



IN2N – Idee, Produkte und Kooperationsmodell

Alexander Haffner

Institutionenübergreifende Kooperation aus Sicht der DNB



- Neue Kooperationsmodelle für nicht-bibliothekarische Partner
- Fremddatenübernahme und -integration bestehender Datensets
 - Problem
 - Daten liegen zunehmend in nicht unterstützten Formaten vor
 - Lösung
 - Aufbau von Infrastruktur für Match&Merge
 - (semi-)automatische Zusammenführung
- kooperative Redaktion über das Web
 - Web-Service anbieten, der einfachen lesenden und schreibenden Zugriff gestattet

Das IN2N-Projekt



- DFG-gefördertes Forschungsprojekt
- Erstellung eines prototypischen Kooperationsmodells
- Deutsches Filminstitut übernimmt Rolle des Pilotpartners für das neu zu entwickelnde Kooperationsmodell

Phasenmodell



- **Phase 1:** Initialer Datenabgleich zwischen dem Datenbestand des Partners und der GND sowie anschließender beiderseitiger Import von Informationen, die durch die Gegenseite erwünscht sind, aber bislang nicht lokal existieren.
- **Phase 2:** Ein redaktioneller Routinebetrieb über das Web, wobei durch den Partner in Echtzeit in der GND gesucht wird sowie Änderungen im Bestand des Partners über Differenzmeldungen in die GND übermittelt werden.



Phase 1

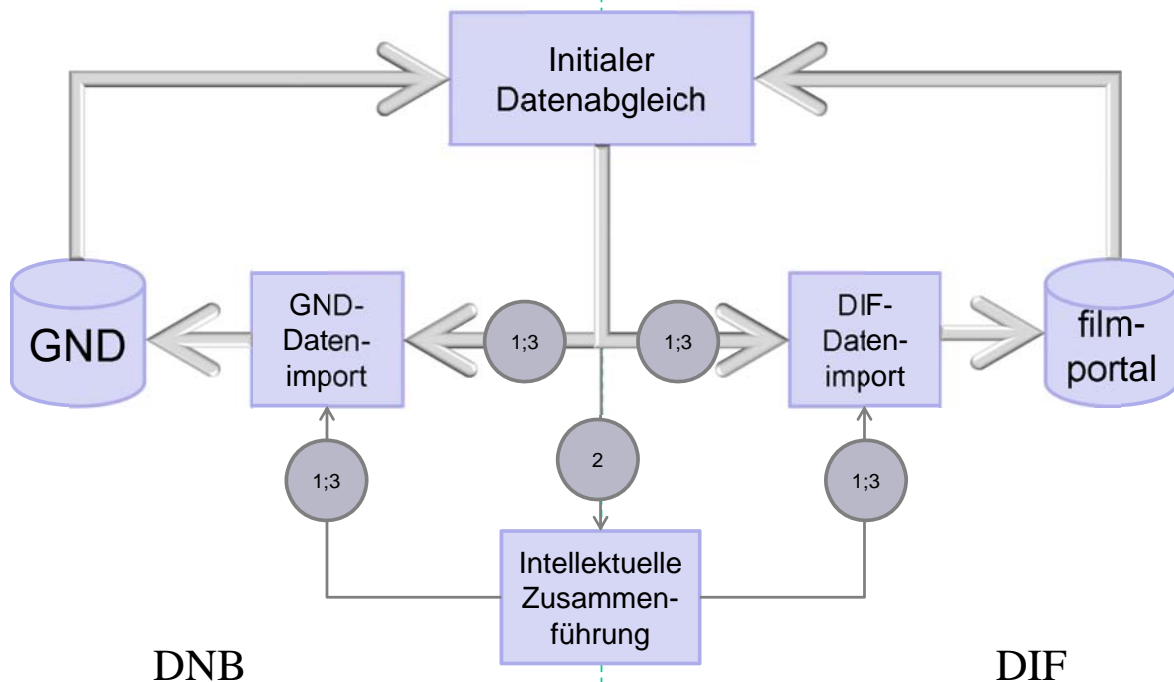
Initialer Datenabgleich und Dateneinspielung

Datenabgleich – Ziel der Ergebnisklassifikation

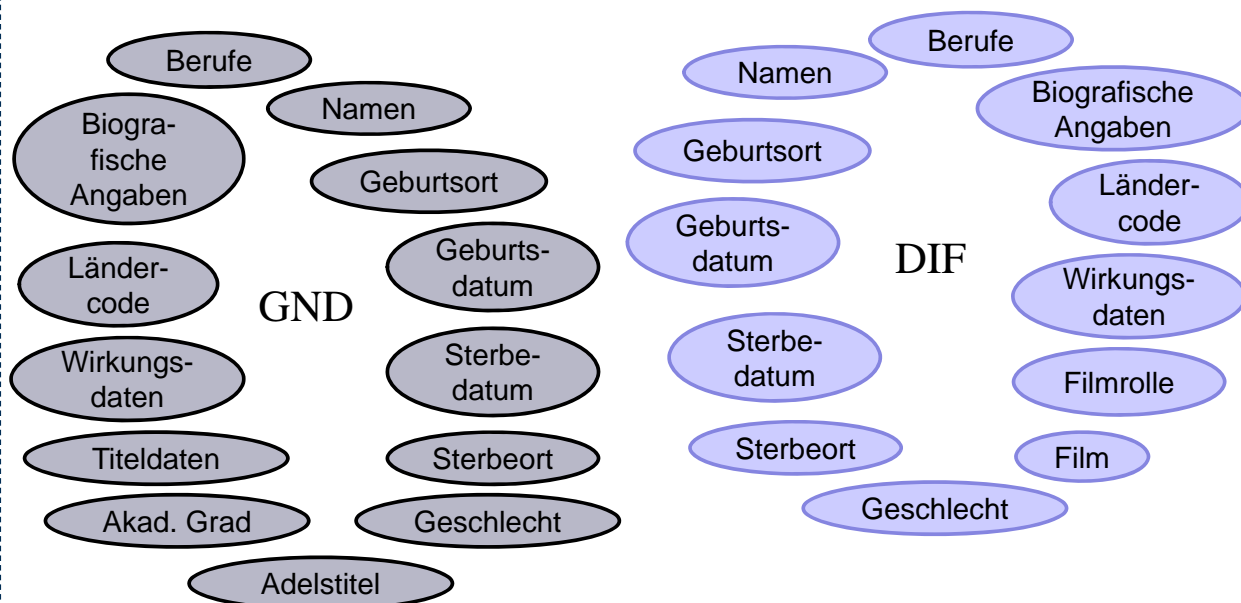


- 1 exakt ein Äquivalent in der GND zuordenbar ist (Konstellation von Datenelementübereinstimmungen, bei welcher die Personensätze automatisch zu einem gemeinsamen Satz zusammengeführt werden können),
- 2 eine oder mehrere potenzielle Äquivalente in der GND existieren (Konstellation, bei welcher die möglichen Treffer als „Kandidaten“ für eine Zusammenführung gekennzeichnet werden) oder
- 3 kein Äquivalent in der GND gefunden werden konnte (Konstellation, bei welcher keine Übereinstimmung angenommen werden soll).

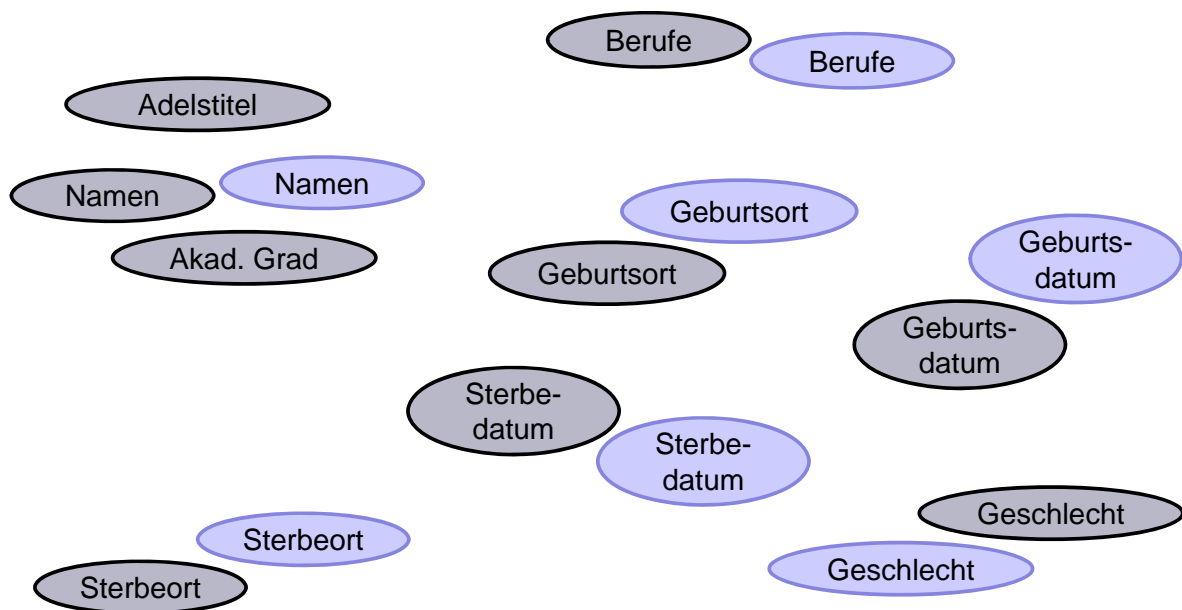
Initialer Match&Merge



Initialer Datenabgleich - Charakteristika



Initialer Datenabgleich - Charakteristika



Initialer Datenabgleich - Matchbedingungen

Elementtyp	Vorgehen
Namen	<ul style="list-style-type: none"> • Identifikation von Namenstypen • Zerlegung in Bestandteile • Zuordnung von GND- und DIF-Bestandteilen • Erstellung kontrollierter Listen für akademische Grade und Adelstitel
Datum	<ul style="list-style-type: none"> • Syntaxanpassung • Umgang mit ungenaueren Angaben
Orte	<ul style="list-style-type: none"> • Zuordnung von DIF-Strings zu verlinkten GND-Orten über vorhandene GND-DIF-Paare • Unscharfer Vergleich von DIF- und GND-Strings
Berufe	<ul style="list-style-type: none"> • Konkordanzerstellung mit semantischer Übereinstimmung
Geschlecht	<ul style="list-style-type: none"> • Wertanpassung

Vorgaben für automatische Zusammenführung



Fall	Kurzbeschreibung	Fall	Kurzbeschreibung	Fall	Kurzbeschreibung
1a	1 Name + exakte Lebensdaten + 1 Beruf	3c	1 Name + GND-Namensvorkommen > 3 + exakte Lebensdaten	5a	1 Name + Lebensdaten Jahresübereinstimmung + Geburts- oder Sterbeort
1b	1 Name + exaktes Geburtsdatum oder exaktes Sterbedatum + 1 Beruf	4a	1 Name + GND-Namensvorkommen <= 3 + Lebensdaten Jahresübereinstimmung + 1 Beruf	5b	1 Name + Geburtsjahr oder Sterbejahr Jahresübereinstimmung + Geburts- oder Sterbeort
2a	1 Name + exakte Lebensdaten + Geburts- oder Sterbeort	4b	1 Name + GND-Namensvorkommen > 3 + Lebensdaten Jahresübereinstimmung + 1 Beruf	6	1 Name + GND-Namensvorkommen <= 3 + Lebensdaten Jahresübereinstimmung
2b	1 Name + exaktes Geburtsdatum oder exaktes Sterbedatum + Geburts- oder Sterbeort	4c	1 Name + GND-Namensvorkommen <= 3 + Geburtsjahr oder Sterbejahr + 1 Beruf	7	1 Name + GND-Namensvorkommen <= 3 + Geburts- oder Sterbeort + 1 Beruf
3a	1 Name + GND-Namensvorkommen <= 3 + exakte Lebensdaten	4d	1 Name + GND-Namensvorkommen > 3 + Geburtsjahr oder Sterbejahr + 1 Beruf	8	1 Name + Namensvorkommen = 1 + 2 Berufe aus filmportal
3b	1 Name + GND-Namensvorkommen <= 3 + exaktes Geburtsdatum oder exaktes Sterbedatum				

IN2N-Workshop | 30.09.2014

Verfahren für potenzielle Äquivalente



- Gewichtung von Charakteristika zur Berechnung einer Match-Score
- Entwicklung einer webbasierten Anwendung zur benutzerfreundlichen Auflösung von potentiellen Äquivalenten in Kategorie 1 und 3

Josh Broecker ([8761C24263394FC3B6E7C3BC871583DA](#))
01.02.1963 Wuppertal -
Darsteller, Regie, Drehbuch

Broecker, Josh
[142197742](#)
1963 Wuppertal -
Regisseur
Regisseur, Drehbuchautor,
Darsteller
MS: 8 ✓ ✗

Kai Christiansen ([A897A45083A04E2FB3BD55E4D570D6BF](#))
1968
Regie, Drehbuch

Christiansen, Kai
[122721039](#)
1968-
-
Dt. Regisseur
MS: 3 ✓ ✗

Christiansen, Kai
[122591488](#)
1969-
-
Diss. Mathematisch-
Naturwissenschaftliche
Fakultät
MS: 1 ✓ ✗

Christiansen, Kai
[122293878](#)
1962-
-
Diss. Biologie
MS: 1 ✓ ✗

IN2N-Workshop | 30.09.2014

Evaluation des Datenabgleichverfahrens

- durch intellektuelle Prüfung
 - iterative Verfeinerung der Fallbeschreibungen und Match-Score-Berechnung
- durch existierende Cross-Konkordanzen
 - aus filmportal.de, aus Wikipedia etc.
- Untersuchung, ob GND-Anreicherungen aus externen Quellen zu besseren Ergebnissen führen
 - fehlende GND-Angaben z.B. aus Wikipedia-Templates oder VIAF in Match-Verfahren einbeziehen

Wikipedia-Personendaten

- [Dieter Pfaff](#) bei [filmportal.de](#)
- [Dieter Pfaff](#) in der [Internet Movie Database](#)

* {{Filmportal.de Name|http://www.filmportal.de/person/dieter-pfaff_53293e23e66545e1903423dcf34d6a90}}

* {{IMDb Name|0679294|Dieter Pfaff}}

Normdaten (Person): [GND: 140176667](#) | [LCCN: no2009034963](#) | [VIAF: 103722749](#) | [Wikipedia-Personensuche](#)

```
{{Normdaten|TYP=p|GND=140176667|LCCN=no/2009/34963|VIAF=103722749}}
```

```
{{Personendaten
|NAME=Pfaff, Dieter
|ALTERNATIVNAMEN=
|KURZBESCHREIBUNG=deutscher Schauspieler und Regisseur
|GEBURTSDATUM=2. Oktober 1947
|GEBURTSORT=[[Dortmund]]
|STERBEDATUM=5. März 2013
|STERBEORT=[[Hamburg]]
}}
```

Initiale Dateneinspielung

- Vorbereitende Maßnahmen - Festlegung welche Elemente:
 - aus filmportal in die GND übernommen werden
 - aus der GND in filmportal übernommen werden

- November 2014
 - alle Klasse 1 und 3 Daten aus filmportal werden in den Bestand der GND eingespielt (Master)
 - Damit auch als Linked Data verfügbar
 - Synchronisation der filmportal-Daten auf Basis des GND-Bestandes (Slave)

- ab November 2014
 - permanente intellektuelle Auflösung von potentiellen Äquivalenten in Klasse 1 und 3
 - alle Klasse 2 Daten verbleiben in nicht-öffentlichem Bereich

Eingesetzte Infrastruktur

- Modulare Systemkomponenten

- Datenabgleich
 - in relationaler Datenbank (MySQL)
 - jedes Importformat möglich
 - austauschbar durch beliebige Datenabgleichsysteme

- Dateneinspielung
 - Standard-GND-Ingest-Verfahren in DNB
 - Partner verwendet Synchronisationsmodul aus Phase 2

- Software für intellektuelle Zusammenführung
 - Match-Ergebnisse aus Fremdsystemen importierbar



Phase 2

Redaktioneller Routinebetrieb

Use Case – Suche



Use Case – Suche



Use Case – Ändern

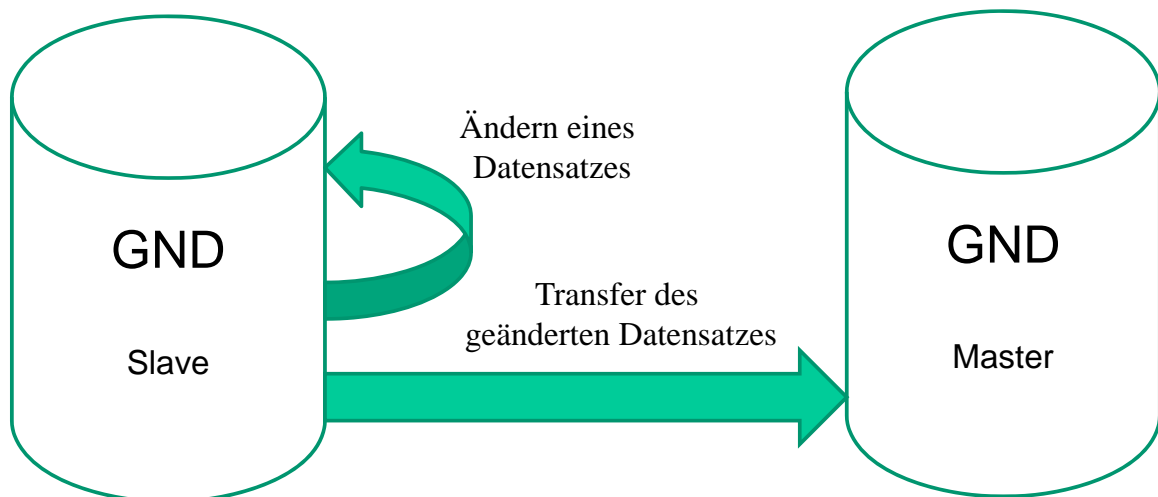


Ist-Stand bibliothekarischer Datentransfer

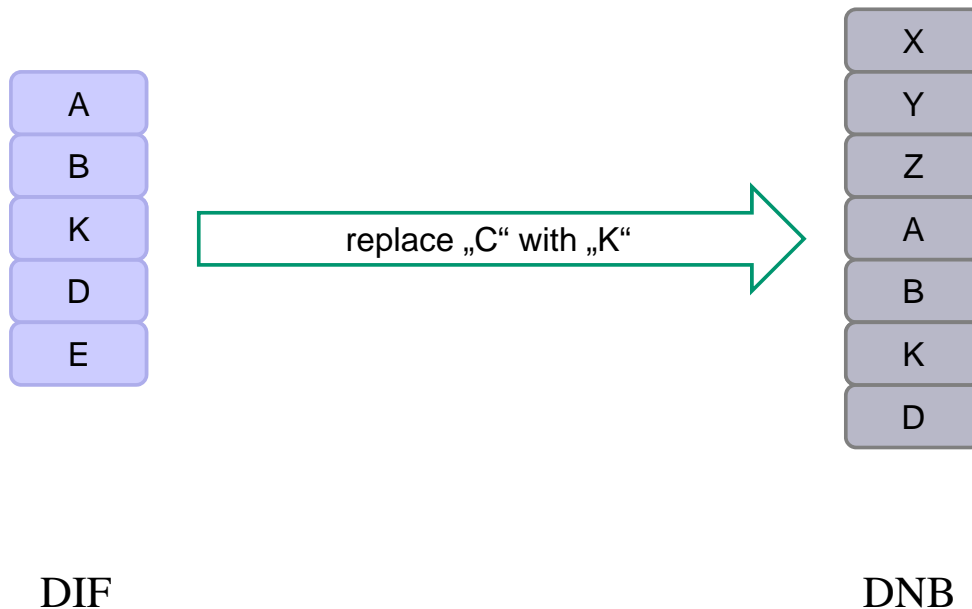


fast ausschließlich in MARC 21

Ist-Stand bibliothekarischer Datentransfer



Soll-Stand IN2N - Inkrementelle Updates

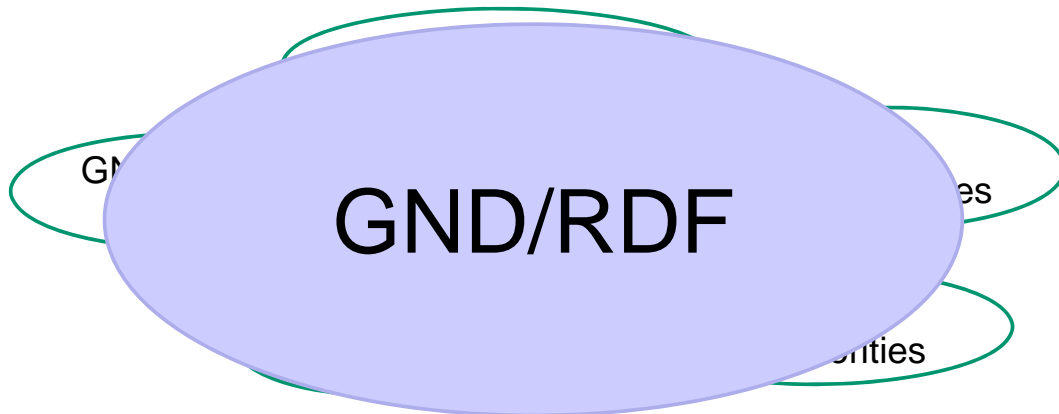


Inkrementelle Updates

- Innovation gegenüber bisherigen Verfahren
 - kein Austausch gesamter Datensätze
 - keine Deckungsgleichheit der Datensätze notwendig
 - Property-basiertes „add“, „remove“ und „replace“
- Ermöglicht sogar Akteuren, die den Inhalt eines GND-Datensatzes nicht kennen Änderungen vorzunehmen
 - bspw. Wikipedia, ReseachGate
 - Kontrolle über Aufnahme GND-seitig notwendig

Transferformate

- Zielgruppe
 - nicht-bibliothekarische Einrichtungen, die eine kooperative Normdatenpflege aus ihrer Arbeitsumgebung verrichten möchten
- Anforderung:
 - Intuitiv verständliches Format, das Web-Standards genügt



DNB-Schnittstellen

- Synchronisation
 - OAI-Protocol for Metadata Harvesting
 - URL-basierte Set-Anfragen
 - webbasierte Schnittstelle des Bibliothekswesens
 - XML-basiertes Containerformat in dem XML-Transferformate eingebettet werden
 - bisher: komplette Spiegelung der GND
 - neu: auf Partner-zugeschnittenes Datenset
- gezielte Suche
 - Search/Retrieve via URL (SRU)
 - URL-basierte Suchanfragen
 - Webschnittstelle des Bibliothekswesens
 - XML-basiertes Containerformat für Results in dem Datensätze in beliebigen XML-Transferformate eingebettet werden können
 - bisher: ungenaue Suchresultate durch gegebene Indexkonfiguration
 - neu: Verbesserung der Indices für gezieltere Result-Sets

Schnittstelle für inkrementelle Updates

- Simple Catalog Update Service Interface (SCUSI) - Arbeitstitel
- Verwendung etablierter Web-Schnittstellen
- Komplettlösung für Update und Neuanlage
 - Inkrementeller Ansatz
 - Property-basiertes Schreiben
- Reduzierter Formatumfang des Transferformatates
 - Konzentration auf ein gebräuchliches Subset an Datenelementen
 - Adressierung der Properties über Values
- Verwendung von JSON + GND/RDF

SCUSI Request: GND-Eintrag erzeugen

```
POST /scusi/v1/gnd
Host: services.dnb.de
Content-Type: application/json
Accept: application/json
Authorization: Token 585e02d222f908fe37f6c3407b7a8c4c
```

```
{
  "@Vocabulary" : "gnd",
  "NameOfThePerson" : [
    {
      "forename" : "Peter R.",
      "surname" : "Adam"
    },
    {
      "forename" : "Peter",
      "surname" : "Adam"
    }
  ],
  "dateOfBirth" : "1957-05-29",
  "placeOfBirth" : "http://d-nb.info/gnd/4046145-2",
  "professionOrOccupation" : [
    "http://d-nb.info/gnd/4274581-0",
    "http://d-nb.info/gnd/4375554-9"
  ],
  "professionOrOccupationAsLiteral" : "Aufnahmeleitung"
}
```

Lege in der GND einen Datensatz für Peter R. Adam an, der am 29.05.1957 in Pirmasens geboren wurde. Peter Adam war als Cutter und Toningenieur sowie in der Aufnahmeleitung tätig.

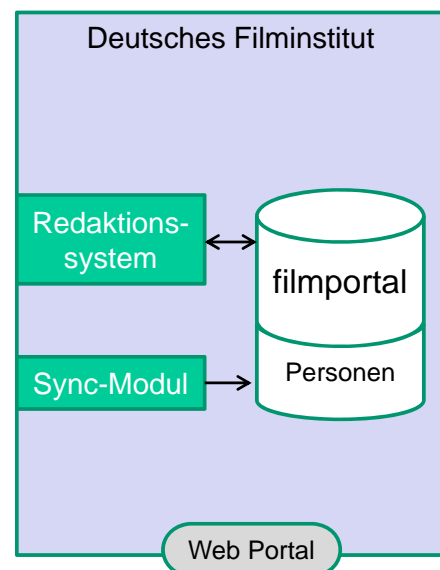
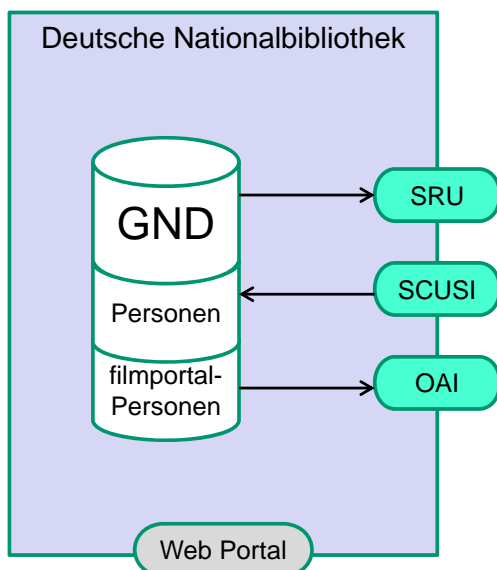
SCUSI Request: GND-Eintrag ändern

```
POST /scusi/v1/gnd/129952788?_method=patch
Host: services.dnb.de
Content-Type: application/json
Accept: application/json
Authorization: Token 585e02d222f908fe37f6c3407b7a8c4c
```

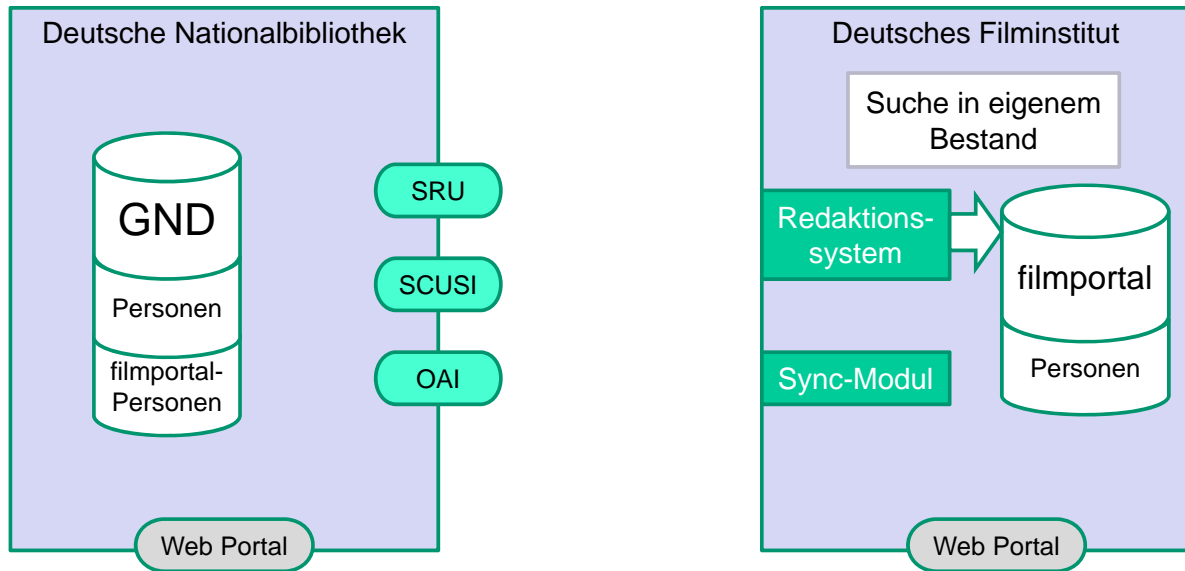
```
{
  "@Vocabulary" : "gnd",
  "operations" : [
    {
      "in" : {
        "NameOfThePerson" : {
          forename : "Wolke A.",
          surname : "Hegenbarth"
        }
      },
      "replace" : {
        forename : "Wolke A."
      },
      "with" : {
        forename : "Wolke Alma"
      }
    },
    {
      "remove" : {
        "placeOfBirthAsLiteral" : "Meerbusch, Deutschland"
      }
    },
    {
      "add" : {
        "placeOfBirth" : "http://d-nb.info/gnd/2029013-5"
      }
    },
    {
      "replace" : {
        "dateOfBirth" : "1980"
      },
      "with" : {
        "dateOfBirth" : "1980-05-06"
      }
    }
  ]
}
```

Ändere in dem GND-Eintrag mit der GND-Nummer 129952788 in „Wolke Alma Hegenbarth“ den Vornamen von „Wolke A.“ auf „Wolke Alma“. Ferner ändere den Geburtsort von „Meerbusch, Deutschland“ auf die URI „http://d-nb.info/gnd/2029013-5“. Ersetze außerdem das Geburtsdatum „1980“ durch „1980-05-06“.

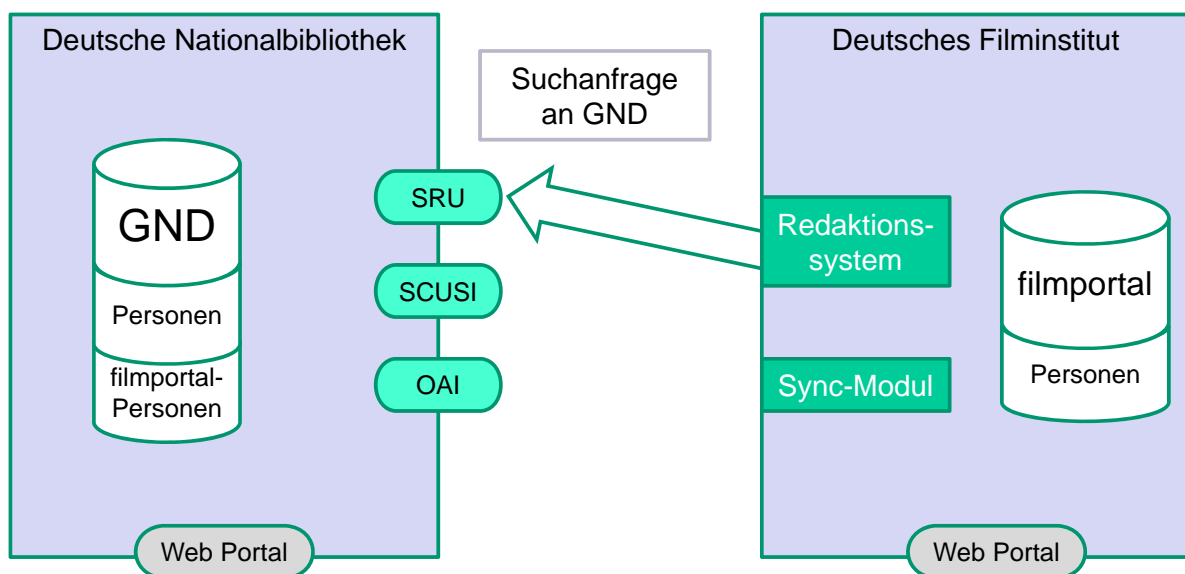
Nutzung der Schnittstellen



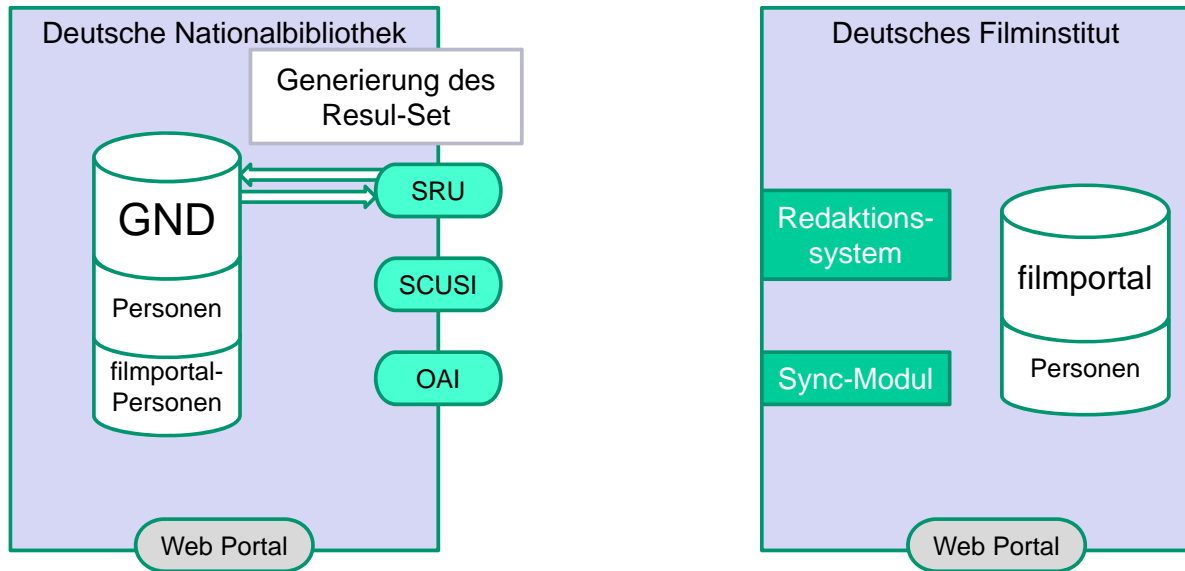
Nutzung der Schnittstellen – Suche & Übernahme



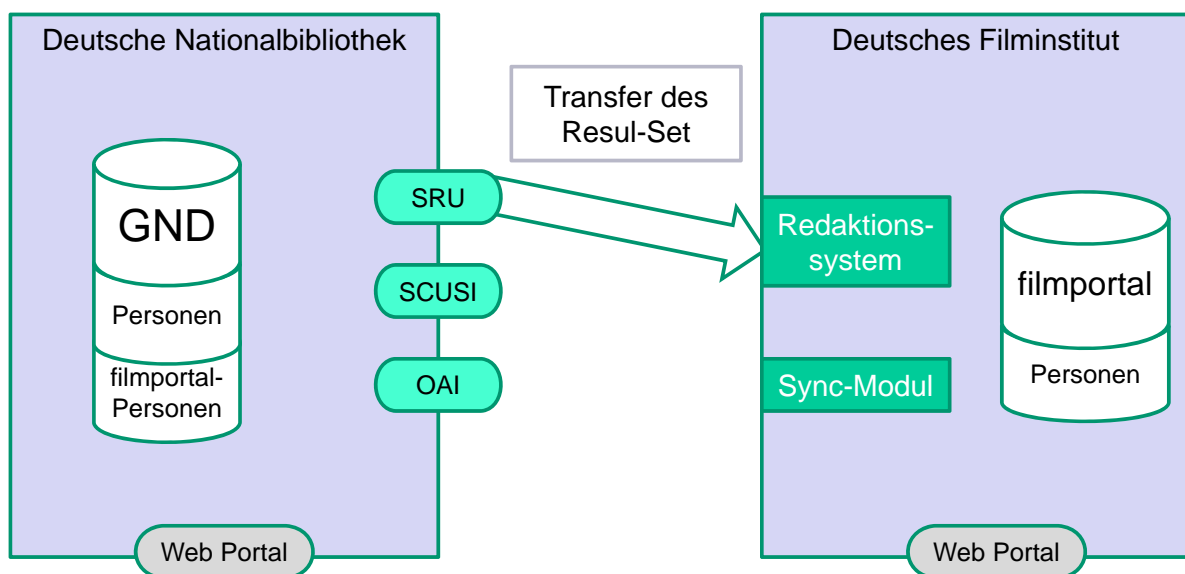
Nutzung der Schnittstellen – Suche & Übernahme



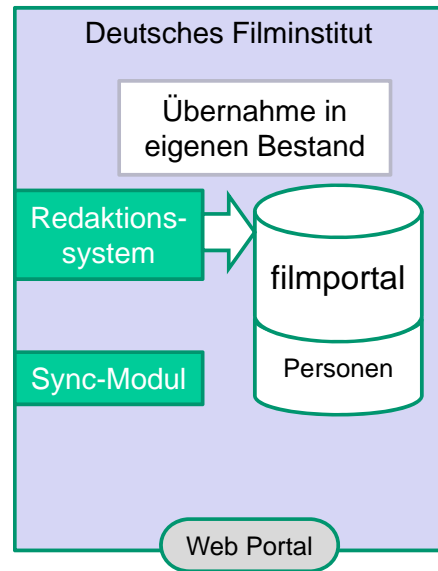
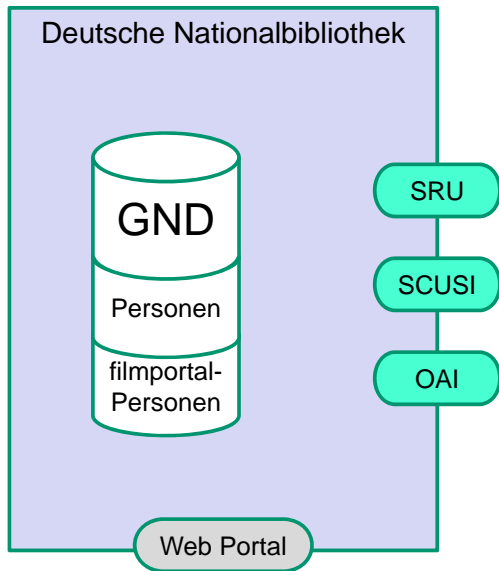
Nutzung der Schnittstellen – Suche & Übernahme



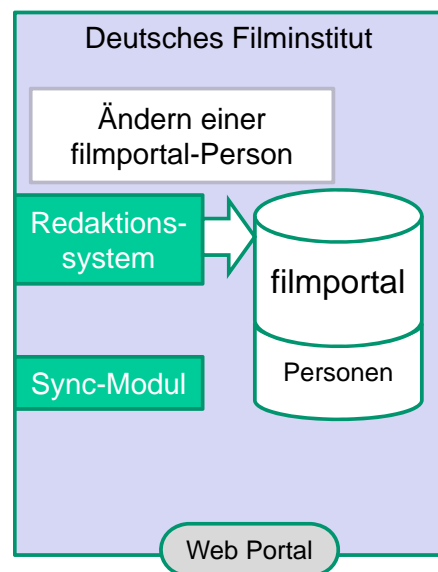
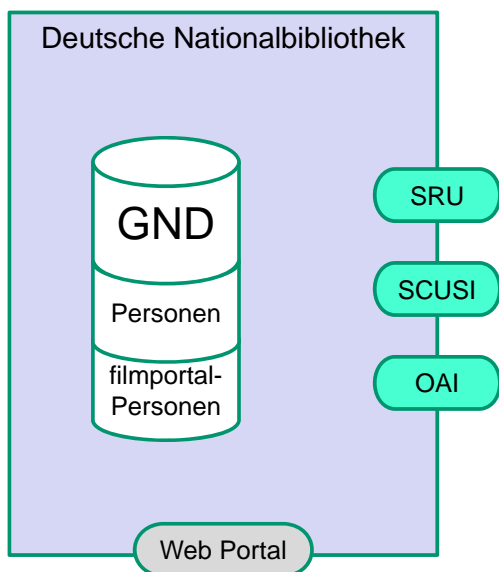
Nutzung der Schnittstellen – Suche & Übernahme



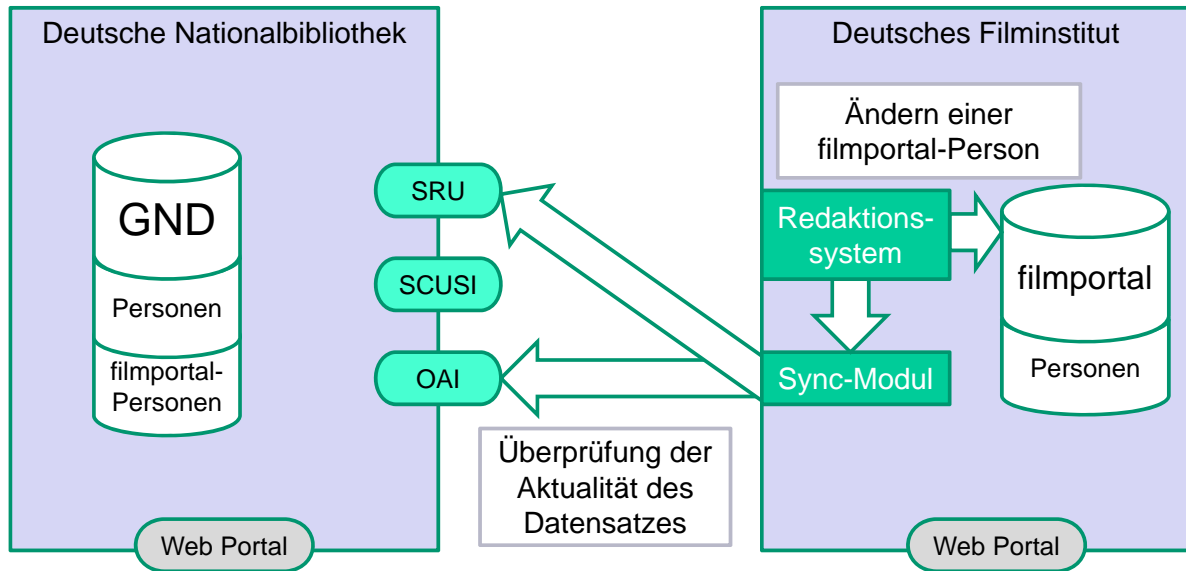
Nutzung der Schnittstellen – Suche & Übernahme



Nutzung der Schnittstellen – filmportal-Änderungen

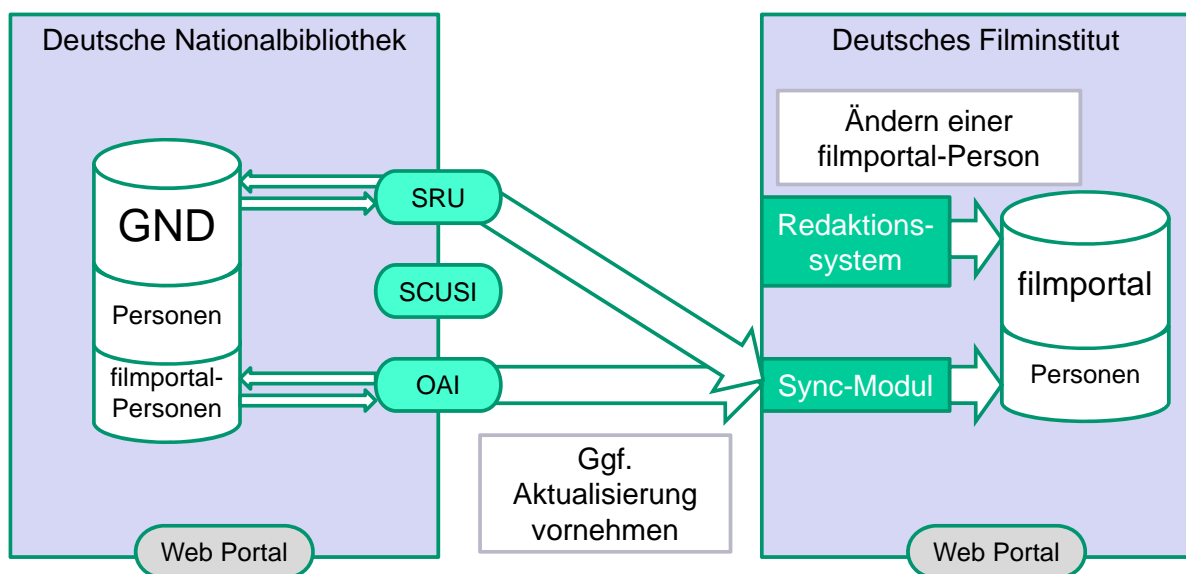


Nutzung der Schnittstellen – filmportal-Änderungen



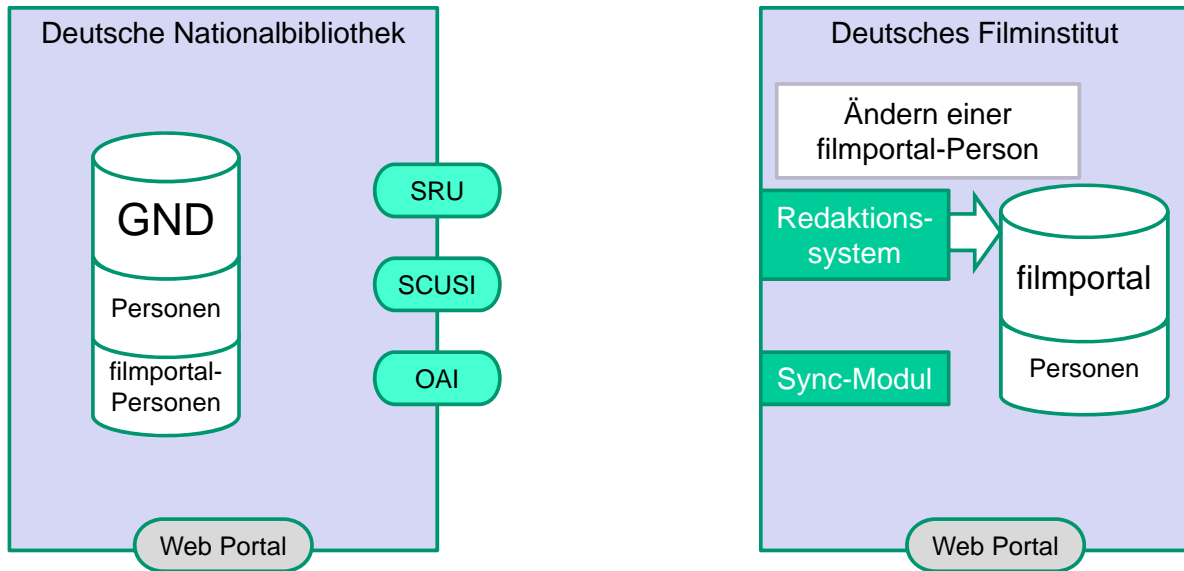
IN2N-Workshop | 30.09.2014

Nutzung der Schnittstellen – filmportal-Änderungen

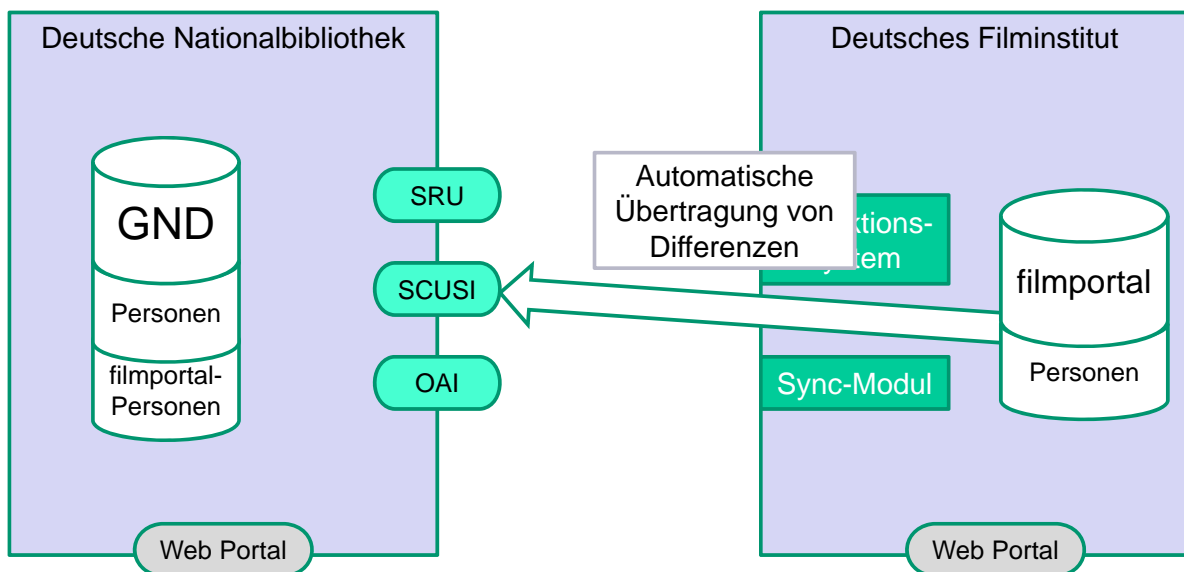


IN2N-Workshop | 30.09.2014

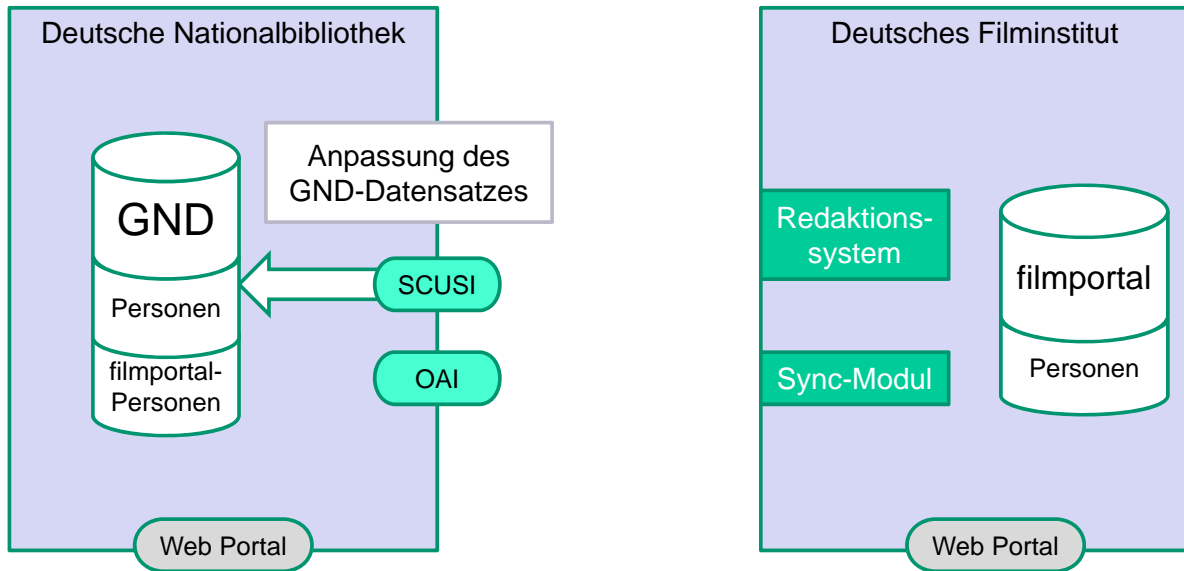
Nutzung der Schnittstellen – filmportal-Änderungen



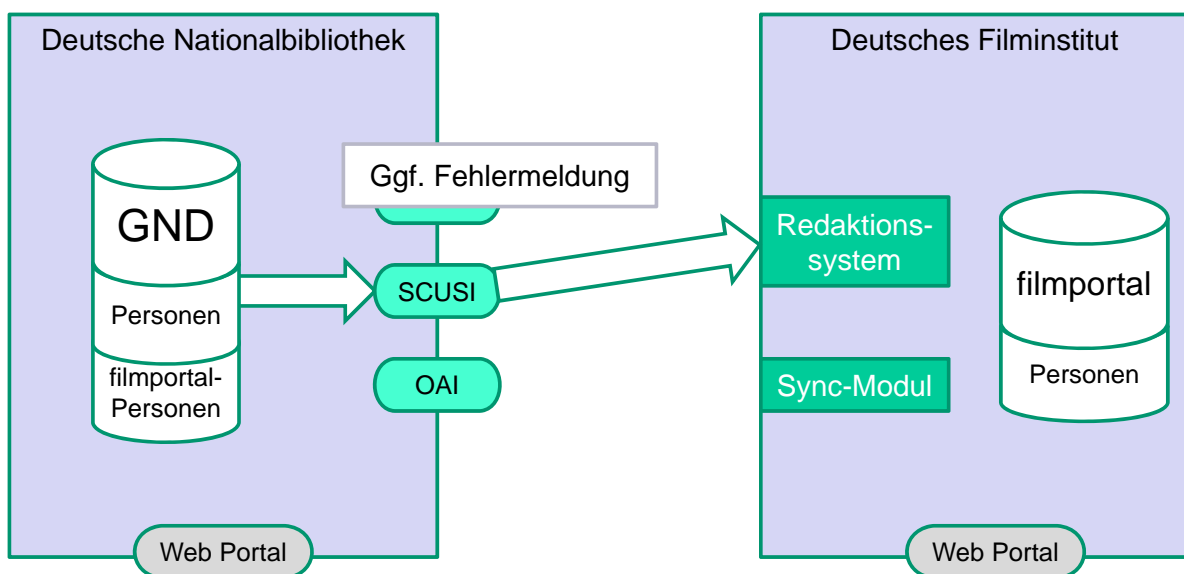
Nutzung der Schnittstellen – filmportal-Änderungen



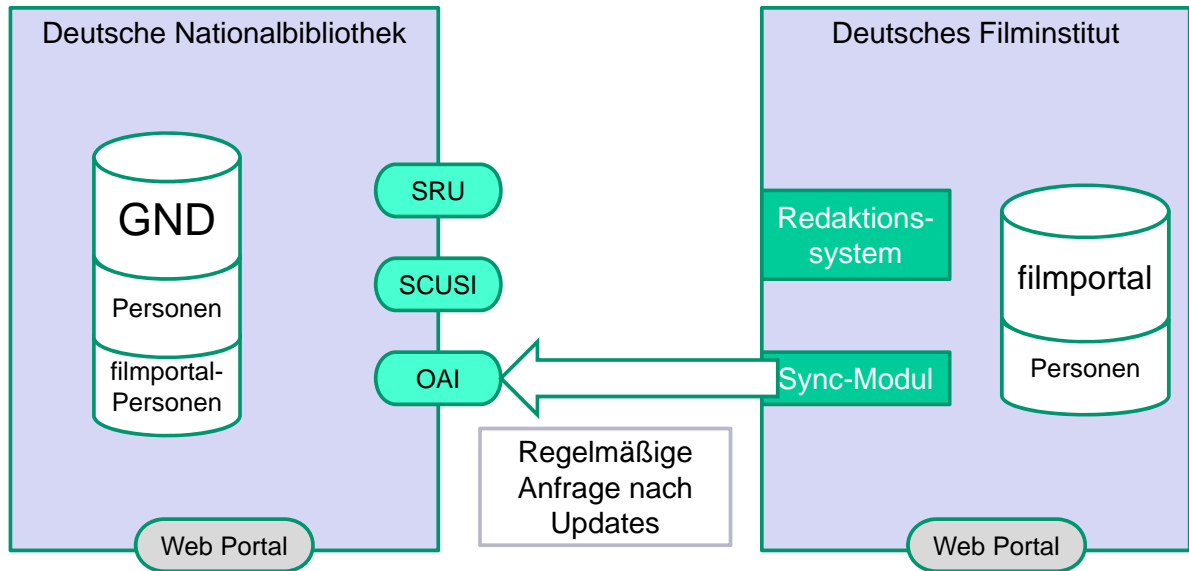
Nutzung der Schnittstellen – filmportal-Änderungen



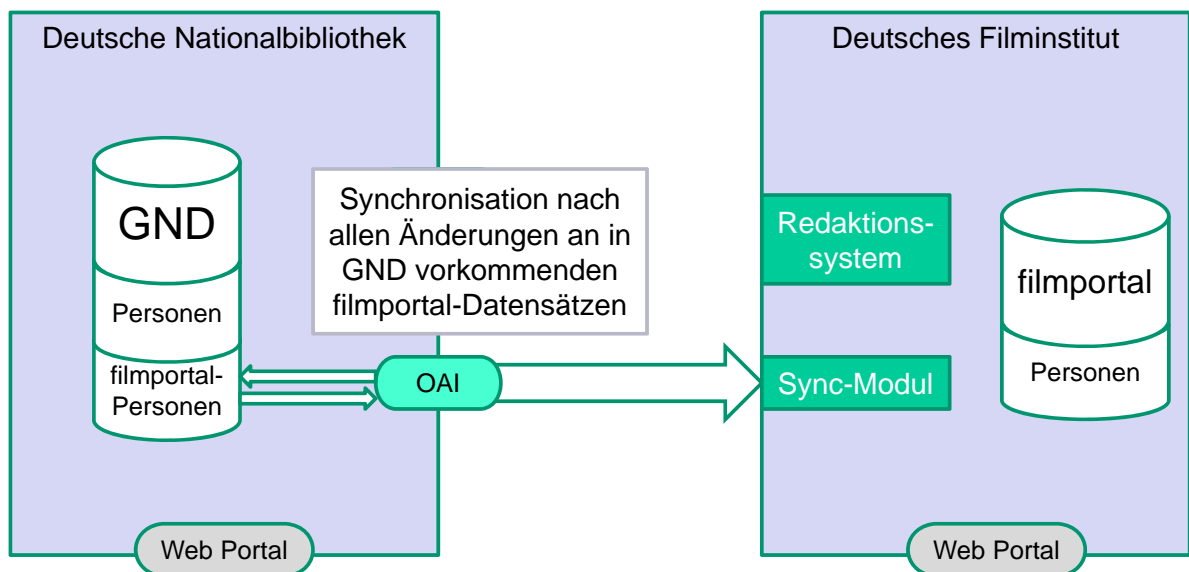
Nutzung der Schnittstellen – filmportal-Änderungen

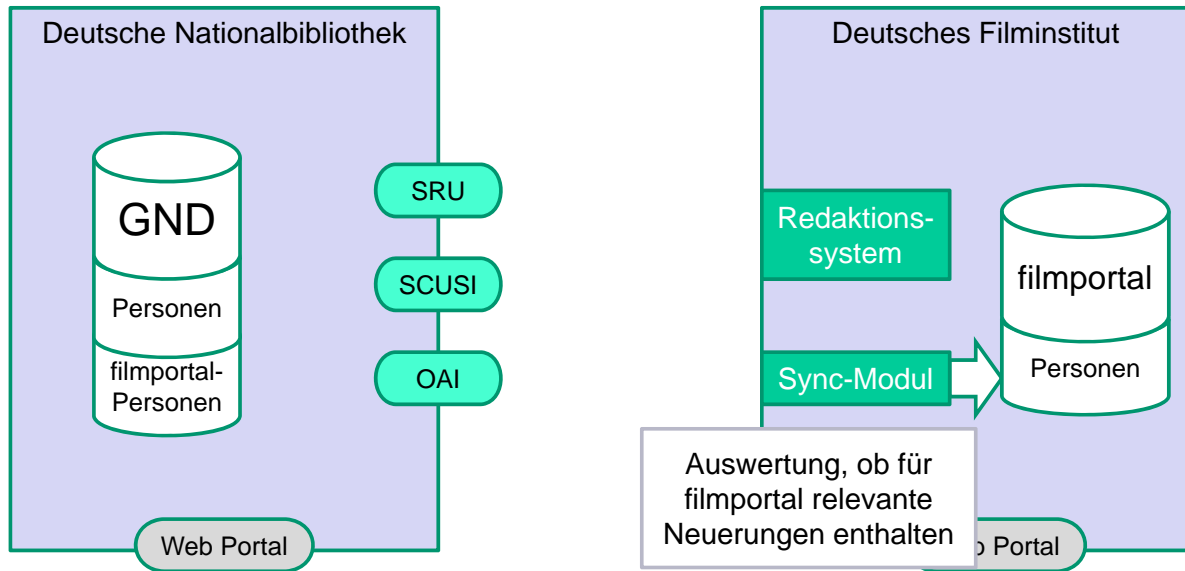


Nutzung der Schnittstellen – Synchronisation



Nutzung der Schnittstellen – Synchronisation





filmportal-Redaktionssystem

- Neue Features
 - Suchen nach GND-Personen
 - Anzeige einzelner GND-Personen
 - Übernahme von Informationen einer GND-Person
 - Optimierung der Darstellungsform
 - Autovervollständigung für kontrollierte Terme
 - Anlegen von Beziehungen zwischen Personen sowie zwischen Personen und Körperschaften
 - Link auf GND-Portal
- Assoziierte Features
 - Einführung von Perma-Links
 - Erweiterung des Austauschformates EAC-CPF

Deutsches Filminstitut (DIF): **Zentrale Filmografie**

Bearbeiter: **sgd**

[Login](#)
[Übersicht](#)
[Suchen](#)
[Status-DB](#)
[KIM-DB](#)
[MIDAS](#)
[Wiki](#)
[Bearbeiter](#)

Person
[Person anlegen]: Neue Person

Neue Person in der GND suchen

Suchen nach:

Adorf, Hans-Jürgen Führungskraft, Geschäftsführer
Adorf, Hendrik (1980-) Physiker
Adorf, Johannes (-1505) Theologe, Pfarrer
Adorf, Margarete Musikern
Adorf, Mario (1930-) Schauspieler, Musiker
Aus: DNB Gemeinsame Normdatei (GND); ID: 119066564 [Importieren](#)
Vorzugsname: **Mario | Adorf**
Geschlecht: **male**
Geboren: 1930-09-08 in Zürich
Berufe: **Schauspieler; Musiker**
Wirkungsort: **München**
Familiäre Beziehung zu: **Verhoeven, Paul**
Andere Quellen: **Wikipedia: Mario_Adorf; dbpedia: Mario_Adorf; VIAF: 84954675**

Adorf, Sigrid (1972-) Kunsthistorikern
Adorf, Stella Maria (1963-) Schauspielerinn
Atorp, Johann (1675-1780) Papiermacher
Johannes (um 1491) Exerctium veteris artis
Leo, Valentin (~1595) Adorf; Beitr. im VD-16
Simon-Adorf, Ute (1963-) Beraterin, Autorin
Unbekannt (~1931-1934)

Prefix	Zeit	Region

[Mehr ...](#)



Zusammenfassung

Was wurde erreicht

- Modularisierter Workflow für den initialen Datenabgleich und die initiale Datenübernahme
 - elaboriertes Abgleichverfahren
 - Unterstützungsmöglichkeit für Redakteure bei unsicheren Entscheidungen der Maschine
- Aufbau einer Infrastruktur für den webbasierten Datenaustausch auf Linked-Data-Basis
 - Anbindung beliebiger Redaktionssysteme aus nicht-bibliothekarischem Umfeld
- Vorbereitung eines generalisierten Kooperationsmodells für die institutionenübergreifende Integration von Nomdaten



Weiteres Vorgehen



Was kommt nach dem IN2N-Projekt

- 2015 Pilotbetrieb mit dem DIF
 - Auftretende Probleme in der Redaktion identifizieren
 - Stabilität der Schnittstellen eruieren
 - Performance evaluieren
- 2015 Kooperationsmodell für Aufnahme weiterer Partner festlegen und Geschäftsgänge etablieren
- Ende 2015 Akquise neuer Kooperationspartner
- 2015 ff. Integration weiterer Partner

IN2N-Workshop | 30.09.2014



Vielen Dank!

discussion is welcome...

www.in2n.de

Alexander Haffner
Deutsche Nationalbibliothek
a.haffner@dnb.de