BZKO: An Ontology for the Card Index of German Post-War Compensation Records

Dilek Yargan[1,*], Jörg Waitelonis[1], Mahsa Vafaie[1,2] and Harald Sack[1,2]

[1]FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany
[2]Karlsruhe Institute of Technology, Institute of Applied Informatics and Formal Description Methods, Kaiserstr. 89, 76133 Karlsruhe, Germany

Abstract

The Central Federal Card Index (*Bundeszentralkartei*) of Germany is a key archival resource documenting compensation claims submitted by victims of National Socialist persecution and their relatives, within the German *Wiedergutmachung* process. To enable semantically enriched representation, integration, and reuse of this historically significant collection, we present the BZK Ontology (BZKO). We propose a two-layer ontology for historical archival data that separates ontologically grounded domain semantics from interoperability-oriented extension constructs. The approach combines BFO-based realism with archival standards (RiC-O, PROV-O, PiCo), enabling provenance-preserving semantic integration, while maintaining logical rigor, modularity, and reuse across digital humanities infrastructures. The proposed approach establishes a reusable semantic foundation for the integration of *Wiedergutmachung* archival materials into digital humanities infrastructures and lays the groundwork for future knowledge graph generation, ontology validation, and the incorporation of additional historical entities and uncertain temporal and spatial information.

The ontology is available on https://github.com/ISE-FIZKarlsruhe/bzko.

Keywords

BZK card index, Bundeszentralkartei, Wiedergutmachung, National Socialist Injustice, ontology, BFO, NFDIcore, archival documents, digital humanities

1. Introduction

The historical process of *"Wiedergutmachung nationalsozialistischen Unrechts" (Wiedergutmachung for National Socialist Injustice)*, left behind voluminous amount of compensation records that constitute one of the most historically significant archival collections related to post-war Germany. *Wiedergutmachung*, literally "making good again", refers to the reparation and restitution process initiated following the fall of the National Socialist regime in Germany, which between 1933 and 1945 systematically persecuted groups of people on the grounds of ethnicity, religion, political belief, and sexual orientation among others [1]. The archival collections related to *Wiedergutmachung* comprise over 100 km of archival items, such as index cards, applications for compensation, restitution, and hardship provisions, as well as administrative files. Although not all the materials are available, either because they are not in digitized form or because there are restrictions under archival law for a subset of the cards, to make the documentary legacy of Wiedergutmachung accessible for searching for persons or groups, organizations, and scientific research, the German Federal Ministry of Finance has initiated the project of *Themenportal Wiedergutmachung*, a central digital portal for the collection of *Wiedergutmachung* files from various state and federal archives in Germany and beyond [2, 3]. To this end, the description and digitalization of archival material, using various information extraction and semantic structuring technologies, have been required to facilitate accessibility and semantic search.

Proceedings of FOIS 2026 Satellite events co-located with the 16th International Conference On Formal Ontology in Information Systems (FOIS 2026), 21–25 September 2026, Vitória, Brazil.

*Corresponding author.
Dilek.Yargan@fiz-Karlsruhe.de (D. Yargan); Joerg.Waitelonis@fiz-Karlsruhe.de (J. Waitelonis); Mahsa.Vafaie@fiz-Karlsruhe.de (M. Vafaie); Harald.Sack@fiz-Karlsruhe.de (H. Sack)
0000-0001-9618-6740 (D. Yargan); 0000-0001-7192-7143 (J. Waitelonis); 0000-0002-7706-8340 (M. Vafaie); 0000-0001-7069-9804 (H. Sack)



Copyright © 2026 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



published 2026-11-27

CEUR-WS.org/Vol-4342/paper7.pdf

The Central Federal Card Index (*Bundeszentralkartei*, abbreviated BZK) is a card index containing over 1.9 million cards, known as *BZK index cards*, or *BZK cards* for short. A BZK card is an administrative record created by a clerk at a compensation office after a person applied for compensation, mostly under the Federal Compensation Act (*Bundesentschädigungsgesetz*, BEG for short) [4]. It contains key personal and procedural details, such as name, birthplace, residential address, date of death, the name of the responsible compensation authority, and a reference number, together with a range of additional personal and administrative details, such as marital status, nationality, information on heirs or surviving dependents, and the claimed type of compensation. The BZK cards have then served as an internal work instrument for compensation authorities, registering compensation applications, facilitating central tracking and cross-referencing of cases across different authorities. In the future, as the BZK itself has now become archival material, these cards will serve as the hub of the other archival materials in the collections, and making them available for the public is of the highest importance for *Themenportal Wiedergutmachung*.

This paper introduces the BZKO (BZK Ontology), designed to semantically represent relevant data from BZK index cards and to serve as the foundation for a knowledge graph that supports semantic search on *Themenportal Wiedergutmachung*. The paper is organized as follows. Section 2 provides an overview of the data contained on the cards to be modeled. Section 3 presents prior work related to modeling and ontology design in the context of the Wiedergutmachung project, and Section 4 details the methodology and tools used to model the ontology. Section 5 suggests the classes, relations, and shortcuts for modeling the data on BZK cards. Section 6 discusses the findings, and Section 7 concludes by sketching plans for future work.

In this paper, class names are written in italics along with their corresponding namespaces, e.g., *bfo:entity*. Relations, together with their namespace, are written in bold, e.g., **rdfs:subClassOf**. Lastly, the logical connectors are in small all-caps, e.g., SOME. Resource-specific naming conventions are preserved, such as the use of PascalCase or camelCase e.g., *picom:PersonObservation* or **rico:hasOrHadIdentifier**.

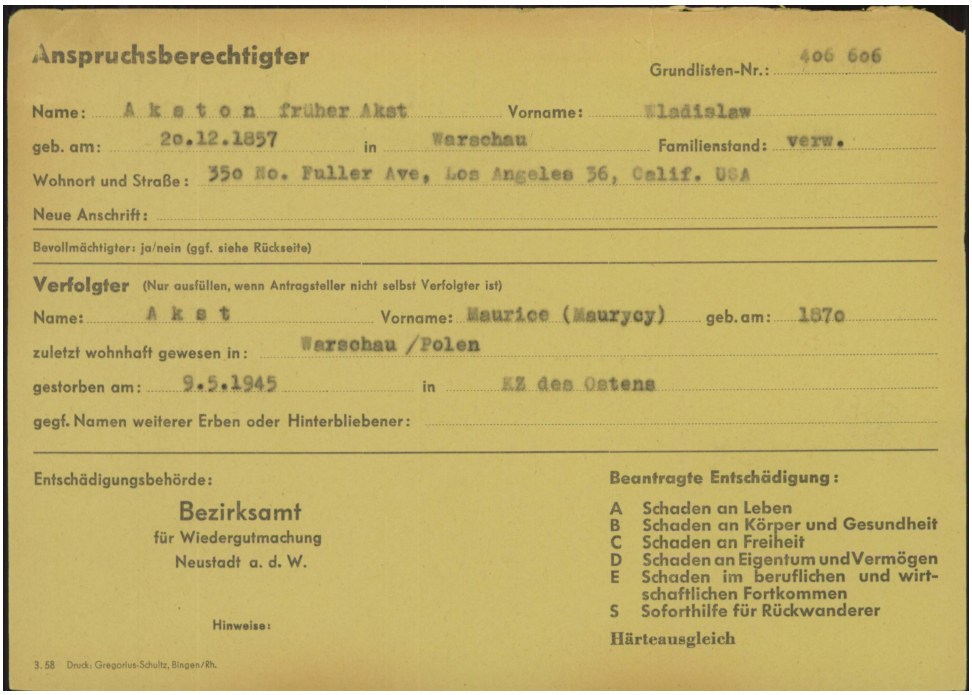
The ontology is publicly available at GitHub on https://github.com/ISE-FIZKarlsruhe/bzko.

2. Information Recorded on the BZK Index Cards

The BZK cards served originally as an internal tool that were created when the case entered the compensation system to identify which compensation claim was filed by whom, on behalf of whom, and which compensation authority handled the case. Thus, the information recorded on the cards was primarily used to identify individuals and cases. Now, the index cards themselves, as archival material, will provide the first basis for semantic search on *Themenportal Wiedergutmachung*.

Vafaie et al.[5] introduced BZKOpen [6], a dataset designed to serve as a ground truth, i.e., the correct reference data for measuring the performance of information extraction algorithms. Its distinct fields were manually annotated by the domain experts from the Federal Archives of Germany. The dataset consists of 516 digitized BZK cards, curated to respect privacy constraints while preserving a diverse range of layout styles. This manually annotated, high-quality sample has been used in this study to build and evaluate the BZK Ontology.

The information recorded on the cards was primarily used to identify individuals, including applicants for compensation and persons persecuted by the Nazi regime, and organizations (more details in Section 4). As such, the cards may include the name of the compensation office, the BZK number (a reference number used for tracking compensation claims), applicant's and/or persecutee's first name, last name, birth name, and other alternative names, date of birth, birthplace, persecutee's date of death, deathplace, and last residential address (see Figure 1). Some cards contain more information about the applicant, the persecutee, and the heirs, such as their marital status or residential address, while others lack structured sections for registering information about the applicant or the persecutee.



Anspruchsberechtigter Grundlisten-Nr.: 406 606

Name: A k s t o n früher Akst Vorname: Wladislaw

geb. am: 20.12.1857 in Warschau Familienstand: verw.

Wohnort und Straße: 350 No. Fuller Ave, Los Angeles 36, Calif. USA

Neue Anschrift:

Bevollmächtigter: ja/nein (ggf. siehe Rückseite)

Verfolgter (Nur ausfüllen, wenn Antragsteller nicht selbst Verfolgter ist)

Name: A k s t Vorname: Maurice (Maurycy) geb. am: 1870

zuletzt wohnhaft gewesen in: Warschau /Polen

gestorben am: 9.5.1945 in KZ des Ostens

gegf. Namen weiterer Erben oder Hinterbliebener:

Entschädigungsbehörde:

Bezirksamt
für Wiedergutmachung
Neustadt a. d. W.

Hinweise:

3. 58 Druck: Gregorius-Schultz, Bingen/Rh.

Beantragte Entschädigung:

A Schaden an Leben
B Schaden an Körper und Gesundheit
C Schaden an Freiheit
D Schaden an Eigentum und Vermögen
E Schaden im beruflichen und wirtschaftlichen Fortkommen
S Soforthilfe für Rückwanderer

Härteausgleich

Figure 1: A sample of a BZK card from BZKOpen [6]. Compensation office: Bezirksamt für Wiedergutmachung Neustadt a. d. W.; BZK number: 406 606; BZK card layout: RLP-Hauptphase; Applicant's first name: Wladislaw; Applicant's last name: Akston; Applicant's birth name: Akst; Applicant's date of birth: 1857-12-20; Applicant's birthplace: Warschau; Applicant's residential address: 390 No. Fuller Ave, Los Angeles 36, Calif. USA; Persecutee's first name: Maurice; Persecutee's last name: Akst; Persecutee's alternative first name: Maurycy; Persecutee's date of birth: 1870; Persecutee's last residential address: Warschau/Polen; Persecutee's date of death: 1945-05-09; Persecutee's deathplace: KZ des Ostens. Source: LAV NRW R, BR 3015.

3. Related Work

Within the context of the *Wiedergutmachung* project and in direct relation with the *Wiedergutmachung* collection, there is some prior research with a focus on modeling and ontology design. Modeling archival collections on the Semantic Web requires ontologies that can jointly address archival structure, provenance, and the historical actors and events embedded in the records.

In [7], Vafaie et al. address the challenge of representing heterogeneous archival documents and justify the reuse of the Records in Contexts Ontology (RiC-O)[1] to model the *Wiedergutmachung* archival documents. They argue that RiC-O's approach of modeling archival hierarchy levels as named individuals offers a more generic strategy that broadens its applicability across archives of different historical periods and geographical contexts. They also point out that RiC-O's inclusion of finer-grained entities within the hierarchy improves findability and allows for a more precise representation of archival resources.

In CourtDocs [8], the authors propose a data model to represent the *Wiedergutmachung* process, starting with the application for compensation. This data model addresses the challenges of representing relationships between documents, the agents involved in their creation, and the procedural events they describe, in a semantically coherent and interoperable way. Heterogeneous document types created in legal proceedings, as well as different administrative and judiciary steps are represented in this data model, drawing on established ontologies, namely RiC-O, the Provenance Ontology (PROV-O)[2], and the CIDOC Conceptual Reference Model (CRM)[3].

[1] https://www.ica.org/standards/RiC/RiC-O_1-1.html
[2] https://www.w3.org/TR/prov-o
[3] https://cidoc-crm.org

4. Two-Layer Ontology Design

To represent the content of the cards, the BZK Ontology (BZKO) is currently built as an application ontology and initially developed in a bottom-up manner. Although a BZK card can contain a wide range of information, such as an applicant's name and date of birth, their marital status, compensation office that issued the claim, and the names of the representative, as an initial step, we have collected all features from the BZKOpen database [6] that constitute the main entities of the ontology. In order to support interoperability, standardization, and ontology reuse, we built upon two established ontologies. The Basic Formal Ontology (BFO), a top-level ontology standardized in ISO/IEC 21838-1 [9, 10, 11], ensures philosophical and formal consistency at scale. It thereby provides clear distinctions between material objects and information artifacts, roles and dispositions, processes, and description-logics-based formality, which are often intentionally blurred in widely used digital humanities tools, such as CIDOC CRM, for practical reasons. Using BFO, thus, minimizes ambiguity in modeling decisions and promotes consistent modeling. NFDIcore, a BFO-based mid-level ontology developed for representing metadata related to the German National Research Data Infrastructure (NFDI), was selected to ensure interoperability within the NFDI ecosystem [12, 13]. NFDIcore reuses existing ontologies, including the Information Artifact Ontology (IAO), the Relations Ontology (RO), and the Ontology for Biomedical Investigations (OBI).

Moreover, we have extended the ontology by aligning it with and importing from the following semantic resources. The Provenance Ontology (PROV-O), a W3C standard, is mainly used to represent and exchange provenance information consistently. Thus, its use enables data traceability across cards and supports accountability. The Records in Contexts Ontology (RiC-O), as mentioned above, is used to represent and manage archival records. Its classes and relations enable the semantically enriched representation of BZK cards and their digitized form in archival terminology. Although RiC-O and NFDIcore have classes to represent persons identified in the BZK cards, we imported the classes from Persons in Context (PiCo)[4] to extend semantic representations of persons, as it is specifically designed to model historical person data and genealogical reconstructions. Lastly, Schema.org[5] data properties are imported, as this semantic resource provides broad interoperability through a widely recognized vocabulary. It thereby enhances interoperability and facilitates data integration.

BZKO adopts a two-layer architecture. The *knowledge layer* contains all the necessary and sufficient classes and relations to model the information recorded on a BZK card. The classes in this layer are primitive (in description logics terminology) and rigid (in OntoClean terminology) classes. In other words, they are the building blocks for representing the content of the BZK cards. Their essential characteristics persist throughout their existence, and they compose other BZKO-related classes. For instance, the "BZK card" class is introduced as a primitive class, since it represents an entity that cannot be defined using existing BZKO classes. It is also considered rigid; its instances cannot cease to be a BZK card without ceasing to exist.

The *extension layer* contains imports from the above-mentioned semantic resources, as well as the defined and non-rigid classes and relations of BZKO. Accordingly, the release artifact of the knowledge layer, `bzk.owl`[6], includes imports from BFO and NFDIcore, which subsume the primitive classes of BZKO. The artifact of the extension layer, `bzk-extension.owl`[7], includes all classes and relations of BZKO and shortcuts and mappings based on the mentioned semantic resources (See Section 5.2 and Section 5.3).

The rationale behind the architecture is twofold. The two-layer structure improves reusability by enabling the knowledge layer to be used independently across different applications. For instance, the knowledge layer of BZKO can be imported on its own and extended with RiC-O, without requiring the extension layer or its PiCo-based classes. This approach also improves maintainability by ensuring that changes to the knowledge layer are propagated in a controlled manner, preventing circular modifications.

[4] https://personsincontext.org
[5] https://schema.org
[6] https://github.com/ISE-FIZKarlsruhe/bzko/blob/main/bzk.owl
[7] https://github.com/ISE-FIZKarlsruhe/bzko/blob/main/bzk-extension.owl

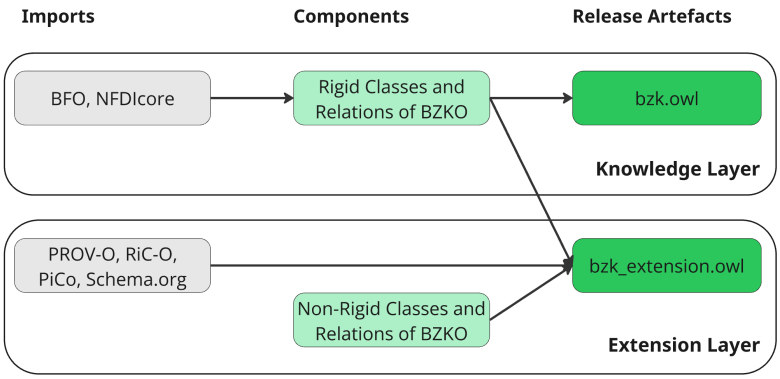


Figure 2: Two-layer design of BZKO, with each layer corresponding to a distinct release artifact. Arrows denote dependency and inclusion relations: each target incorporates or depends on its source.

Figure 2 illustrates the two-layer architecture of BZKO and its main structural components. Each layer is characterized by three elements: imported semantic resources, components, and a corresponding release artifact. The components of the layers were determined based on the ontological statuses of features in BZKOpen [6]. Each feature was examined to determine the most specific BFO and NFDIcore class, or classes, where compositional modeling was required, under which it could be appropriately subsumed. Then, each feature was represented either by a single ontology class or by a composition of classes. For instance, the feature "BZK number" was represented by *bzko:bzk number*, since its being an identifier that persists over time in this domain; whereas, "applicant" was modeled through the composition of *bzko:applicant role*, *nfdi:person*, and **ro:has role**, since roles are dependent on the bearer and are realized in processes: a person can have multiple roles (applicant and persecutee roles) at once or hold a role (applicant) at different times.

Each class and relation was then analyzed to determine whether it could be defined in terms of other classes and relations in BZKO. The primitive classes and relations constitute the components of the knowledge layer, while the others, together with their formal equivalents, either constitute the components of the extension layer or are represented through shortcut relations. Finally, to provide mappings from domain-related semantic resources, relevant external classes were incorporated into the imports of the extension layer.

We are compiling three collections of competency questions in collaboration with experts from the Federal Archives of Germany and the State Archives of Baden-Württemberg. The first collection includes organizational questions that define the scope of the ontology and support the validation of terminology choices. The second collection includes validation questions that assess the ontological correctness with respect to its logical structure. The third collection includes knowledge acquisition questions derived from use cases to evaluate whether the ontology meets project requirements and constraints. The answers of the first collection have been reflected in the ontology design, namely, the coverage of the ontology was approved by the domain experts. On the other hand, the last two are ongoing work. The evaluation of these collections requires data; therefore, they were excluded from the current evaluation and deferred to future work.

For ontology development, we follow the Open Biological and Biomedical Ontology (OBO) Foundry Principles [14], a collection of best practices aimed at supporting orthogonality and interoperability across ontologies [15]. The ontology has been implemented in OWL (Web Ontology Language), using Protégé 5.6.7.

The Ontology Development Kit (ODK) has been used to manage ontology development, dependency imports, and quality control in a reproducible manner [16, 17]. External classes and object properties were imported from reference ontologies using ODK's import pipeline, which extracts only the required terms together with their relevant axioms while preserving references to the source ontologies with

their particular versions. This approach minimizes redundancy and facilitates interoperability. Different ontology variants were managed through the ODK build workflow, enabling the automatic generation of release artifacts and import modules for distinct use cases. Ontology quality was assessed using the validation tools integrated into ODK, including automated reasoning and consistency checks, as well as ontology quality-control tests to identify missing annotations, unsatisfiable classes, and structural inconsistencies prior to release.

5. Results

In the following subsections, the components of each BZKO layer and their relations to the imported classes and relations are discussed.

5.1. The Knowledge Layer of BZKO

The knowledge layer includes classes and relations from BFO and NFDIcore, as well as classes and relations imported from ontologies such as IAO, OBI and RO, which are included in BFO and NFDIcore. It also includes rigid classes of BZKO, which are necessary for defining domain knowledge in the BFO framework.

A BZK index card, a physical archival object, records information about agents, names, dates, places, and identifiers. Its corresponding representation, *bzko:bzk card*, is necessary to establish the provenance of the data. A digitized version of a physical BZK card has been used to extract information, so its corresponding representation is necessary not only to represent the domain correctly within the BFO framework, but also because the file names of the digitized cards have been used in date-extraction procedures, as they are the dates of birth of either applicants or persecutees. Hence, *bzko:digitized bzk card* is also introduced in the ontology.

During the indexing process, most BZK cards were assigned an identifier, called the BZK number, which is a Federal Central Register number referring to a compensation claim. This number may be used to track the corresponding compensation files in the archives holding them today. Consequently, the introduction of *bzko:bzk number* is necessary.

The content of a card includes information about agents, namely, about persons (applicants and persecutees) and an organization (a compensation office). In this context, an applicant is a person who was recorded on a BZK card and applied for compensation for a persecuted person or for themselves; a persecutee, on the other hand, is a person who was subjected to persecution by the Nazi regime, including but not limited to discrimination, dispossession, forced displacement, or violence, on political, racial, religious, or other grounds. A compensation office is a government agency that reviews claims and provides compensation to persecuted people. In order to model these agents within the BFO framework, we need to define their roles, as being an applicant or a compensation office is not per se. These roles depend on a process, which realizes them. For this reason, we introduce *bzko:compensation card indexing*, a subclass of *bfo:process*, along with *bzko:applicant role*, *bzko:persecutee role*, and *bzko:compensation office role*.

The class *bzko:compensation card indexing*, however, is associated with roles that are not actually realized through the indexing process: the applicant role is realized through participation in the compensation application process, while the persecuted-person role is realized through relevant persecution-related processes. In our model, therefore, we intentionally adopt a coarser level of granularity. The purpose of the BZK card is to identify persons as applicants or persecuted persons and organizations as compensation offices. The card-indexing process records the attribution of these roles to the corresponding agents rather than realizing the roles themselves.

A collection of information on the cards contains the names of the agents. On the one hand, representing the names of compensation offices is straightforward, i.e., the name of a compensation office instantiates *bzko:compensation office name*; on the other hand, representing the names of persons requires a more fine-grained modeling. The first and family names of persons are represented by *iao:given name* and *iao:family name*, respectively. Additionally, the family name a woman had before

marriage is represented by *bzko:birth name*. In the context of BZKO, *bzko:alternative given name* and *bzko:alternative family name* represent given names and family names that have been altered or rendered in German or Latin forms, respectively. However, none of these classes in the ontology is domain-specific.

A distinct collection of information on the cards concerns dates and places. Both applicants and persecutees have dates of birth and birthplaces, but some of the persecuted people have dates of death and death places. To represent dates and places, *bzko:birth process boundary* and *bzko:death process boundary* are introduced [9]. How to represent birthplaces, deathplaces, and dates of birth and death based on these process boundaries will be defined in Section 4.2.

Addresses constitute another collection of information on the cards. The class *iao:postal address* is used to represent the addresses of persons during BZK card creation, and *bzko:last residential* address is introduced to represent the last known voluntarily addresses of persons voluntarily during persecution.

There is one class left, which is neither the content of a card nor necessary for the ontological representation of content within the BFO framework: *bzko:bzk card layout*. A layout type refers to the classification of BZK index cards by experts at the Federal State Archive, based on the spatial distribution of information across the card layout. For instance, the layout of the BZK card shown in Figure 1, labelled *RLP-Hauptphase*, provides distinct, structured sections for both the applicant and persecutee, but the layout type *NRW-Innenministerium* lacks such sections. This entity is required in the modeling, as criteria for determining the layout types have been used in the information extraction processes.

5.2. Non-Rigid Classes and Relations of BZKO

A class is *non-rigid* (in the OntoClean sense [18]) if its instances can cease to instantiate it without ceasing to exist. For instance, a person can cease to *be* an applicant while continuing to *exist* as a person. This distinction important in BZKO: roles, places, and dates are all non-rigid, as they depend on a specific context or process.

The first group of non-rigid BZKO classes concerns agents. The people whose names appear on the BZK cards hold applicant and/or persecutee roles within the Wiedergutmachung context. For this reason, they are represented by the classes *bzko:wgm applicant* and *bzko:persecutee of the nazi regime*. The following are examples of defined classes in BZKO.

> *bzko:persecutee of the nazi regime* **rdfs:subClassOf** *nfdi:person* AND (**ro:has role** SOME *bzko:persecutee role*)
>
> *bzko:compensation office* **rdfs:subClassOf** *obi:organization* AND (**ro:has role** SOME *bzko:compensation office role*)

To enhance the semantic representation of the BZK cards, the following relations between the classes are introduced: **bzko:applies for compensation for**, **bzko:applies for compensation to**, and **bzko:issues**. Figure 3 illustrates the domain and the range of these object properties, along with the imported and rigid and non-rigid classes of BZKO, with all superclasses up to the closest BFO class in the taxonomy. All the object and data properties in the BZKO knowledge layer are already defined in NDFIcore: they are either BFO, RO, or IAO relations, or NFDIcore relations.

The second group of non-rigid BZKO classes contains places. A place can be labeled in various forms, such as a birthplace, a hiding place, or a marketplace, but none of them are defined by the ontological status of the place itself (Cf. [18]). A place, then, gains such extrinsic characteristics only within a context. The same holds for the third group of non-rigid BZKO classes, viz., dates. A date is modeled as a temporal instance; however, a date of birth is a temporal instance that depends on a birth process boundary and is therefore context-dependent. Consequently, *bzko:birthplace* and *bzko:deathplace*, and *bzko:date of birth* and *bzko:date of death* are non-rigid.

The first group of relations, namely object properties of BZKO, is introduced to enhance the semantic representation of the BZK cards. As already illustrated in Figure 3, **bzko:applies for compensation**

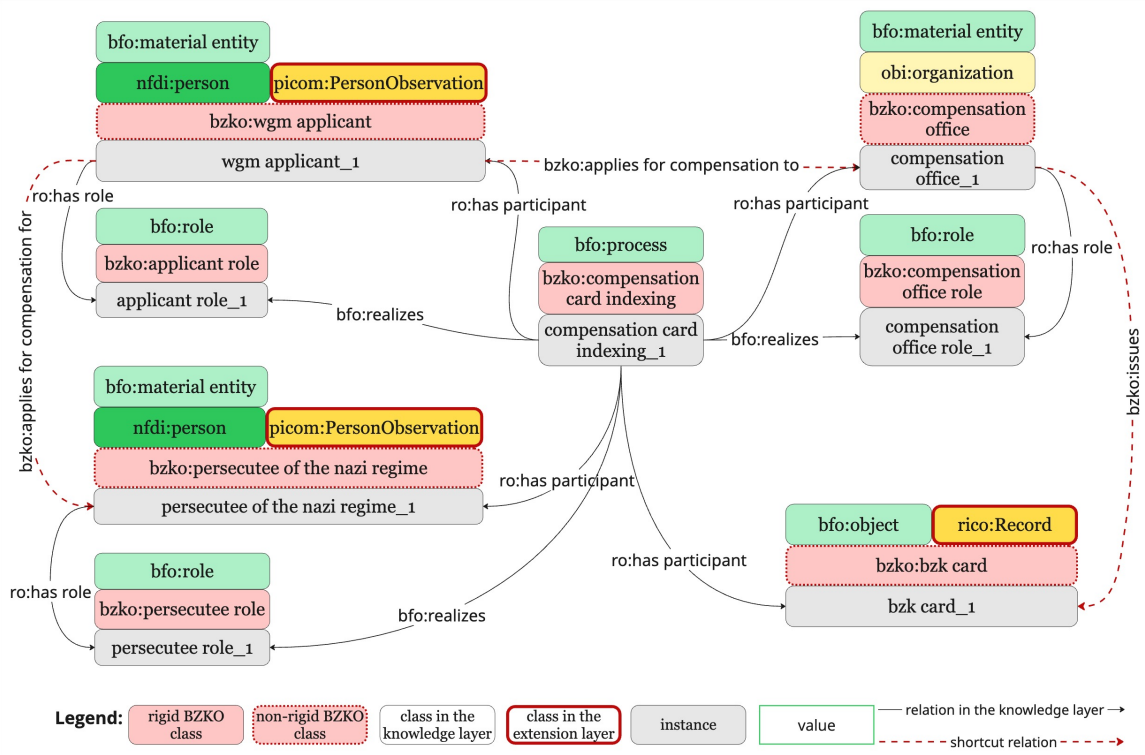


Figure 3: The process of compensation card indexing and its participants, along with all superclasses up to the closest BFO class in the taxonomy, the example instances, and the relations between them.

for, **bzko:applies for compensation to**, and **bzko:issues** relate the classes to expose the domain knowledge beyond the content of the BZK cards.

The second group of relations, namely data properties of BZKO, is introduced for convenience, e.g., to enable direct access to literal values. These relations are redundant since the domain can already be represented using the existing entities. Figure 4 illustrates the relations between a BZK index card, its digitized version, its layout, and its number. The relations, namely data properties, **bzko:has bzk number** and **bzko:has layout type** are added to facilitate semantic search, as they directly relate class instances to their data values. For the same purpose, other BZKO data properties are introduced to relate names and addresses to their corresponding literal strings.

5.3. The Extension Layer of BZKO

The extension layer of BZKO includes (i) all classes and relations from the knowledge layer; namely, all rigid classes and primitive relations (Section 5.1); (ii) the non-rigid classes and relations of BZKO (Section 5.2); and (iii) classes and relations from the imported semantic resources PROV-O, RiC-O, PiCo, and Schema.org, widely used semantic standards in the digital humanities, to extend BZKO.

The classes *rico:Record* and *rico:Instantiation* are used to define *bzko:bzk card* and *bzko:digitized bzk card*, respectively, thereby assigning archival characteristics to the latter classes. BZKO also includes *picom:PersonObservation*, a subclass of *picom:Person*. Since *picom:Person* refers to human beings, we interpret *picom:PersonObservation* as referring to persons whose personal details are recorded in source documents. A more appropriate label would therefore be "Observed Person". Under this interpretation, a person mentioned on a BZK card instantiates both *picom:PersonObservation* and *nfdi:person*.

For the future integration of additional Wiedergutmachung documents, which also contain information about people, organizations, and addresses, into the Themenportal Wiedergutmachung, tracking the source of the data becomes crucial. In this respect, **prov:hadPrimarySource** is added to the ontology, and SPARQL CONSTRUCT queries are defined to link the classes associated with data properties with the corresponding digitized BZK cards through this relation, hereby enabling instances and their associated

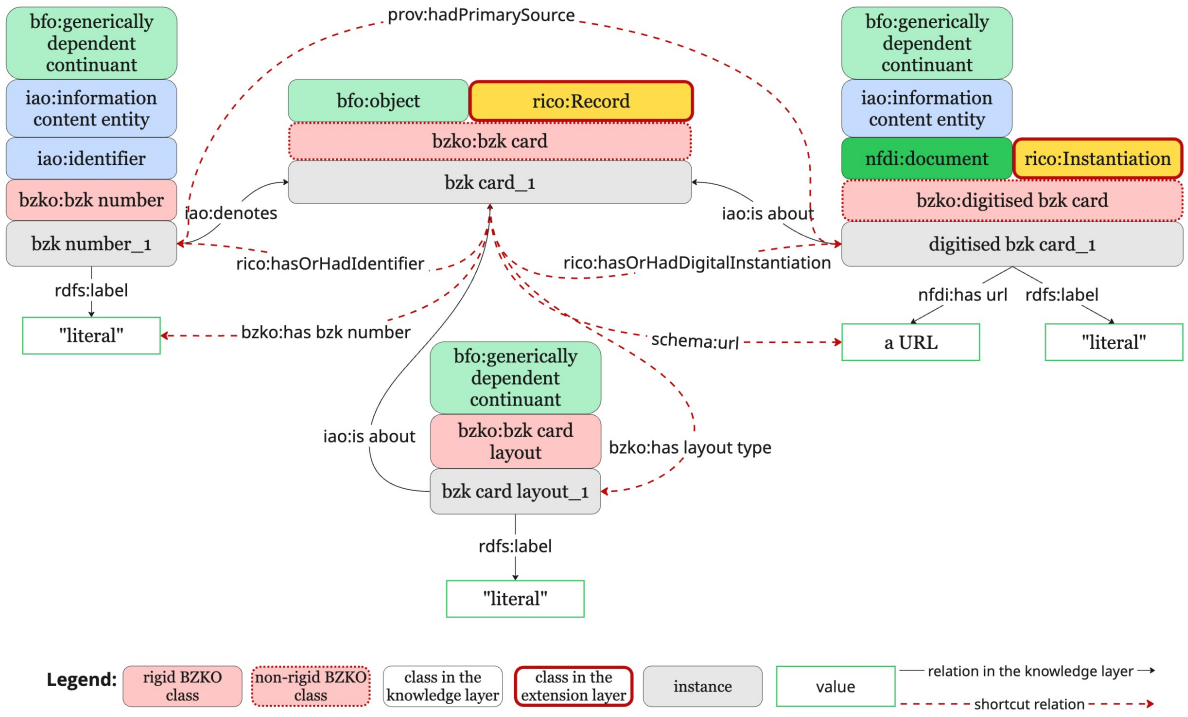


Figure 4: A diagram showing the relationships between a BZK index card, its digitized representation, its layout, and its number, as well as relationships in the extension layer, including the imported property **prov:hadPrimarySource** and a shortcut property **bzko:has layout type**.

property values to be traced back to their corresponding digitized BZK cards. Figure 4 illustrates how these entities are used in the modeling, along with rigid and non-rigid classes and relations of BZKO.

To express the relation between shortcut relations and their corresponding ontological patterns, we have defined a set of rules. The Semantic Web Rule Language (SWRL) rules were not used because they use transitive properties. As a result, not all reasoners, such as HermiT [19], fully support them. Moreover, SWRL rules may not allow the creation of new individuals, which is required when mapping shortcut relations to more complex ontological patterns. To ensure broad support across OWL reasoners, we defined rules in the form of SPARQL CONSTRUCT queries. These queries establish explicit mappings, allowing seamless translation in both directions; from shortcuts to full patterns and vice versa. By encoding these relationships as CONSTRUCT queries, we ensure that the mappings are formally defined, machine-readable, and consistently applicable across datasets based on the ontology. The same approach applies to complex mappings to external vocabularies, where CONSTRUCT queries are likewise used to define the rules governing the translation. As a simple example, consider **bzko:has bzk number** in Figure 4. Assuming all the prefixes are provided, which are available in the OWL file, a CONSTRUCT query is designed to retrieve the BZK number directly from the card as follows.

```
CONSTRUCT {
   ?bzk_card1 bzko:BZK_0000003 ?label .}   #has bzk number
WHERE {
   ?bzk_card1 a bzko:BZK_0000002: .        #bzk card
   ?bzk_number1 a bzko:BZK_0000003 .       #bzk number
   ?bzk_number1 iao:denotes ?bzk_card1 .
   ?bzk_number1 rdfs:label ?label .
}
```

6. Discussion

The BZK Ontology is being developed not merely as a graph representation of the OpenBZK, but as a semantic model that captures entities underlying the BZK cards. It aims at balancing interoperability with domain-specific expressiveness by being BFO-compliant and reusing prominent semantic resources in digital humanities, viz., RiC-O, PiCo, and PROV-O. As a result, it enables representation of archival data beyond what is explicitly stored in the database and also the maintenance of traceability to the contents of the cards. This traceability is implemented by linking all data values (agent names, dates, places, BZK numbers, addresses) to *bzko:digitized bzk card* via **prov:hadPrimarySource**.

Although **prov:wasDerivedFrom**, which is the superproperty of **prov:hadPrimarySource**, has been subsumed by **ro:causally influenced by** in [20] (Cf. [21]), we do not adopt the latter relation in the knowledge layer of BZKO. On the one hand, causality is the central notion underlying **ro:causally related to** and its subproperties [22], on the other hand, provenance concerns linking the data values to the digitized BZK cards. In BZKO, thus, provenance relations are not used in the sense of causal influences (Cf. [23]).

At present, *bzko:compensation card indexing* models the creation of the card and identification of agents. A fuller model should distinguish among index card creation, application submission, and prosecution-related processes, and include relations that link each realizable entity to the process in which it is realized. A more fine-grained representation of these processes is left for future work.

BZK card layouts are modeled as named individuals in the ontology for two reasons. First, we cannot speak of an instance of a layout type: the design of the card, which is contingent, determines the layout type rather than instantiating it. Second, the archivists created a list of criteria to determine the card layout labels and provided a ground truth [6] in which cards are associated with their corresponding layout. This suggests that cards are mapped to an item of a list.

In the ideal case, birth and/or death dates are modeled using **time:inXSDDate**, with values in the "YYYY-MM-DD" format. However, our review of the cards reveals that dates of birth and/or death may contain uncertain dates (e.g., "unbekannt" (*Eng.* "unknown")), approximate dates (e.g., "ca. 1900"), incomplete dates (e.g., "Im Juli 1942"), or ambiguous dates (e.g., "1943 oder 1944", "14/26. März 1880", "75.70.1953", "1886–1865"). The ontology, then, must be extended to semantically represent all these cases. However, this is left as future work.

Some values recorded in the deathplace field do not denote a determinate place: they may refer to events (e.g., "in Deportation"), death-related annotations (e.g., "verstorben" (*Eng.* "deceased")), or place names qualifed by uncertanty (e.g., "wahrscheinlich Auschwitz" (*Eng.* "probably Auschwitz")), rather than plain assertions of location. Therefore, representing deathplaces as structured locations would not adequately capture the data. Consequently, these values are currently modeled as literal strings.

In the knowledge layer, we chose **ro:has participant** over **bfo:has participant** (Cf. [20]), as the domain of the former covers *bfo:process boundary*, which is required to model birth and death as occurrences. Furthermore, *bzko:compensation office name* is subsumed under *iao:institutional identification*. However, its definition ("A textual entity intended to identify a particular institution") appears to suggest that it would more appropriately be classified under *iao:identifier*, rather than *iao:textual entity*. We adhere to the IAO design choices, although we find this modeling decision debatable.

We intentionally avoid committing imported classes to specific BFO branches, since they do not provide a sufficiently specific BFO commitment. For instance, *rico:Record*, the superclass of *rico:RecordResource*, may refer to either an information content entity, as an instance of *bfo:generically dependent continuant*, or a material entity, as an instance of *bfo:independent continuant*. Similarly, *picom:PersonObservation* refers to the persons who are observed on a *picom:Source*, which is, again, can be an information content entity or a material entity. Thus, in order to avoid future implementation errors and inconsistencies, these classes are maintained as mapping-layer constructs rather than integrated into the BFO taxonomy.

The high-level patterns in the knowledge layer are not immediately clear when viewing the ontology. So, we introduced equivalent classes and shortcuts in the extension layer to compact complex expressions into simpler ones [24]. The BZKO relations, covering **bzko:issues**, **bzko:has layout type**, and **bzko:has**

alternative given name, are introduced as shortcuts. Although in the knowledge layer complex multihop relation paths already express the semantics, they are not easy to understand for a non-ontology expert. Thus, BZKO relations are introduced to not only improve usability and readability by providing a straightforward, natural-language alignment but also to facilitate faster querying.

The extension layer also contains imported relations from RiC-O, PROV-O, and PiCo as shortcuts. These relations are preserved in the ontology when their semantics cannot be fully represented by the corresponding relations in the knowledge layer. For instance, **rico:hasOrHadInstantiation** extends the semantics of the model by characterizing a digitized BZK card as an archival instantiation rather than merely relating an instance of *bfo:object* to an instance of *iao:information content entity*. Similarly, **rico:hasOrHadIdentifier**, defined as a relation that "connects a Thing to one of its past or present Identifiers", whose definition reads "[c]onnects a [owl:]Thing to one of its past or present [rico:]Identifiers.", was not modeled as a subproperty of **iao:denoted by**, since the label of the RiC-O relation explicitly includes temporal aspects of identification, a separate mapping was provided in order to preserve its intended semantics. Additionally, to the best of our understanding ot the RiC-O documentation [25], the *hasOrHad-* naming convention used throughout RiC-O denotes a relation that may hold now or may have held at some point in the past, rather than committing to a specific formal theory of time. As such, these relations are compatible with the BFO framework, as they do not introduce any OWL DL temporal semantics.

7. Conclusion and Future Work

This paper illustrates the development of BZKO, a BFO-compliant ontology for index cards in the Wiedergutmachung documents. The ontology distinguishes between a knowledge layer, which captures ontologically grounded entities and relations, and an extension layer, which introduces shortcuts and mappings for imported classes and relations from RiC-O, PROV-O, and PiCo to preserve archival semantics, support provenance tracking, and facilitate integration with existing digital humanities infrastructures. Thus, we aimed at balancing ontological rigor with archival practicality.

The development of the BZK Ontology is still ongoing, and three directions for future work remain: ontology validation, knowledge graph generation and data enrichment, as well as an extension of the domain coverage.

The ontology requires further validation through competency questions. To this end, meetings with archivists are being conducted to refine and complete the competency questions on validation and use cases. Once these have been established, the ontology can be systematically evaluated against the underlying database. The resulting competency questions will also provide a foundation for developing SWRL rules and SPARQL queries.

Several research activities are currently underway within the *Themenportal Wiedergutmachung* Project. For instance, address parsing, the correction of place names, modeling vague spatial references, and the normalization and linking of location information to GeoNames are currently under investigation. These developments will require corresponding extensions to the ontology, particularly regarding the representation of historical places and geographical entities. As the results of these studies become available, the ontology will be revised to accommodate the newly identified requirements.

Similarly, ongoing research on the normalization of temporal information will inform future revisions of the ontology. The current model represents dates as literals and does not yet capture the full range of uncertain, incomplete, approximate, and ambiguous temporal data found on the index cards. Future results in temporal data normalization will therefore guide the extension of the ontology's temporal representation.

Finally, as mentioned above, the BZK cards contain additional information that has not yet been modeled. Examples include familial relationships between applicants and persecutees, nationality, and companies, political parties, trade unions, religious communities, and similar organizations appearing as applicants and/or persecutees. Incorporating these entities and relations will require further extensions of both the ontology and the underlying database, as well as the formulation of new competency

questions. Although these developments are considered long-term objectives, they remain part of the broader research agenda.

Acknowledgments

This work is funded by the German Federal Ministry of Finance (*Bundesministerium der Finanzen*). We are indebted to Inger Louise Banse-Strobel, Kevin Dubout, and Dominik Sturm for the many valuable discussions on the archival domain, for their contributions to the development of the competency questions, and for the comments on the earlier version of this paper. We thank the anonymous reviewers for their valuable comments and constructive suggestions, which helped improve the quality and clarity of this paper.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] German Federal Ministry of Finance, Wiedergutmachung-provisions relating to compensation for national socialist injustice, 2026. URL: https://www.bundesfinanzministerium.de/Content/EN/Standardartikel/Press_Room/Publications/Brochures/2018-08-15-entschaedigung-ns-unrecht-engl.pdf?__blob=publicationFile&v=3.

[2] Archivportal-D, Themenportal wiedergutmachung, 2024. URL: https://www.archivportal-d.de/themenportale/wiedergutmachung.

[3] German Federal Ministry of Finance, Archivierungsprojekt wiedergutmachung, 2021. URL: https://www.bundesfinanzministerium.de/Monatsberichte/2021/01/Inhalte/Kapitel-3-Analysen/3-7-archivierungsprojekt-wiedergutmachung.html, Monthly Report January 2021, in German.

[4] Düsseldorf District Government, Information Sheet on Compensation for Nazi Injustice (English), https://www.brd.nrw.de/document/20230201_1_15_BZK_Merkblatt_englisch.pdf, 2023. Available via the Compensation for Nazi Injustice information page: https://www.brd.nrw.de/Ueber-uns/Die-Bezirksregierung/Entschaedigung-fuer-Naziunrecht.

[5] M. Vafaie, S. Hertling, I. Banse-Strobel, K. Dubout, H. Sack, End-to-end information extraction from archival records with multimodal large language models, in: Proceedings of the 34th ACM International Conference on Information and Knowledge Management, 2025, pp. 6075–6083. doi:10.1145/3746252.3761503.

[6] M. Vafaie, Bzkopen dataset, 2025. URL: https://huggingface.co/datasets/MahsaVafaie/BZKopen, Hugging Face dataset repository.

[7] M. Vafaie, B. Oleksandra, N. Pilz, D. Dessí, H. Sack, et al., Modelling archival hierarchies in practice: Key aspects and lessons learned, in: CEUR workshop proceedings, volume 2981, CEUR-WS, 2021.

[8] M. Vafaie, O. Bruns, N. Pilz, J. Waitelonis, H. Sack, Courtdocs ontology: Towards a data model for representation of historical court proceedings, in: Proceedings of the 12th Knowledge Capture Conference 2023, K-CAP '23, Association for Computing Machinery, New York, NY, USA, 2023, p. 175–179. doi:10.1145/3587259.3627562.

[9] R. Arp, B. Smith, A. D. Spear, Building Ontologies with Basic Formal Ontology, MIT Press, Cambridge, MA, 2015.

[10] International Organization for Standardization, ISO/IEC 21838-2:2020 Information Technology — Top-Level Ontologies (TLO) — Part 2: Basic Formal Ontology (BFO), 2020. URL: https://www.iso.org/standard/74572.html.

[11] Basic Formal Ontology GitHub Repository, 2025. URL: https://github.com/bfo-ontology.

[12] National Research Data Infrastructure, 2024. URL: https://www.nfdi.de.
[13] NFDIcore Ontology GitHub Repository, 2025. URL: https://github.com/ISE-FIZKarlsruhe/nfdicore.
[14] Open Biological and Biomedical Ontology Foundry Principles, 2024. URL: https://obofoundry.org/principles/fp-000-summary.html.
[15] R. Jackson, N. Matentzoglu, J. A. Overton, R. Vita, J. P. Balhoff, P. L. Buttigieg, S. Carbon, M. Courtot, A. D. Diehl, D. M. Dooley, W. D. Duncan, N. L. Harris, M. A. Haendel, S. E. Lewis, D. A. Natale, D. Osumi-Sutherland, A. Ruttenberg, L. M. Schriml, B. Smith, C. J. Stoeckert Jr., N. A. Vasilevsky, R. L. Walls, J. Zheng, C. J. Mungall, B. Peters, Obo foundry in 2021: operationalizing open data principles to evaluate ontologies, Database 2021 (2021) baab069. doi:10.1093/database/baab069.
[16] N. Matentzoglu, D. Goutte-Gattat, S. Z. K. Tan, J. P. Balhoff, S. Carbon, A. R. Caron, W. D. Duncan, J. E. Flack, M. Haendel, N. L. Harris, W. R. Hogan, C. T. Hoyt, R. C. Jackson, H. Kim, H. Kir, M. Larralde, J. A. McMurry, J. A. Overton, B. Peters, C. Pilgrim, R. Stefancsik, S. M. Robb, S. Toro, N. A. Vasilevsky, R. Walls, C. J. Mungall, D. Osumi-Sutherland, Ontology development kit: a toolkit for building, maintaining and standardizing biomedical ontologies, Database 2022 (2022) baac087. doi:10.1093/database/baac087.
[17] INCATools, Ontology development kit, 2024. URL: https://incatools.github.io/ontology-development-kit/.
[18] N. Guarino, C. A. Welty, An overview of ontoclean, Handbook on ontologies (2009) 201–220.
[19] B. Glimm, I. Horrocks, B. Motik, G. Stoilos, Z. Wang, Hermit: an owl 2 reasoner, Journal of automated reasoning 53 (2014) 245–269. doi:10.1007/s10817-014-9305-1.
[20] Relations Ontology Project, Relations in RO and BFO, 2026. URL: https://github.com/oborel/obo-relations/blob/8154e83fd3e1b5203929dda24adbd8983ac00715/docs/ro-and-bfo.md.
[21] T. Prudhomme, G. De Colle, A. Liebers, A. Sculley, P. Xie, S. Cohen, J. Beverley, Prov-to-bfo: v2025-01-19, 2025. doi:10.5281/zenodo.14692262, [Data set] Zenodo].
[22] Relations Ontology, Ro: causally related to (ro_0002410), 2024. URL: http://purl.obolibrary.org/obo/RO_0002410.
[23] Relations Ontology, Causal relations in the obo relation ontology, 2025. URL: https://github.com/oborel/obo-relations/blob/master/docs/causal-relations.md.
[24] C. Mungall, A. Ruttenberg, D. Osumi-Sutherland, Taking shortcuts with owl using safe macros, Nature Precedings (2010) 1–1. doi:https://doi.org/10.1038/npre.2010.5292.1.
[25] Expert Group on Archival Description (EGAD), Records in Contexts Ontology (RiC-O) Version 1.1, Technical Report, International Council on Archives, 2023. URL: https://www.ica.org/standards/RiC/RiC-O_1-1.html.