



Einführung in das CERIF-Datenmodell und Vergleich mit dem Datenmodell des Kern- datensatz Forschung (KDSF)

Helpdesk Kerndatensatz Forschung

31. März 2020

Inhaltsverzeichnis

1	Einleitung	3
2	Was ist CERIF?	3
	2.1 CERIF-Komponenten	3
	2.2 Das konzeptionelle Datenmodell von CERIF	4
3	Das KDSF-Basisdatenmodell	8
4	Konzeptioneller Vergleich der Architektur CERIF – KDSF	9
5	Mapping CERIF – KDSF	10
6	Referenzen	18

Das dieser Veröffentlichung zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung und aller sechszehn Bundesländer unter dem Förderkennzeichen KDS2016 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Dieses Dokument (Einführung in das CERIF-Datenmodell und Vergleich mit dem Datenmodell des Kerndatensatz Forschung), das dargestellte Mapping von Kerndatensatz Forschung und CERIF sowie das Logo des Kerndatensatz Forschung auf der Titelseite dieses Dokuments sind lizenziert unter einer [Creative Commons Namensnennung - Weitergabe unter gleichen Bedingungen 4.0 International Lizenz](https://creativecommons.org/licenses/by-sa/4.0/).



1 Einleitung

Dieses Dokument bietet eine kurze Einführung in das Common European Research Information Format – kurz CERIF. Die Architektur des konzeptionellen Datenmodells dieses europäischen Standards für Forschungsinformationen wird mit dem Datenmodell des Kerndatensatz Forschung (KDSF) strukturell verglichen. Schließlich wird ein Mapping der beiden konzeptionellen Datenmodelle vorgenommen, in dem die Übereinstimmung der jeweiligen Objekte und ihrer Eigenschaften geprüft wird.

Es richtet sich hauptsächlich an Nutzer*innen von Forschungsinformationen und Forschungsinformationssystemen, Beauftragte zur Umsetzung des Kerndatensatz Forschung, Forschungsreferent*innen, Verantwortliche in Bibliotheken und Rechenzentren an deutschen Universitäten und außeruniversitären Forschungseinrichtungen, die ein technisches Grundverständnis der Datenmodelle von CERIF und des KDSF erlangen möchten und sich für die Kompatibilität dieser beiden Standardspezifikationen interessieren.

2 Was ist CERIF?

CERIF ist ein freiwilliger Standard für die Erfassung, Vorhaltung und den Austausch von Forschungsinformationen im europäischen Forschungsraum. Er wurde ursprünglich mit Unterstützung der Europäischen Kommission entwickelt und den Mitgliedstaaten zur Verwendung empfohlen. Im Jahr 1991 wurde die erste Version von CERIF veröffentlicht. Seitdem wird der Standard fortlaufend von euroCRIS, einem internationalen gemeinnützigen Verband, der den europaweiten Wissensaustausch zu Forschungsinformationen und Forschungsinformationssystemen fördert¹, betreut und weiterentwickelt (Asserson et al., 2002). Die euroCRIS-Website bietet Zugriff auf eine Reihe von Materialien (z.B. Tutorials und Dokumentation) zu CERIF. Aktuell wird die Nutzung der Versionen 1.5. und 1.6. empfohlen.

CERIF zielt darauf ab, eine Datenintegration sowie einen konsistenten und qualitativ hochwertigen Datenaustausch zwischen verschiedenen Informationssystemen (z.B. Forschungsinformationssystemen (FIS), im internationalen Raum häufig *Current Research Information Systems* (CRIS) genannt) und vergleichbaren Anwendungen, wie zum Beispiel institutionellen Repositorien, durch Standardisierung zu unterstützen und zu ermöglichen.

2.1 CERIF-Komponenten

CERIF setzt sich zusammen aus

- einem **konzeptionellen** Datenmodell, welches als CERIF Full Data Model (FDM) bezeichnet wird. Es beinhaltet eine formale, abstrakte Beschreibung der Objekte bzw. Konzepte, ihrer Eigenschaften und ihrer Beziehungen zueinander,
- einem **Austauschformat** basierend auf XML Schemata und Linked Data sowie
- einem **physischen** Datenmodell, d.h. einer relationalen Datenbank, die in Form von SQL-Skripten verfügbar ist und auf gängigen Entity-Relationship-Model-Konstrukten (ERM) basiert (Dvořák & Brasse, 2019).

Diese Einführung konzentriert sich auf das konzeptionelle CERIF-Datenmodell.

¹ <https://www.eurocris.org/>

2.2 Das konzeptionelle Datenmodell von CERIF

Das Modell beinhaltet verschiedene Typen von Objekten, so genannte Entitäten (entity types) und deren Beziehungen.

Zunächst werden einige zentrale Begriffe wie Entität, Entitätstyp, Attribute und Beziehungen erläutert (Meier, 1998):

- Eine *Entität* repräsentiert ein abstraktes Objekt der realen Welt und ist ein Element einer Menge (z.B. Personen, Projekte, Bücher, etc.).
- Ein *Entitätstyp* ist eine Klasse von Entitäten, die eindeutig durch einen Entitätstypnamen benannt ist und zu der dieselben Informationen im Wesentlichen auf dieselbe Weise gespeichert und verarbeitet werden.
- *Attribute* sind Eigenschaften oder Charakteristika, die eine Entität besitzt (wie z.B. das Geschlecht oder Geburtsdatum einer Person).
- *Beziehungen* kennzeichnen die konkrete Verknüpfung zwischen zwei beteiligten Entitäten (z.B. eine Person hat eine Publikation). Es gibt verschiedene Typen von Beziehungen. Grundsätzlich lassen sich {1:1}, {1:n} oder {1...0,*} sowie {n:m} oder {0,*...0,*} Beziehungen (Kardinalitäten) unterscheiden. Eine 1:1-Beziehung bedeutet, dass eine Entität des einen Entitätstyps mit höchstens einer Entität des anderen Entitätstyps in einer Beziehung steht und umgekehrt. Eine 1:n-Beziehung bedeutet, dass eine Entität des einen Entitätstyps beliebig vielen Entitäten zugeordnet ist und umgekehrt eine der vielen Entitäten höchstens einer anderen Entität. In der n:m-Beziehung kann jede Entität mit beliebig vielen weiteren Entitäten in einer Beziehung stehen und umgekehrt.

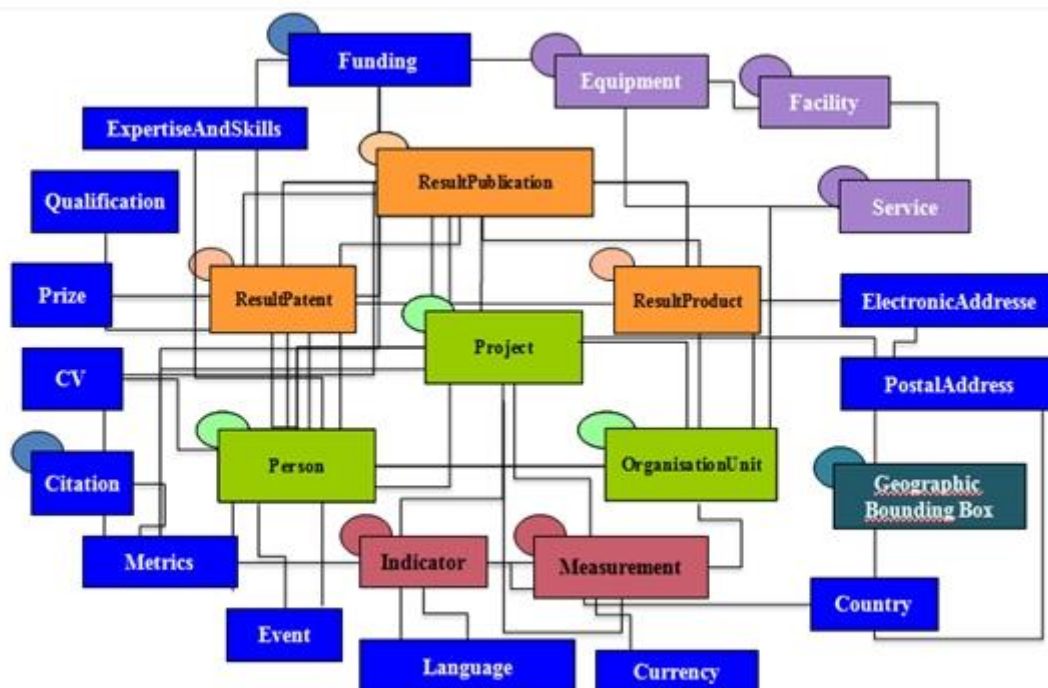


Abbildung 1. CERIF-Entitäten und ihre Beziehungen²

² EuroCRIS (2020): Entity-Relationship-Model CERIF. Verfügbar unter <https://www.eurocris.org/cerif/main-features-cerif>. © Alle Rechte vorbehalten.

Die Kernstruktur des konzeptionellen Datenmodells von CERIF basiert auf den drei Basisentitäten Forschungsprojekt, Person und Organisationseinheit. Diese verkörpern die zentrale Funktion der Standardspezifikation, nämlich die Erfassung und Vorhaltung von Informationen über Forschungsaktivitäten in Form von Forschungsprojekten individueller und kollektiver Akteure.

Auf der konzeptuellen Ebene werden die Objekte ausgeschrieben, während sie auf der physischen Ebene, d.h. bei ihrer Implementation in einer relationalen Datenbank abgekürzt werden mit dem Prefix „cf“ (für CERIF), z.B. cfProj, cfPers, cfOrgUnit. Alle Entitäten besitzen eine Grundmenge an Eigenschaften (Attribute), die entweder objektspezifisch sind (wie birth date bei Personen oder head count bei Organisationseinheiten) oder bei allen Objekten vorkommen, wie z.B. ein Identifikator (z.B. cfProjID) und eine zeitliche Begrenzung mittels Start- und Enddatum (time stamp) (Russell, 2011).

- Base entities (grün dargestellt): Die Basisentitäten sind **Person**, **OrganisationUnit** und **Project**. Jede Basisentität ist mit sich selbst rekursiv verbunden und unterhält Beziehungen zu vielen anderen Entitäten.
- Result entities (orange): Die Ergebnisentitäten sind **ResultPublication**, **ResultPatent** und **ResultProduct**. Sie ermöglichen die Darstellung von Forschungsergebnissen.
- 2nd level entities (blau): Die Entitäten der 2. Ebene sind **Funding**, **ExpertiseAndSkills**, **Qualification**, **Prize**, **CV**, **Citation**, **Event**, **Language**, **Currency**, **Country**, **PostalAddress** und **ElectronicAddress**. Sie ermöglichen die Darstellung des Forschungsumfelds und Forschungskontextes durch die Verknüpfung von Basis- und Ergebnisentitäten. Die Verknüpfung wird mittels so genannter Verknüpfungsentitäten (link entities) hergestellt.
- Link entities: Verknüpfungsentitäten stellen die Beziehungen zwischen Basis-, Ergebnis- und Entitäten der zweiten Ebene her. Die Verknüpfungen können zwischen zwei verschiedenen Entitätstypen, wie zum Beispiel Person 'is author of' ResultPublication bzw. ResultPublication 'is funded by' Funding Programme oder rekursiv zu einer Entität zu sich selbst (OrgUnit_OrgUnit) bestehen. Beziehungen sind zeitlich definiert: Jeder Verknüpfungseintrag ist mit einem Start- und Enddatum versehen. Das ermöglicht die Abbildung zeitlich begrenzter Beziehungen zwischen und innerhalb von Entitäten. So kann eine Person Mitglied eines Projektes und einer Organisationseinheit zu unterschiedlichen Zeitpunkten sein. Anhand der Abbildung 2 werden einige beispielhafte Verknüpfungsentitäten (durch Pfeile gekennzeichnet) mit möglichen Rollen veranschaulicht.
- Infrastructure entities (violett): Die Infrastruktureinheiten sind **Equipment**, **Facility** und **Service**. Jede Infrastrukturentität ist mit sich selbst rekursiv und zusätzlich mit anderen Infrastrukturentitäten verknüpft.
- Mit der neuesten Version 1.6 hat CERIF zum einen Indicator and Measurement entities (purpurrot) eingeführt, um quantitative Messungen von Forschungsleistungen zu ermöglichen. Zum anderen wurde im Kontext von Forschungsinfrastrukturen eine Geographic Bounding Box entity (jägergrün) für die geografische Lokalisierung ergänzt (Dvořák & Brasse, 2019; Jörg, 2010; Russell, 2011; Simons, 2014).

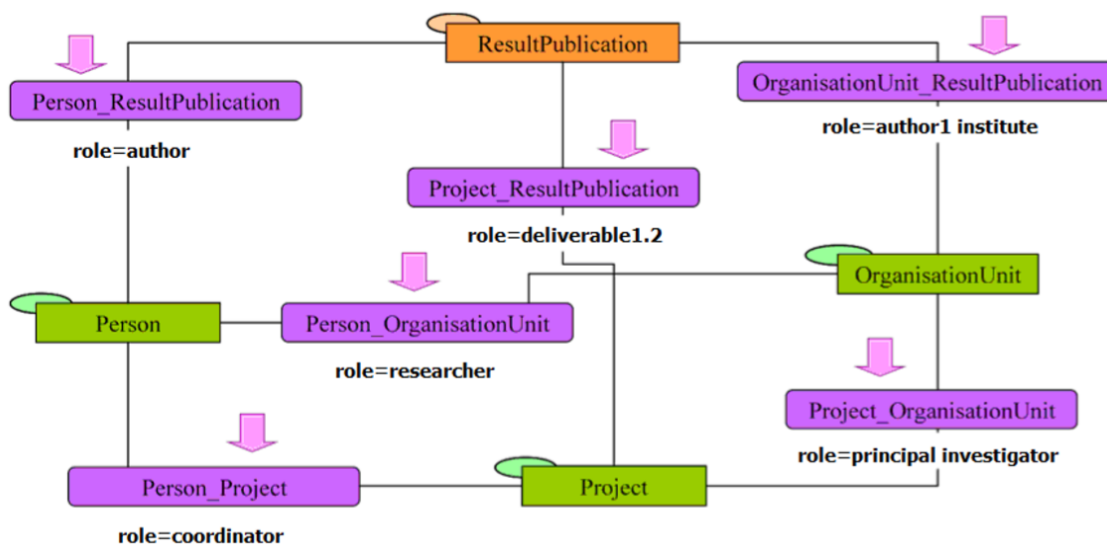


Abbildung 2. CERIF-Beispielverknüpfungsentitäten mit Rollen (Russell, 2011)

Neben den Entitäten und ihren Beziehungen enthält CERIF so genannte Features.

- Multilingual features:** Viele Forschungsinformationen müssen in mehr als einer Sprache dargestellt werden. Die Unterstützung mehrsprachiger Funktionen ist sehr wichtig, vor allem in Ländern, in denen mehrere Amtssprachen gesprochen werden. CERIF unterstützt mehrere Sprachfunktionen für names (cfName), titles (cfTitle), subtitle (cfSubtitle), descriptions (cfDescr), keywords (cfKeyw), abstracts (cfAbstract), abbreviation (cfName-Abbrev), etc. Diese sind im CERIF Datenmodell gelb markiert. Die codierte Sprache wird mit dem Attribut cfLangCode gespeichert, (z.B. cfLangCode=en, cfLangCode=de, cfLangCode=fr, etc.).
- CERIF Semantic Layer (Semantic Features):** Die CERIF Semantic Layer beinhaltet formale oder inhaltliche Schemata und Strukturen zur Klassifizierung von Entitäten in Form kontrollierter Vokabularen, wie zum Beispiel Keyword-Listen, Thesauri, Klassifikationsschemata, Taxonomien und Ontologien, welche Beziehungsarten, Rollen, Typen, Fächerlisten etc. darstellen. Die Semantic Layer ist offen und erweiterbar. Sie besteht aus den beiden Entitäten Klassifikation (cfClass) und Klassifizierungsschema (cfClassScheme). Darüber hinaus können mehrsprachige Begriffe z.B. (cfClassTerm) und Klassenbeschreibungen (cfClassDescr) dargestellt werden (Dvořák & Brasse, 2019; Russell, 2011). Die Klassifizierungsschemata können als Excel-Datei im CERIF 1.5 Vocabulary auf der euroCRIS-Website aufgerufen werden.

Die Semantic Layer und die Verknüpfungsentitäten werden als zwei zentrale Merkmale und Stärken des Datenmodells von CERIF betrachtet. Anstelle Eigenschaften als Attribute von Objekten zu konzipieren, werden diese in den Link Entities als Relationen zum Ausdruck gebracht. Auf diese Weise können beispielsweise multiple Rollen für jede Entität und damit mehr Eigenschaften abgebildet werden. Auch die Semantic Layer erlaubt eine hohe Flexibilität durch die Inklusion einer unbegrenzten Anzahl an Klassifikationsschemata pro Objekt und zur Abbildung verschiedener Klassifikationsschemata aufeinander. Gleichzeitig wird durch die Ergänzung der Semantic Layer das Datenmodell selbst vereinfacht (Dvořák & Brasse, 2019; Simons, 2014).

Nach der Vorstellung der CERIF-Entitätstypen und ihrer mehrsprachigen und semantischen Merkmale (Features), wird das CERIF-Datenmodell zur Illustration als Beispieltabelle (siehe Tabelle 1) dargestellt (in Anlehnung an Jörg, 2010).

Entität / Tabelle	Type	Attribute	Semantic Layer (CERIF Semantics)	
cfProj cfProj cfProj cfProj cfProj	base/core base/core base/core base/core base/core	cfProjId cfStartDate cfEndDate cfAcro cfURI	Classification	Classification Scheme
cfProjTitle cfProjAbstr cfProjKeyw	lang(en) lang(en) lang(en)	cfTitle cfAbstr cfKeyw		
cfProj_ResPubl cfProj_OrgUnit cfProj_Fund	link link link	cfResPublId cfOrgUnitId cfFundId	is originator of is coordinated by is funded by	cfProj_ResPubl cfProj_OrgUnit cfProj_Fund
cfResPubl cfResPubl cfResPubl cfResPubl cfResPubl cfResPubl cfResPubl cfResPubl cfResPubl cfResPubl cfResPubl cfResPubl cfResPubl	result result result result result result result result result result result result result	cfResPublId cfResPublDate cfNum cfVol cfEdition cfSeries cfIssue cfStartPage cfEndPage cfTotalPages cfISBN cfISSN cfURI		
cfResPublTitle cfResPublSubtitle cfResPublAbstr cfResPublKeyw	lang(en) lang(en) lang(en) lang(en)	cfTitle cfSubtitle cfAbstr cfKeyw		
cfPers_ResPubl cfOrgUnit_ResPubl cfProj_ResPubl cfResPubl_Event	link link link link	cfPersId cfOrgUnitId cfProjId cfEventId	is author of is publisher of is originator of is presented at	cfPers_ResPubl cfOrgUnit_ResPubl cfProj_ResPubl cfResPubl-Event

Tabelle 1. Beispiel eines CERIF-Datensatzes

Tabelle 1 stellt beispielhaft einen CERIF-Datensatz für Projekte und Publikationen dar, in dem die allgemeinen (Basis/Kern, Ergebnis) und mehrsprachigen (lang) Attribute in den oberen Zeilen abgebildet sind. Die unteren Zeilen zeigen einige Beziehungen (Link), einschließlich ihrer formalisierten kontextuellen Semantik. Die Verknüpfung wird physisch durch IDs (z.B. cfResPublId, cfOrgUnitId, cfFundId, cfPersId, cfOrgUnitId, cfProjId, cfEventId) hergestellt, wie in der oberen Zeile der Attributspalte zu sehen ist. Die Spalte Typ gibt den konzeptionellen Entitätstyp an (base/core, link, lang). Die formalisierten semantischen Werte (z.B. „ist Urheber von“, „wird koordiniert von“, „wird finanziert von“, „ist Autor von“, etc.) werden in der Spalte Klassifikation dargestellt, die zur semantischen Ebene (Semantic Layer) gehört, wo jeder Wert einem vordefinierten Klassifikationschema zugewiesen wird (z.B. „cfProj_ResPubl“, „cfProj_OrgUnit“, „cfProj_Fund“, „cfPers_ResPubl, etc.).

3 Das KDSF-Basisdatenmodell

Der KDSF ist eine Spezifikation zum Austausch von Aggregatdaten. Diese setzen jedoch die einrichtungsinterne Vorhaltung von Basisdaten voraus. Entsprechend gliedert sich das einrichtungsinterne Datenmodell des KDSF konzeptuell in zwei Teilbereiche auf: Ein Teilmodell, das die Basisdaten definiert und ein Teilmodell für die Aggregationsebene. Zum Vergleich mit dem europäischen Standard CERIF ist nur das Basisdatenmodell von Relevanz, welches im Folgenden kurz eingeführt wird. Weitergehende detaillierte Informationen sind im Spezifikationsdokument zu finden (Institut für Forschungsinformation und Qualitätssicherung (ifQ) et al., 2015b) und auf der Webseite https://kerndatensatz-forschung.de/version1/technisches_datenmodell/.

Das Basisdatenmodell enthält sieben Objekte mit ihren jeweiligen objektspezifischen Attributen: (1) Personen, (2) Drittmittelprojekt, (3) Strukturiertes Promotionsprogramm, (4) Patent, (5) Ausgründung, (6) Publikation und (7) Forschungsinfrastruktur.

Neben den objektspezifischen Attributen gibt es noch zwei weitere zentrale Typen von Attributen: Verknüpfungsattribute, welche zwei oder mehr Basisdaten zueinander in Beziehung setzen und so genannte Zuordnungsattribute. Diese stellen eine Besonderheit des Datenmodells des KDSF dar. Es gibt drei Zuordnungsattribute, welche Objekte zu 1. Organisationseinheiten, 2. zu Fächern und 3. zu interdisziplinären und problemorientierten Forschungsfeldern zuordnen. Dies soll den Datenaustausch zwischen Datenabfragern und Dateneignern erleichtern.

Für die Erfassung dieser jeweiligen Attributstypen gelten verschiedene Regeln u.a. zur Erfassung von Mehrfachausprägungen, welche im Spezifikationsdokument niedergelegt sind.

Das Datenmodell des KDSF kann, genau wie das konzeptionelle Modell von CERIF, in Form eines Entity-Relations-Modells (ERM) abgebildet werden (Institut für Forschungsinformation und Qualitätssicherung (ifQ) et al., 2015b).

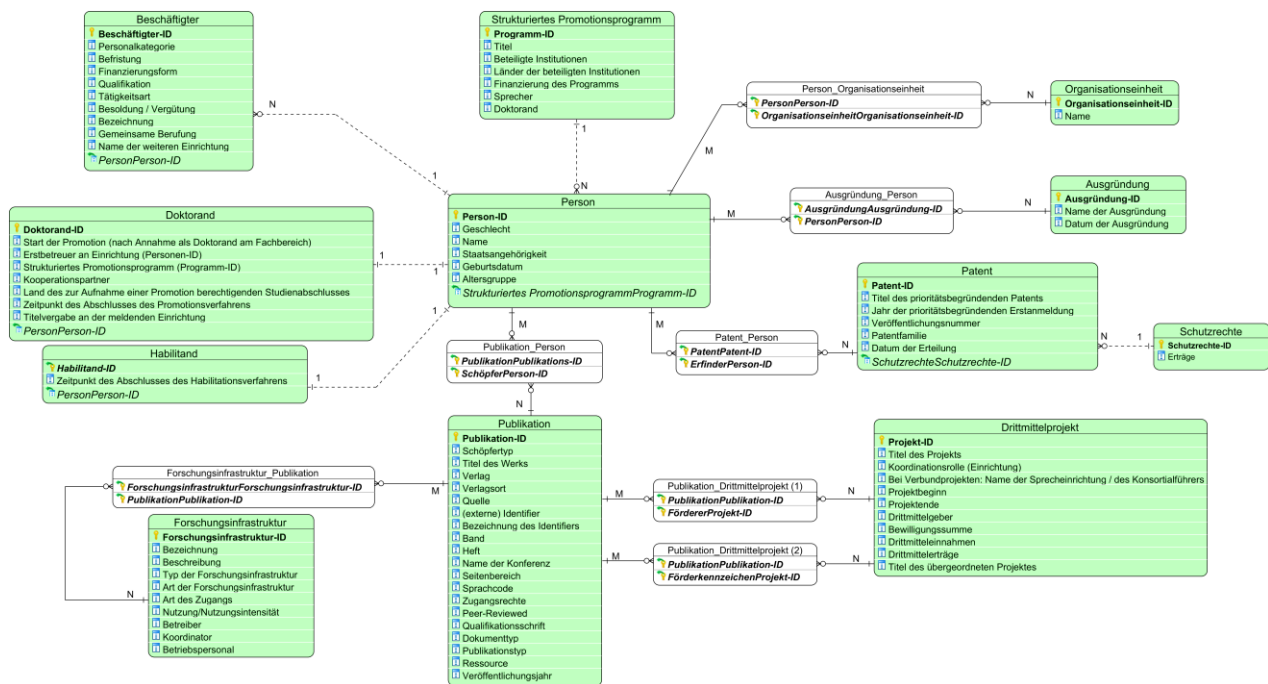


Abbildung 3. ERM des KDSF

Weiterführende Informationen zum ERM des KDSF finden sich auf der Plattform https://kerndatensatz-forschung.de/version1/technisches_datenmodell/ER-Modell.html

4 Konzeptioneller Vergleich der Architektur CERIF – KDSF

Ogleich sowohl CERIF als auch der KDSF als Standardspezifikationen dasselbe Ziel, nämlich die Harmonisierung und Standardisierung der Abbildung von Forschungsinformationen verfolgen, so wird an ihren jeweiligen konzeptionellen Datenmodellen sichtbar, welche unterschiedlichen Schwerpunkte hierbei gelegt werden. Das Datenmodell von CERIF ist aus historischer Perspektive dadurch gekennzeichnet, dass zunächst die Darstellung von Forschungsprojekten zentral war, zu der erst in Folgeversionen die weiteren Basisentitäten Organisationseinheit und Personen dazu kamen (Asserson et al., 2002). Im Gegensatz dazu zeichnet sich das Basisdatenmodell des KDSF durch das Primat der Person aus, welche qua Affiliation oder Arbeitsvertrag mit einer Forschungseinrichtung verbunden ist.

Auch der Umgang mit Beziehungen zwischen den Objekten und deren temporären Dimensionen unterscheidet sich deutlich zwischen beiden Datenmodellen: In CERIF sind eine deutlich höhere Anzahl an Objekten rekursiv miteinander und untereinander verbunden und jeweils alle mit zeitlichen Begrenzungen versehen, um beispielsweise aufeinander folgende Projekte oder parallele Beschäftigungen an verschiedenen Organisationen in verschiedenen Rollen deutlich zu machen (Asserson et al., 2002). Der KDSF hat als Ausgangspunkt die berichtende Einrichtung und die Grundannahme, dass keine neuen Berichtszwecke geschaffen, sondern bestehende Berichtsformate harmonisiert werden. Zu diesem Zweck werden ausschließlich aggregierte Daten ausgetauscht. Dies bedingt eine deutlich geringere Anzahl an Basisobjekten und deren Verknüpfungen miteinander als vergleichsweise in CERIF.

5 Mapping CERIF – KDSF

Ein Mapping, d.h. das Aufeinander Abbilden von Datenelementen, kann auf Ebene von Klassifikationsschemata erfolgen oder auf Ebene der Datenmodelle mit ihren jeweiligen Objekten, Attributen und Beziehungen. Der vorliegende Abgleich von CERIF mit dem KDSF bezieht nur die Datenmodelle ein und im Falle des KDSF nur das Basisdatenmodell, da es eine zu den Aggregatdaten vergleichbare Ebene in CERIF nicht gibt (Institut für Forschungsinformation und Qualitätssicherung et al., 2015a).

Ein vollständiges Mapping von CERIF mit dem KDSF ist bisher noch nicht veröffentlicht worden, mit Ausnahme der intensiven Vorarbeiten in dem Projekt zur Spezifikation des Kerndatensatz Forschung (Institut für Forschungsinformation und Qualitätssicherung et al., 2015a; Quix & Riechert, 2017). Im Gegensatz dazu wurden jeweils CERIF und der KDSF bereits zum Datenmodell der Open Source Anwendung VIVO, einem Forschungsinformationssystem basierend auf Linked Open Data, gemappt (Lezcano et al., 2013; Walther et al., 2019).

Der Abgleich von CERIF mit dem KDSF basiert auf einem manuellen Vergleich von Äquivalenzbeziehungen zwischen den Kernobjekten und ihren jeweiligen Eigenschaften. Dazu wurden die ERM-Modelle beider Standardspezifikationen herangezogen.

In einem ersten Schritt wurde ein quantitativer Vergleich der Modellmetriken, d.h. der Anzahl an Entitäten, deren Attribute sowie Verknüpfungen und Klassifikationsschemata, vorgenommen.

Das Ergebnis ist in Tabelle 2 abgebildet.

10

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

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

Tabelle 2. Modellmetriken CERIF und KDSF im Vergleich

Der Überblick der Modellmetriken zeigt substantielle Unterschiede auf, die in der Konstruktionslogik beider Standards begründet sind. Der KDSF basiert auf dem Prinzip der Datensparsamkeit und dem Einsatz einer überschaubaren Anzahl vielseitig einsetzbarer Kerndaten, welche für deutsche Forschungseinrichtungen berichtet werden sollen. Auf Basis der sieben Basisobjekte und ihrer Eigenschaften können verschiedene Aggregate gebildet werden. Im Gegensatz dazu gibt es in CERIF eine deutlich höhere Anzahl an Entitäten, um möglichst zahlreiche Facetten der Forschungsaktivitäten europäischer Länder umfassend und flexibel abbilden zu können. Das Datenmodell von CERIF erlaubt es zudem, dass Attribute auch als Verknüpfungen dargestellt werden können in Form von Link Entities.

Die Anzahl von Klassifikationsschemata in CERIF ist flexibel, nicht standardisiert und kann als Teil der Semantic Layer erweitert werden. Das Datenmodell des KDSF dagegen basiert auf einem kleineren, vordefinierten Set an Klassifikationsschemata für zum Beispiel Publikations- und Dokumenttypen, Personalkategorien oder Besoldungsgruppen oder Arten von Zuwendungsgebern, um das Berichtswesen vermehrt zu harmonisieren (Institut für Forschungsinformation und Qualitätssicherung et al., 2015a).

In einem zweiten Schritt wurden die Objekte und deren Eigenschaften inhaltlich miteinander verglichen und geprüft, inwiefern das konzeptionelle Datenmodell von CERIF die Elemente des KDSF abbilden kann. Es wurde dabei nicht berücksichtigt, ob durch Ergänzungen oder Transformationen des Datenmodells von CERIF eine Abbildbarkeit geleistet werden kann.

Die Ergebnisse werden in Tabelle 3 abgebildet. Für die Darstellung der Abgleichsbedingungen für jeden Bereich oder jedes Objekt des KDSF sowie die Elemente und Attribute des CERIF wurden zwei verschiedene Farben ausgewählt. Dies ist in Abbildung 4 wie folgt veranschaulicht.

Elemente sind in beiden Datenmodellen enthalten

Elemente sind in KDSF, nicht in CERIF

Abbildung 4. Abgleichsbedingungen

KDSF			CERIF
Area	Attribute	Entity	Mapping
Person (person)	Altersgruppe (age group)	person	not modeled
	Geburtsdatum (birth-date)	person	cfPers.cfBirthdate
	Geschlecht (gender)	person	cfPers.cfGender
	Name (name)	person	fPersName_Pers, cfPers-Name
	Staatsangehörigkeit (citizenship)	person	cfPers_Country
	hat Promotionsberechtigung aus (has doctoral eligibility from)	person	cfPers_Country
	Qualifikation (qualification)	employee	cfPers_Qual, cfQualTitle
	Besoldung (salary)	professors	not modeled
	Bezeichnung der Professur (professorial title)	professors	not modeled
	Gemeinsame Berufung (joint appointment)	professors	not modeled
	Abschlusszeitpunkt (completion date)	qualification procedure	not modeled

	Altersgruppe bei Abschluss (age group at graduation)	qualification procedure	not modeled
	Kooperationspartner bei kooperativer Promotion (cooperation partner in cooperative doctorate)	doctoral candidate	not modeled
	Start der Promotion - titelvergebende Einrichtungen (start of the doctorate - institution awarding the title)	doctoral candidate	not modeled
	Start der Promotion - nicht titelvergebende Einrichtungen (start of the doctorate - non-title awarding institution)	doctoral candidate	not modeled
	Zeitpunkt des Abschlusses des Promotionsverfahrens (time of completion of the doctoral procedure)	doctoral candidate	not modeled
	hat Organisationseinheit (has organizational unit)	person	cfPers_OrgUnit
	hat Qualifikationsverfahren (has qualification procedure)	person	not modeled
	hat Beschäftigung (has employment)	employee	not modeled
	hat Befristung (has limited term contract)	employment	not modeled
	hat Fach (has discipline)	employment	not modeled
	hat Finanzierungsform (has form of financing)	employment	not modeled
	hat Forschungsfeld (has research field)	employment	not modeled
	hat Personalkategorie (has staff category)	employment	not modeled
	hat Tätigkeitsart (has type of work)	employment	not modeled
	hat Bezeichnung (has name)	professors	not modeled

	hat gemeinsame Berufung (has joint appointment)	professors	not modeled
	hat Erstbetreuer (has supervisor)	doctoral candidate	not modeled
	hat Kooperationspartner (has cooperation partner)	doctoral candidate	not modeled
	hat Strukturiertes Promotionsprogramm (has structured doctoral program)	doctoral candidate	not modeled
Drittmittelprojekt (third-party funded project)	Bewilligungssumme (amount of funding)	third-party funded project	cfProj_Fund::cfAmount
	Drittmittelleinnahmen (third-party funding simple entry accountancy)	third-party funded project	not modeled
	Drittmittelerträge (third-party funding double entry accountancy)	third-party funded project	not modeled
	Förderkennzeichen (grant number)	third-party funded project	not modeled
	Koordinationsrolle (coordinating role)	third-party funded project	not modeled
	KoordinatorEinrichtung (coordinating institution)	third-party funded project	cfProj_OrgUnit
	Projektbeginn (start of project)	third-party funded project	cfProj.cfStartDate
	Projektende (end of project)	Third-party funded project	cfProj.cfEndDate
	Titel des Projekts (title of project)	third-party funded project	cfProjTitle
	hat Fach (has discipline)	third-party funded project	cfProj_Class
	hat Forschungsfeld (has research field)	third-party funded project	cfProj_Class

	hat Mittelgeber (has type of funder)	third-party funded project	cfProj_Class
	hat Organisationseinheit (has organizational unit)	third-party funded project	cfProj_cfOrgUnit
	hat übergeordnetes Projekt (has superordinate project)	third-party funded project	cfProj_Proj
Strukturiertes Promotionsprogramm (structured doctoral program)	Beteiligte Institutionen (participating institutions)	structured doctoral program	not modeled
	Länder der beteiligten Institutionen (country of participating institutions)	structured doctoral program	not modeled
	Titel des Promotionsprogramms (title of doctoral program)	structured doctoral program	not modeled
	hat Fach (has discipline)	structured doctoral program	not modeled
	hat Finanzierungsform (has form of financing)	structured doctoral program	not modeled
	hat Forschungsfeld (has research field)	structured doctoral program	not modeled
	hat Organisationseinheit (has organizational unit)	structured doctoral program	not modeled
	hat Sprecher (has speaker)	structured doctoral program	not modeled
	hat laufende Promotion (has ongoing doctorate)	structured doctoral program	not modeled
Patent (patent)	Datum der prioritätsbegründenden Erstanmeldung (date of first priority application)	patent	cfResPat.cfRegistrDate
	Titel des Patents (title of patent)	patent	cfResPatTitle
	Veröffentlichungsnummer (publication number)	patent	cfResPat.cfPatentNum

	Anzahl der Patentfamilien (number of patent families)	patent	not modeled
	Datum der erteilten Patente (date of granted patents)	patent	cfResPat.cfApprovDate
	hat Erfinder (has inventor)	patent	cfPers_ResPat
	hat Fach (has discipline)	patent	cfResPat_Class
	hat Forschungsfeld (has research field)	patent	cfResPat_Class
	hat Organisationseinheit (has organizational unit)	patent	cfOrgUnit_ResPat
Ausgründung (spin-off)	Name der Ausgründung (name of spin-off)	Spin-off	not modeled
	Datum der Ausgründung (date of spin-off)	Spin-off	not modeled
	hat Fach (has discipline)	Spin-off	not modeled
	hat Forschungsfeld (has research field)	Spin-off	not modeled
	hat Organisationseinheit (has organizational unit)	Spin-off	not modeled
Publikation (publication)	Förderer (funding organisation of publication)	publication	not modeled
	Förderkennzeichen (grant number of publication)	publication	not modeled
	Identifier (identifier)	publication	cfFedId, cfFedId_Class, cfDCResourceIdentifier
	ist Peer Reviewed (is peer-reviewed)	publication	cfResPubl_Class
	Ressource (resource)	publication	cfDCResourceType
	Sprachcode (language code)	publication	cfDCLanguage
	Titel der Publikation (title of publication)	publication	cfResPublTitle
	Veröffentlichungsjahr (publication year)	publication	cfResPubl.cfResPublDate
	Zugangsrechte (access rights)	publication	cfDCRightsMMAccessRights

	hat Dokumenttyp (has document type)	publication	cfResPubl_Class
	hat Name der Quelle (has name of source)	publication	not modeled
	hat Fach (has discipline)	publication	cfResPubl_Class
	hat Format (has format)	publication	cfResPubl_Event, cfEvent_Class, cfResPubl.cfVol, cfResPubl.cfIssue, cfResPubl.cfStartPage, cfResPubl.cfEndPage
	hat Forschungsfeld (has research field)	publication	cfResPubl_Class
	hat Organisationseinheit (has organizational unit)	publication	cfOrgUnit_ResPubl
	hat Publikationstyp (has publication type)	publication	cfResPubl_Class
	hat Qualifikationsschrift (has thesis)	publication	cfResPubl_Class
	hat Schöpfer (has creator)	publication	cfPers_ResPub
	hat Unterstützung durch (has support by)	publication	not modeled
	hat Forschungsinfrastruktur (has research infrastructure)	publication	cfResPubl_Equip
	hat Verlag (has publisher)	publication	cfOrgUnit_ResPubl
Forschungsinfrastruktur (research infrastructure)	Beschreibung der Forschungsinfrastruktur (description of research infrastructure)	research infrastructure	cfEquipDescr
	Bezeichnung der Forschungsinfrastruktur (name of research infrastructure)	research infrastructure	cfEquipName
	hat Art der Forschungsinfrastruktur (has kind of research infrastructure)	research infrastructure	cfEquip_Class
	hat Betreiber (has operator)	research infrastructure	cfOrgUnit_Equip

	hat Organisationseinheit (has organizational unit)	research infrastruc- ture	cfOrgUnit_Equip
	hat Betriebspersonal (has operating staff)	research infrastruc- ture	cfPers_Equip
	hat Koordinator (has coordinator)	research infrastruc- ture	cfOrgUnit_Equip
	hat Nutzung / Nutzungs- intensität (has use inten- sity)	research infrastruc- ture	not modeled
	hat Publikation (has publication)	research infrastruc- ture	cfResPubl_Equip
	hat Typ der Forschungs- infrastruktur (has type of research infrastructure)	research infrastruc- ture	cfEquip_Class
	hat Zugangsart (has type of access)	research infrastruc- ture	cfEquip_Class

Tabelle 3. Mapping KDSF mit CERIF

Das Mapping ergibt, dass ein Großteil der Kerndaten des KDSF in CERIF abgebildet wird und der KDSF damit weitgehend CERIF-kompatibel ist. Jedoch können der Bereich der Nachwuchsförderung und der Teilbereich Ausgründungen nicht abgebildet werden. Einige Ausdifferenzierungen von Personen, wie Doktorand*innen oder Beschäftigte, insbesondere deren Ausdifferenzierungen in Personalkategorien des deutschen Wissenschaftssystems, werden aktuell nicht abgebildet, da CERIF als europäischer Standard für Forschungsinformationen naturgemäß nur begrenzt die Spezifika nationaler Forschungssysteme berücksichtigen kann. Es wäre jedoch möglich, Klassifikations-schemata für Personalkategorien, die der deutschen Forschungslandschaft entsprechen, in die Semantic Layer zu integrieren.

6 Referenzen

Asserson, A., Jeffery, K. G., & Lopatenko, A. (2002). CERIF: Past, Present and Future: An Overview. In „*Gaining Insight from Research Information*“: *Proceedings of the 6th International Conference on Current Research Information Systems* (S. 33–40). euroCRIS; Kassel University Press.

Dvořák, J., & Brasse, V. (2019, November 18). *CERIF Tutorial: Autumn 2019 Münster strategic membership meeting*. Autumn 2019 euroCRIS Strategic Membership Meeting (WWU Münster, Germany, Nov 18-20, 2019), Münster.

Institut für Forschungsinformation und Qualitätssicherung (ifQ), Fraunhofer-Institut für Angewandte Informationstechnik (FIT), & Geschäftsstelle des Wissenschaftsrats (2015a). *Ergebnisbericht zum Projekt „Kerndatensatz Forschung“*. https://www.kerndatensatz-forschung.de/version1/Ergebnisbericht_Projekt_KDSF.pdf

Institut für Forschungsinformation und Qualitätssicherung (ifQ), Fraunhofer-Institut für Angewandte Informationstechnik (FIT), & Geschäftsstelle des Wissenschaftsrats. (2015b). *Spezifikation des Kerndatensatz Forschung—Version 1.0*. https://www.kerndatensatz-forschung.de/version1/Spezifikation_KDSF_v1.pdf

Jörg, B. (2010). CERIF: The Common European Research Information Format Model. *Data Science Journal*, 9, CRIS24–CRIS31. <https://doi.org/10.2481/dsj.CRIS4>

Lezcano, L., Jörg, B., Lowe, B., & Corson-Rikert, J. (2013). Promoting International Interoperability of Research Information Systems: VIVO and CERIF. *Journal of Universal Computer Science*, 19(12), 14.

Meier, A. (1998). *Relational Databases—An Introduction to Practice* (3. Aufl.).

Quix, C., & Riechert, M. (2017). Modelling national research information contexts based on CERIF. *Procedia Computer Science*, 106, 253-259. <https://doi.org/10.1016/j.procs.2017.03.023>

Russell, R. (2011). *An introduction to CERIF*. UKOLN, University of Bath. www.ukoln.ac.uk/rim/documents/Introduction_to_CERIF_1.0.pdf

Simons, E. (2014, April 2). *EuroCRIS and CERIF. The Importance of an International Standard Metadata Model for Research Information*. SK-CRIS Event (CVTI SR, Bratislava, Apr 2nd, 2014), Bratislava.

Walther, T., Hauschke, C., & Kasprzik, A. (2019). The Research Core Dataset (KDSF) in the Linked Data context. *Procedia Computer Science*, 146, 29–38. <https://doi.org/10.1016/j.procs.2019.01.074>