefore still in progress.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Employment/ has employee	not specified (In RCD version 1.0, the specification of a uniform classification scheme is not available and therefore still in progress.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Employment/ has organizational unit	not specified (In RCD version 1.0, the specification of a uniform classification scheme is not available and therefore still in progress.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Employment/ has form of financing	External Funds; Main Funds; etc. (In RCD version 1.0, a separate classification scheme with new classification terms is specified.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Employment/ has limited term contract	Limited; Unlimited (In RCD version 1.0, a separate classification scheme with new classification terms is specified.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Employment/ has personnel category	Administrative Staff; Professor; Scientific and Artistic Staff; etc. (In RCD version 1.0, a separate classification scheme with new classification terms is specified.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Employment/ has type of work	Full-time; Part-time (In RCD version 1.0, a separate classification scheme with new classification terms is specified.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)

Table 2 (Continued).

	Employment/ has subject	Mathematics; Physics; etc. (In RCD version 1.0, a separate classification scheme with new classification terms based on the subject classification of the German Federal Statistical Office (Destatis) is specified.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Employment/ has research area	not specified (In RCD version 1.0, the specification of a uniform classification scheme is not available and therefore still in progress.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Professorial Employment/ has salary	B1; …; B11; C2; W1; …; W3; (In RCD version 1.0, a separate classification scheme with new classification terms is specified.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Professorial Employment/ has professorial title	Professorship; Junior Professorship; etc. (In RCD version 1.0, a separate classification scheme with new classification terms is specified.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Professorial Employment/ shared vocation	not specified (In RCD version 1.0, the specification of a uniform classification scheme is not available and therefore still in progress.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Qualification Procedure/ has person	Doctorate; Habilitation (In RCD version 1.0, a separate classification scheme with new classification terms is specified.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Doctorate/ has supervisor	not specified (In RCD version 1.0, the specification of a uniform classification scheme is not available and therefore still in progress.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)

Table 2 (Continued).

	Doctorate/ has cooperation partner	University; University of Applied Sciences; etc. (In RCD version 1.0, a separate classification scheme with new classification terms is specified.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Doctorate/ has structured doctoral program	not specified (In RCD version 1.0, the specification of a uniform classification scheme is not available and therefore still in progress.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
(Drittmittelprojekt) **Third-part of funded project**	Third-party funded project/ has organizational unit (In RCD version 1.0, the CERIF linking entity is reused.)	not re-specified (In RCD version 1.0, the CERIF classification schemes and terms are reused.)	cfProj_OrgUnit	Applicant; Coordinator; Participant; etc.
	Third-party funded project/ has founder (In RCD version 1.0, the CERIF linking entity is reused.)	EU; Federation Federal States, etc. (In RCD version 1.0, a separate classification scheme with new classification terms is specified.)	cfProj_Class	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Third-party funded project/ has subject (In RCD version 1.0, the CERIF linking entity is reused.)	Mathematics; Physics; etc. (In RCD version 1.0, a separate classification scheme with new classification terms based on the subject classification of the German Federal Statistical Office (Destatis) is specified.)	cfProj_Class	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Third-party funded project/ has research area (In RCD version 1.0, the CERIF linking entity is reused.)	not specified (In RCD version 1.0, the specification of a uniform classification scheme is not available and therefore still in progress.)	cfProj_Class	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)

Table 2 (Continued).

(Strukturiertes Promotionsprogramm) **Structured doctoral program**	Structured doctoral program/ has organizational unit	not specified (In RCD version 1.0, the specification of a uniform classification scheme is not available and therefore still in progress.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Structured doctoral program/ has form of financing	External Funds; Main Funds; etc. (In RCD version 1.0, a separate classification scheme with new classification terms is specified.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Structured doctoral program/ has speaker	not specified (In RCD the specification of a uniform classification scheme is currently still in progress.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Structured doctoral program/ has ongoing doctorates	not specified (In RCD the specification of a uniform classification scheme is currently still in progress.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Structured doctoral program/ participating institutions	not specified (In RCD the specification of a uniform classification scheme is currently still in progress.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Structured doctoral program/ has subject	Mathematics; Physics; etc. (In RCD version 1.0, a separate classification scheme with new classification terms based on the subject classification of the German Federal Statistical Office (Destatis) is specified.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Structured doctoral program/ has research area	not specified (In RCD version 1.0, the specification of a uniform classification scheme is not available and therefore still in progress.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)

Table 2 (Continued).

(Patente und Ausgründungen) **Patents and Spin-offs**	Patent/ has organizational unit (In RCD version 1.0, the CERIF linking entity is reused.)	not re-specified (In RCD version 1.0, the CERIF classification schemes and terms are reused.)	cfOrgUnit_ResPat	Author Institution; Publisher Institution; etc.
	Patent/ has subject (In RCD version 1.0, the CERIF linking entity is reused.)	Mathematics; Physics; etc. (In RCD version 1.0, a separate classification scheme with new classification terms based on the subject classification of the German Federal Statistical Office (Destatis) is specified.)	cfRestPat_Class	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Patent/ has research area (In RCD version 1.0, the CERIF linking entity is reused.)	not specified (In RCD version 1.0, the specification of a uniform classification scheme is not available and therefore still in progress.)	cfRestPat_Class	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Spin-off/ has organizational unit	not specified (In RCD version 1.0, the specification of a uniform classification scheme is not available and therefore still in progress.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Spin-off/ has subject	Mathematics; Physics; etc. (In RCD version 1.0, a separate classification scheme with new classification terms based on the subject classification of the German Federal Statistical Office (Destatis) is specified.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Spin-off/ has research area	not specified (In RCD version 1.0, the specification of a uniform classification scheme is not available and therefore still in progress.)	not modeled	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
(Publikation) **Publication**	Publication/ has organizational unit (In RCD version 1.0, the CERIF linking entity is reused.)	not re-specified (In RCD version 1.0, the CERIF classification schemes and terms are reused.)	cfOrgUnit_ResPubl	Author Institution; Publisher Institution; etc.

Table 2 (Continued).

	Publication/ identifier (In RCD version 1.0, the CERIF linking entity is reused.)	not re-specified (In RCD version 1.0, the CERIF classification schemes and terms are reused.)	cfFedId cfFedId_Class	DOI; ISBN; etc.
	Publication/ has thesis (In RCD version 1.0, the CERIF linking entity is reused.)	Doctoral Thesis; Habilitation Thesis (In RCD version 1.0, a separate classification scheme with new classification terms is specified.)	cfResPubl_Class	Doctoral Thesis
	Publication/ has research infrastructure (In RCD version 1.0, the CERIF linking entity is reused.)	not specified (In RCD version 1.0, the specification of a uniform classification scheme is not available and therefore still in progress but differs from existing schemes in CERIF version 1.6.)	cfResPubl_Equip	Usage Document; Access Policy Document; etc.
	Publication/ has document type (In RCD version 1.0, the CERIF linking entity is reused.)	Editorial; Meeting Abstract; Review; etc. (In RCD version 1.0, a separate classification scheme with new classification terms is specified.)	cfResPubl_Class	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Publication/ has publication type (In RCD version 1.0, the CERIF linking entity is reused.)	Article; Book; ePaper; etc. (In RCD version 1.0, a separate classification scheme with new classification terms is specified.)	cfResPubl_Class	Article; Book; Manual; etc.
	Publication/ is peer-reviewed (In RCD version 1.0, the CERIF linking entity is reused.)	not re-specified (In RCD version 1.0, the CERIF classification schemes and terms are reused.)	cfResPubl_Class	Yes; No; Unknown
	Publication/ has subject (In RCD version 1.0, the CERIF linking entity is reused.)	Mathematics; Physics; etc. (In RCD version 1.0, a separate classification scheme with new classification terms based on the subject classification of the German Federal Statistical Office (Destatis) is specified.)	cfResPubl_Class	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)

Table 2 (Continued).

	Publication/ has research area (In RCD version 1.0, the CERIF linking entity is reused.)	not specified (In RCD version 1.0, the specification of a uniform classification scheme is not available and therefore still in progress.)	cfResPubl_Class	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
(Forschungsinfrastruktur) **Research infrastructure**	Research infrastructure/ has organizational unit (In RCD version 1.0, the CERIF linking entity is reused.)	not re-specified (In RCD version 1.0, the CERIF classification schemes and terms are reused.)	cfOrgUnit_Equip	Owner; User; etc.
	Research infrastructure/ has publication (In RCD version 1.0, the CERIF linking entity is reused.)	not specified (In RCD version 1.0, the specification of a uniform classification scheme is not available and therefore still in progress but differs from existing schemes in CERIF version 1.6.)	cfResPubl_Equip	Usage Document; Access Policy Document; etc.
	Research infrastructure/ has type of research infrastructure (In RCD version 1.0, the CERIF linking entity is reused.)	Large Equipment and Instrument; Information and communication infrastructure, etc. (In RCD version 1.0, a separate classification scheme with new classification terms is specified.)	cfEquip_Class	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)
	Research infrastructure/ has kind of research infrastructure (In RCD version 1.0, the CERIF linking entity is reused.)	Local; Distributed; Virtual; (In RCD version 1.0, a separate classification scheme with new classification terms is specified.)	cfEquip_Class	Cluster; Single Location; Network; Virtual; etc.
	Research infrastructure/ has type of access (In RCD version 1.0, the CERIF linking entity is reused.)	Open Access; Shared Access; User Access (In RCD version 1.0, a separate classification scheme with new classification terms is specified.)	cfEquip_Class	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)

Table 2 (Continued).

	Research infrastructure/ has coordinator (In RCD version 1.0, the CERIF linking entity is reused.)	Coordinator (In RCD version 1.0, a separate classification scheme with a new classification term is specified.)	cfOrgUnit_Equip	not specified (In CERIF version 1.6, no equivalent classification scheme exists.)

Our mapping looks at two categories:

(1) Comparison of the basic data of RCD with CERIF
(2) Comparison of the semantics of RCD with CERIF.

The results of these categories between RCD and CERIF are shown in Tables 1 and 2.

The results of a mapping (basis data, semantics and link entities) of RCD and CERIF show that the elements of the RCD are mappable to the CERIF data model and have a common vocabulary, and that these two standards allow the exchange between different research information systems. The RCD and CERIF formats provide models to structure the research area into relevant objects and their relationships, while allowing their high-quality integration and interoperability into the RIS in a common format. These are not only beneficial for information management, but also for analyzing data and accessing data, information and knowledge. In addition, the two standards provide clarity in the collection of research information and to reduce the administrative burden and to improve the data quality of the research information and to represent sound and transparent decisions.

4. Conclusion

Summing up one can say that the two data models RCD and CERIF support the interoperability of research information in different formats, e. g. exchange, merge, sharing and mapping of data. CERIF and RCD can be considered as a basic data format and thus increase the flexibility of RIS. However, for better integration and compatibility between CERIF and RCD, the changes outlined above should be implemented in RCD version 2.0.

Acknowledgements

This work has been funded by the German Center for Higher Education Research and Science Studies (DZHW) and by the German Federal Ministry of Education and Research (BMBF) in the context of the project “Helpdesk to facilitate implementation of the Research Core Dataset” (https://kerndatensatz-forschung.de/) (project period: 2017–2019; grant number: KDS2016).

References

[1] O. Azeroual, G. Saake and J. Wastl, Data measurement in research information systems: metrics for the evaluation of data quality, *Scientometrics* **115**(3) (2018), 1271–1290. doi:10.1007/s11192-018-2735-5.

[2] O. Azeroual, J. Schöpfel and D. Ivanovic, Influence of Information Quality via Implemented German RCD Standards in Research Information Systems. *Data* **5**(2) (**2020**), 30. doi: 10.3390/data5020030.
[3] S. Biesenbender and S. Hornbostel, The Research Core Dataset for the German science system: Challenges, processes and principles of a contested standardization project, *Scientometrics* **106**(2) (2016), 837–847. doi:10.1007/s11192-015-1816-y.
[4] S. Biesenbender and S. Hornbostel, The Research Core Dataset for the German science system: Developing standards for an integrated management of research information, *Scientometrics* **108**(1) (2016), 401–412. doi:10.1007/s11192-016-1909-2.
[5] Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), KDSF Technische Datenmodelle. URL: http://www.kerndatensatz-forschung.de/version1/technisches_datenmodell/. Retrieved April 2, 2019.
[6] German Council of Science and Humanities, Recommendations on a research core dataset (Drs. 2855-13), 2013, Berlin, Germany. URL: http://www.wissenschaftsrat.de/download/archiv/2855-13.pdf. Retrieved April 12, 2019.
[7] German Council of Science and Humanities, Recommendations on a research core dataset (Drs. 5066-16), 2016, Berlin, Germany. URL: http://www.wissenschaftsrat.de/download/archiv/5066-16.pdf. Retrieved April 12, 2019.
[8] B. Jörg, K. Jeffery, J. Dvořák, N. Houssos, A. Asserson and G. van Grootel, CERIF 1.3 Full Data Model (FDM) - Introduction and Specification, January 2012. URL: http://www.ict.nsc.ru/xmlui/handle/ICT/1865. Retrieved April 14, 2019.
[9] B. Jörg, CERIF: The common European research information format model, *Data Sc. J.* **9**: (2010), 24–31. doi:10.2481/dsj.CRIS4.
[10] B. Jörg, O. Krast, K. Jeffery and G. van Grootel, CERIF 2008 – 1.0 XML Data Exchange Format Specification, May 2009. URL: https://www.eurocris.org/cerif/downloads/cerif-2008. Retrieved April 10, 2019.
[11] Institut für Forschungsinformation und Qualitätssicherung (iFQ), Spezifikation des Kerndatensatz-Forschung, Berlin, Germany 2015. URL: http://kerndatensatzforschungde/version1/Spezifikation_KDSF_v1.pdf. Retrieved April 4, 2019.
[12] D. Ivanovic, D. Surla and M. Racković, A CERIF data model extension for evaluation and quantitative expression of scientific research results, *Scientometrics* **86**(1) (2011), 155–172. doi:10.1007/s11192-010-0228-2.
[13] L. Lezcano, B. Jörg and M.A. Sicilia, Modeling the context of scientific information: Mapping VIVO and CERIF. in: *Advanced Information Systems Engineering CAISE 2010 International Workshops*, M. Bajec and J. (Hrsg.) Eder (eds), Vol. 112, Gdansk, Poland, 2012, pp. 123–129. doi:10.1007/978-3-642-31069-0_11.