

# Der Langzeitarchivierungsservice der Deutschen Nationalbibliothek

Karlheinz Schmitt

18.06.2013

**W**arum ein Langzeitarchivierungsdienst?

**W**as kann der Dienst für Sie leisten?

**W**ann steht der Dienst zur Verfügung?

---

**W**ieviel kostet der Langzeitarchivierungsdienst?

# W arum ein Langzeitarchivierungsdienst?

Digitale  
Langzeitarchivierung  
an der DNB





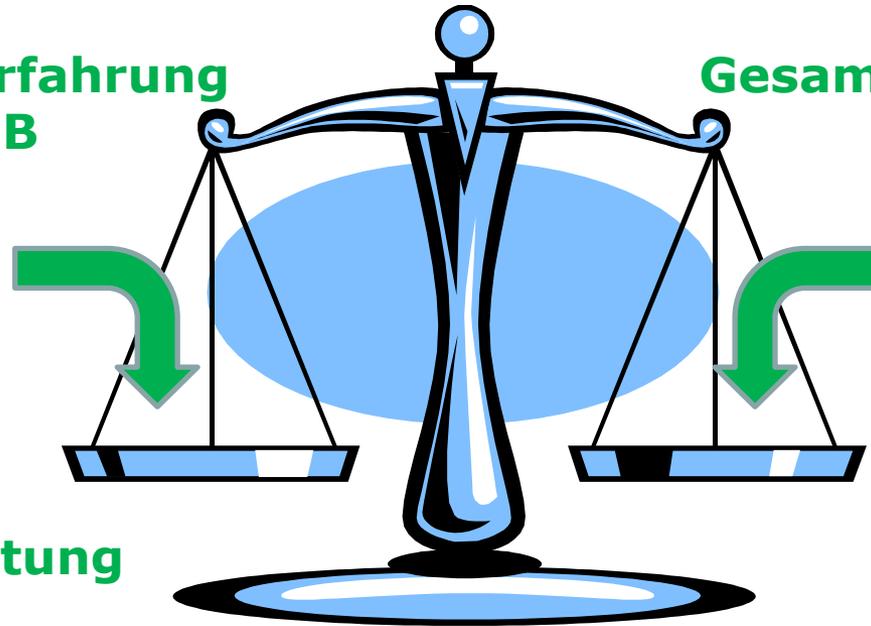
## Notwendigkeiten

- **Erfahrungsaustausch**
- **Zusammenarbeit**
- **Grenzen überwinden**

## Kooperative Langzeitarchivierung

Gesammelte Erfahrung  
der DNB

Gesammelte Erfahrungen  
der Nutzer



als Dienstleistung

als Gegenleistung



## Der Service-Katalog

### Basis-Dienst:

- Bereitstellung von Massenan- und -auslieferungsschnittstellen
- Integritätsprüfung aller digitalen Objekte
- Generierung technischer Metadaten
- Qualitätsprüfung (technisch) aller digitaler Objekte
- Ortsunabhängige Mehrfachspeicherung
- Bereitstellung von Zugriffs- und Suchfunktionen
- Bereitstellung von Kontrollmöglichkeiten
- Preservation Planning: Risikomanagement



## Automatisierte Massenschnittstellen

### ● Anlieferung:

- Hotfolder
- OAI-PMH-Protokoll

### ● Auslieferung:

- SOAP-Schnittstelle
- REST-Schnittstelle
- WEB-GUI (Benutzerschnittstelle)



## Was kann der Dienst für Sie leisten?

### Austauschobjekte

#### Anlieferungspakete:

- Beliebige Zusammenstellung zu „semantischen Einheiten“
- Beliebige Dateiformate
- Mitlieferung von deskriptiven Metadaten möglich
- Checksummen pro digitales Objekte und/oder pro Anlieferungspaket

#### Rücklieferungsobjekte

Je nach Bedarf können einzelne Archivpakete, oder ganze Migrationshistorien samt generierte Metadaten zurückgeliefert werden.



# Was kann der Dienst für Sie leisten?

## Rollendefinition

Im Zuge der  
Pflichtabgabe



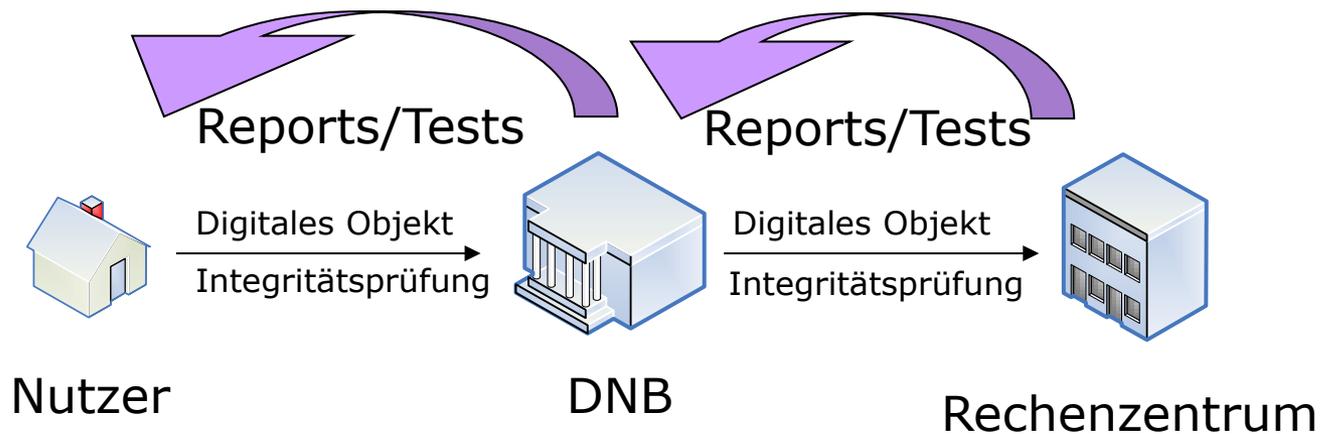
Separat  
(Reine Dienstleistung)

Auch das ist möglich, sprechen Sie uns an.

## Gemeinsame Sicherung des Bitstreams

- Integritätsprüfung aller digitaler Objekte, bei der Aufnahme und für die Dauer der Archivierung

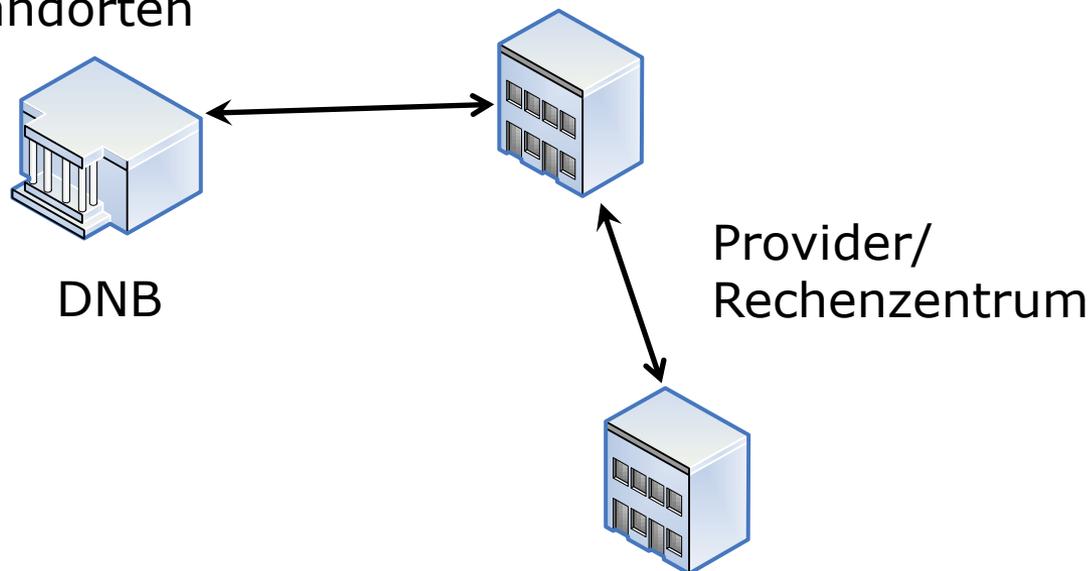
Muss über alle Stationen erfolgen:



## Gemeinsame Sicherung des Bitstreams

### ● Archivierung & Ausfallsicherheit

- Mehrfachspeicherung
- An mehreren Standorten



- Statusreports
- Backuptests

## Generierung technischer Metadaten

- FITS-Framework
- didigo-Tool (Eigenentwicklung)
- aus Informationssystemen

Aufgestellte  
Metadaten-Kernsets  
pro Dateiformat



Generierte  
technische Informationen



**Normalisierung der Ergebnisse**



# Risikomanagement

## Schritt 1: Das Erkennen von Risiken

### 1.) Technische Qualitätsprüfung

- Anhand aller technischer Metadaten und
- Vereinbarter gemeinsamer Workflows

### 2.) Klassifizierung in Ingest-Level

Ein Ingest-Level repräsentiert das Ergebnis eines mehrstufigen Prüfverfahrens, welches zurzeit durch mehrere aufeinander aufbauende Dienste innerhalb der DNB realisiert wird.



**Was kann der Dienst für Sie leisten?**

**Risikomanagement**

# **Ingest-Level - Erklärung**

Der Ingest-Level ist eine Abschätzung bzgl. der wahrscheinlichen Langzeitarchivierbarkeit digitaler Objekte, die auf Grundlage momentan zur Verfügung stehender technischer Mittel und den etablierten Geschäftsprozessen der DNB getroffen wird. Der Ingest-Level stellt jedoch keine Aussage über die allgemeine Qualität der geprüften digitalen Objekte dar.



**Risikoabschätzung**



## Schritt 2: Das Einleiten von Gegenmaßnahmen am Beispiel der DNB

### Filterung

	DI	ID	BF	MD	V
Level 0	X	O	O	O	O
Level 1	X	X	O	O	O
Level 2	X	X	X	O	O
Level 3	X	X	X	X	O
Level 4	X	X	X	X	X



**as kann der Dienst für Sie leisten?**

**Vertrauen gut – Kontrolle besser!**

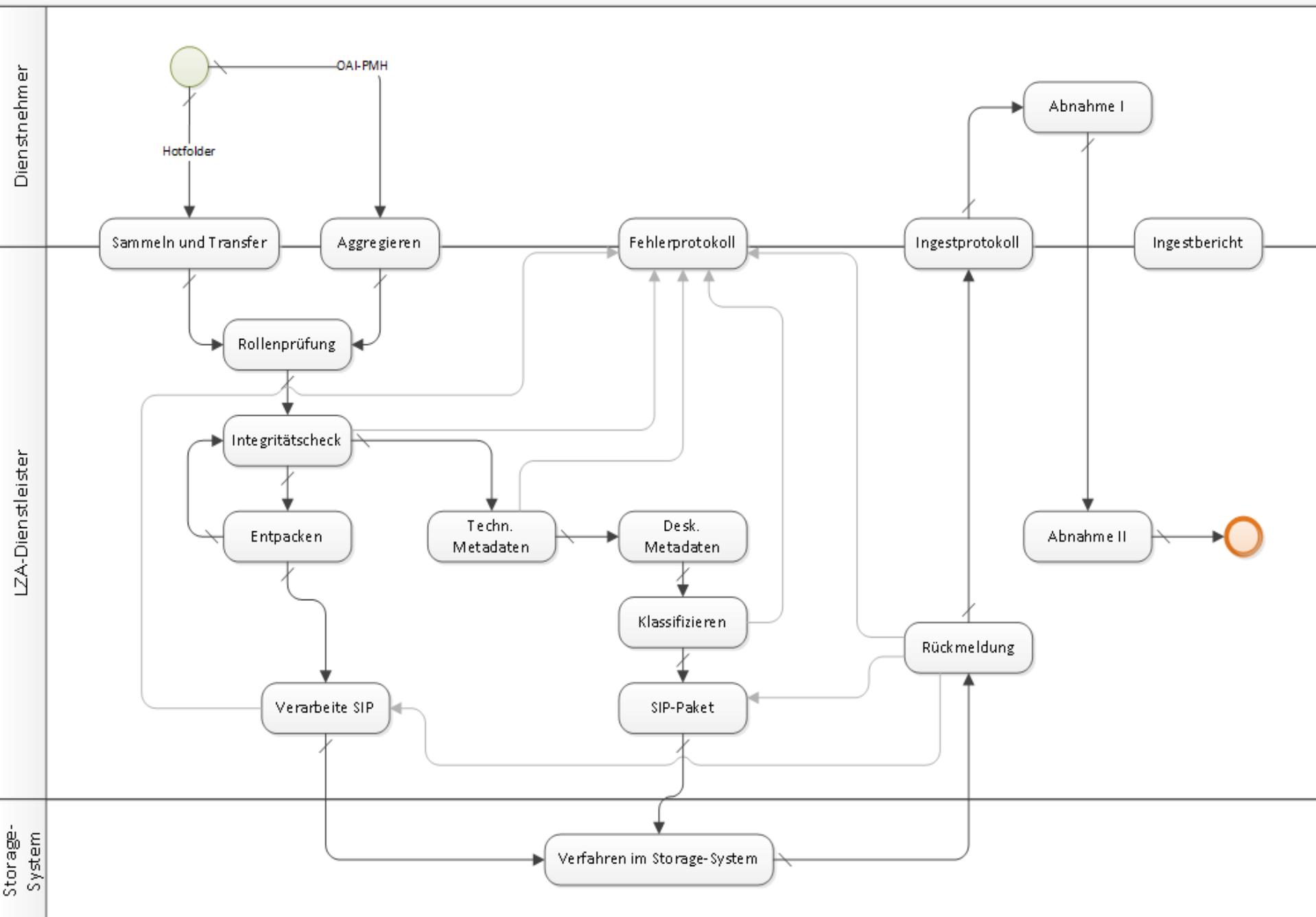


**Berichtswesen**



Lückenloses Berichtswesen soll den Aufbau einer vertrauensvollen Zusammenarbeit unterstützen.

Workflow-Konzept zum Ingest digitaler Objekte in ein Langzeitarchiv – vereinfachte Darstellung





# Was kann der Dienst für Sie leisten?



## Ingest-Verlauf

Sie sehen hier den aktuellen Verlauf einer Archivierung durch den Ingest-Workflow. Sollten Fehler oder Probleme bei gewissen Schritten innerhalb der Prozesskette aufgetreten sein, werden diese hier vermerkt.

(ID: 20120214\_11\_err.zip)

ID: 20120214\_11\_err.zip

### Ausgabe

Modul	Event-Information
<b>Start</b>	<b>Startzeit: 14.02.2012 14:59:57</b>
<b>ZipChecksumCheck</b>	Transferpaket: 20120214_11_err.zip abgeschlossenen am: 14.02.2012 14:59:57 Paketbestandteil: 20120214_11_err.zip Status: DONE Modulnachricht: package file 20120214_11_err.zip is valid
<b>Unpack</b>	Transferpaket: 20120214_11_err.zip abgeschlossenen am: 14.02.2012 14:59:58 Paketbestandteil: 20120214_11_err.zip Status: DONE Modulnachricht: package file 20120214_11_err.zip successful unpacked
<b>ItemChecksumCheck</b>	Transferpaket: 20120214_11_err.zip abgeschlossenen am: 14.02.2012 15:00:04 Paketbestandteil: 20120214_11_err.zip Status: DONE Modulnachricht: alle Paketbestandteile sind valide
<b>DcMetadataExtractor</b>	Transferpaket: 20120214_11_err.zip abgeschlossenen am: 14.02.2012 15:00:06 Paketbestandteil: 20120214_11_err.zip Status: DONE Modulnachricht: DC/Metadaten erfolgreich verarbeitet
<b>DuplicationCheck</b>	Transferpaket: 20120214_11_err.zip abgeschlossenen am: 14.02.2012 15:00:09 Paketbestandteil: 20120214_11_err.zip Status: DONE Modulnachricht: 94899d8c1cc84e999199e4357e7970388c2f64402
<b>Difigo</b>	Transferpaket: 20120214_11_err.zip abgeschlossenen am: 14.02.2012 15:00:14 Paketbestandteil: 20120214_11_err.zip Status: DONE Modulnachricht: alle Paketbestandteile erfolgreich geprüft
<b>SubmitErrorProtocol</b>	Transferpaket: 20120214_11_err.zip abgeschlossenen am: 14.02.2012 15:00:15 Paketbestandteil: UZL_UP_19_11_11.pdf Status: ERROR Modulnachricht: Ingest-Policy: IngestPolicy für CP4lib-Verlag (10185000) - Verfügformid:10185000, Formid:10185000, Verfügformid:10185000 Ingest-Modul 3 für Format:fm10 liegt unter der Policy: Vergabe (4) - Abbruch der Verarbeitung - error:Message: No file trailer offset=81942880
<b>SubmitErrorProtocol</b>	Transferpaket: 20120214_11_err.zip abgeschlossenen am: 14.02.2012 15:00:15 Paketbestandteil: UZL_UP_19_11_11.pdf Status: DONE Modulnachricht: Error-Protokoll an Https://Op4lib.longsight.com/ingest/ (reportable) versendet. message: protocol_type=error; delivery_role=CP4lib; to_name=20120214_11_err.zip; error_module=IngestLevelCheck; error_message=Ingest-Policy: IngestPolicy für CP4lib-Verlag (10185000) - Verfügformid:10185000, Formid:10185000, Verfügformid:10185000 Ingest-Modul 3 für Format:fm10 liegt unter der Policy: Vergabe (4) - Abbruch der Verarbeitung - error:Message: No file trailer offset=81942880

**Ingest-Verlauf**

Sie sehen hier den aktuellen Verlauf einer Archivierung durch den Ingest-Workflow. Sollten Fehler oder Probleme bei gewissen Schritten innerhalb der Prozesskette aufgetreten sein, werden diese hervorgehoben.

ID: 20120214\_11\_err.zip

**Ausgabe**

Didigo

Transferpaket: 20120214\_11\_err.zip  
 abgeschlossen am: 14.02.2012 15:00:14  
 Paketbestandteil:  
 Status: DONE  
 Modulenachricht: alle Paketbestandteile erfolgreich gepueft

IngestLevelCheck

Transferpaket: 20120214\_11\_err.zip  
 abgeschlossen am: 14.02.2012 15:00:15  
 Paketbestandteil: LVZ\_LPZ\_19\_11\_11.pdf  
 Status: ERROR  
 Modulenachricht: Ingest-Policy: IngestPolicy für DP4lib-Verlag (151552000) - Wohlgeformtheit bzw. Formatvalidität verletzt -> ermitteltes Ingest-Level 3 fuer Format fmt/18 liegt unter der Policy-Vorgabe (4) - Abbruch der Verarbeitung - error-Messages: No file trailer offset=62942880

SubmitErrorProtocol

Transferpaket: 20120214\_11\_err.zip  
 abgeschlossen am: 14.02.2012 15:00:15  
 Paketbestandteil: LVZ\_LPZ\_19\_11\_11.pdf  
 Status: DONE  
 Modulenachricht: Error-Protokoll an <https://dp4lib.langzeitarchivierung.de/report.php> versendet. message: protocol\_type=error; delivery\_role=DP4lib; tp\_name=20120214\_11\_err.zip; error\_module=IngestLevelCheck; error\_message=Ingest-Policy: IngestPolicy für DP4lib-Verlag (151552000) - Wohlgeformtheit bzw. Formatvalidität verletzt -> ermitteltes Ingest-Level 3 fuer Format fmt/18 liegt unter der Policy-Vorgabe (4) - Abbruch der Verarbeitung - error-Messages: No file trailer offset=62942880

## ACCESS

## DP4lib - LTP Access Service

**Liste aller abgelieferten Objekte**

Sie sehen hier alle von Ihnen abgelieferte Objekte und haben die Möglichkeit sich weitere Informationen anzeigen zu lassen.

angemeldet als DP4lib-Verlag  
(ID: 151453696)

**Ausgabe**

urn	Ingest-Transaktions-ID	Eingangsdatum	Transferpaketname	Status	Retrieval
<a href="#">urn:nbn:de:101:1-201202141614</a>	1329242058461479	14.02.2012 18:54:21	20120214_17.zip		
<a href="#">urn:nbn:de:101:1-201112254501</a>	13292286364151323	14.02.2012 15:10:41	20120214_11.zip		

Bei Fragen wenden Sie sich bitte an Stefan Hein (s.hein@dnb.de) oder Karlheinz Schmitt (k.schmitt@dnb.de)  
Version 0.9.4 (13.02.2012)



## Detailinformationen für Archiveinheit

Sie sehen hier Detailinformationen (bibliografische Metadaten und Transaktionen) für eine konkrete Archiveinheit.

**ID: urn:nbn:de:101:1-201112254501**

### Ausgabe

#### bibliografische Metadaten

Element	Wert
Creator	[]
Title	[Leipziger Volkszeitung / 118. Jahrgang, Nr. 269]
Identifizier	[NP997738642_20111119_7, urn:nbn:de:101:1-201112254501]
Date	[2011-11-19]
Language	[de]
Subject	[Ausgabe Stadt Leipzig, Leipzig, Sachsen, Deutschland]
Format	[application/pdf]
Rights	[Deutsches Urheberrecht]
Type	[Tageszeitung, Online-Ressource, Elektronische Ressource]
Source	[]
Contributor	[]
Description	[]
Coverage	[]
Relation	[Druckausg.10128559951]
Publisher	[Leipzig: Leipziger Verl.-u.-Dr.-Ges.]

#### INGEST-Transaktionen

Nr.	Transaktions-ID	Datum	Transferpaketname	Dateiliste	Retrieval
1	<a href="#">13292286364151323</a>	14.02.2012 15:10:41	20120214_11.zip	<a href="#">Dateiliste zeigen</a>	

#### ACCESS-Transaktionen

Nr.	Transaktions-ID	Datum	Status
1	1329235076992	14.02.2012 16:57:55	
2	1329234968876	14.02.2012 16:56:08	

#### MIGRATION-Transaktionen

Dublin Core  
Metadaten

Transaktionen



## Der Service-Katalog

### Basis-Dienst: (Grundfunktionalitäten)

- Ingest
- Curation
- Access

### Optionale Erweiterungen

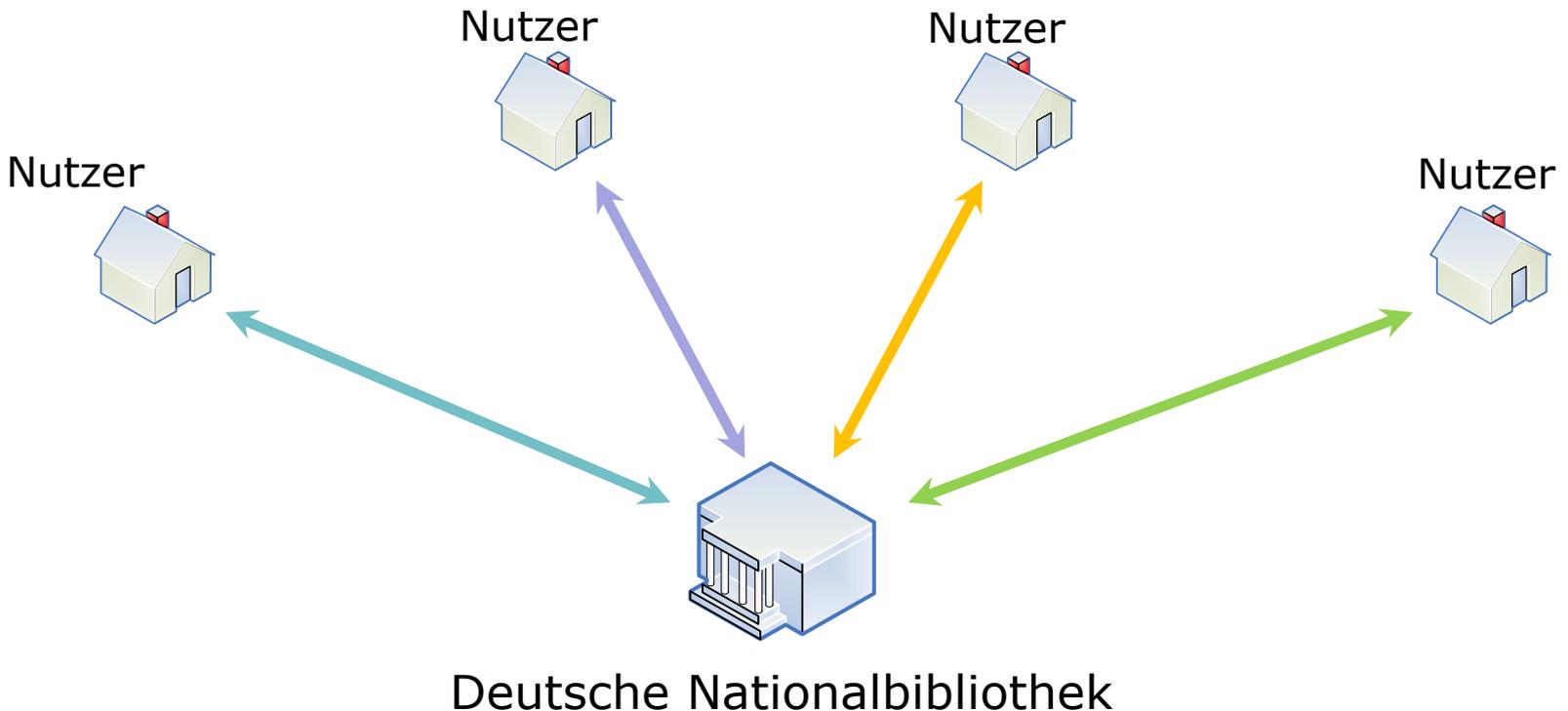
#### Ingest:

- Verarbeitung von DC-Simpe-Dateien
- Verarbeiten weiterer Zusatzinformationen

#### Curation:

- Ausführliche Statusreports

**Mit jedem Nutzer kann im Rahmen des Service-Katalogs  
ein individueller LZA-Workflow  
vereinbart und umgesetzt werden**





## Preis der Dienstleistung

Durch das Kostenmodell ist die DNB in der Lage eine auf einen Kunden speziell angepasste Leistungsverrechnung durchzuführen.

Abhängig vom individuell zusammengestellten Langzeitarchivierungsdienst für einen Dienstnehmer, wird nach einer gemeinschaftlich festgelegten Einführungsphase der jährliche Preis festgelegt.



## Momentan zu berechnende Kosten:

- Der Nutzer hat die Möglichkeit Speicherkontingente bei der DNB zu mieten.
- Pro gemietetes GB fallen pro Jahr zur Zeit **1,49 €** Kosten an, die berechnet werden müssen.
- Leistungen, die vom Nutzer zur kooperativen LZA eingebracht werden, können in die gegenseitige Leistungsverrechnung eingehen

**W**ann steht dieser Dienst zur Verfügung?

**Ab sofort!**

Sprechen Sie uns an!

Karlheinz Schmitt / [k.schmitt@dnb.de](mailto:k.schmitt@dnb.de)