

## **DiPS.kommunal**

### **(Digital Preservation Solution)**

#### Leistungsbeschreibung

21.11.2016

## Dokumentenstatus

<b>Erstellt am</b>	06.03.2015
<b>Zuletzt geändert</b>	21.11.2016
<b>Bearbeitungszustand</b>	Zwischen LWL und StK abgestimmt
<b>Fertig gestellt</b>	21.11.2016
<b>Version/Stand</b>	Version: 1.1

## Änderungsverzeichnis

Version	Datum	Name	Änderung
0.1	31.03.2015	[REDACTED]	Initiale Version
0.2	27.04.2015	[REDACTED]	Überarbeitete Version
0.3	11.05.2015, 13.05.2015	[REDACTED]	Übernahme Änderungen LWL, weitere Anpassungen
0.4	19.05.2015 – 21.05.2015	[REDACTED]	Per StK und LWL Videokonferenz abgestimmte Änderungen, Formatlisten vereinheitlicht
1.1	21.11.2016	[REDACTED]	DiPS.kommunal Neuerungen ergänzt/aktualisiert

## Abnahmeverzeichnis

Version	Datum	Teilnehmer bei Doku-Review	Hinweise / Datum Doku-Review
1.0	22.05.2015	[REDACTED]	
	23.07.2015		Anpassung zu Produktname „DiPS.kommunal“

## Inhaltverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>4</b>
1.1	Leistungsübersicht	5
1.2	Datenschutzklassifizierung	6
<b>2</b>	<b>DiPS.kommunal - Funktionen und Komponenten</b>	<b>6</b>
2.1	Pre-Ingest / PIT.plus	7
2.1.1	Begriffsdefinitionen	8
2.1.2	Repräsentationen	10
2.1.3	Struktureller Aufbau des AIP	10
2.1.4	Lieferung	11
2.1.5	Übergabe der Lieferung	11
2.2	Ingest	11
2.2.1	Integritätsprüfung	12
2.2.2	Formaterkennung und -validierung	12
2.2.3	Metadatenergänzung	13
2.2.4	Bildung der AIPs und Export der Metadaten für Erschließungssoftware	13
2.3	Archivspeicher	13
2.4	Preservation Management	13
2.5	Benutzung	13
<b>3</b>	<b>Software</b>	<b>14</b>
3.1	Betriebssysteme	14
3.2	Java (JDK)	14
3.3	Ingest-Komponente (technisch)	15
3.4	DROID	15
3.5	Modul Fileextension	16
3.6	JHOVE	16
3.7	Pre-Ingest-Toolset (PIT.plus)	16
3.8	DOXIS 4	17
3.8.1	Webcube (Ingest)	17
3.8.2	Content Service Bus (CSB)	17
3.8.3	Workflow-Service	18
3.9	Rendition-Server	18
3.10	Erschließungssoftware / Archivverwaltungssoftware (Optional ACTApro)	18
3.11	Überwachung der eingesetzten Komponenten	18
<b>4</b>	<b>Qualitätssicherung</b>	<b>19</b>
4.1	Archivspeicher	19
4.2	Infrastruktur	19
4.2.1	Kommunikation innerhalb des Services	20
4.2.2	Ausfallsicherheit der Infrastruktur	20
4.2.3	Sicherungsstrategie der Server	20
4.3	Betrieb der technischen Infrastruktur	20
4.3.1	Rechenzentren	20
4.3.2	Wartung der Hard- und Software	21
4.3.3	Überwachung der Hardware	21
4.3.4	DiPS.kommunal Produktiv und Testumgebung	21
	Anhang 1	22

## 1 Einleitung

Unter der Federführung des Ministeriums für Familie, Kinder, Jugend, Kultur und Sport des Landes NRW (MFKJKS) wurde 2009 ein Projekt unter dem Namen „Digitales Archiv NRW“ initiiert, das eine träger-, sparten- und institutionsübergreifende kooperative Lösung für die dauerhafte Erhaltung und Bereitstellung des digitalen Kultur- und Archivgutes in NRW entwickeln soll. Die Lösung wird nach Fertigstellung von allen teilnehmenden Institutionen in NRW zur Sicherung ihrer digitalen Bestände genutzt werden können.

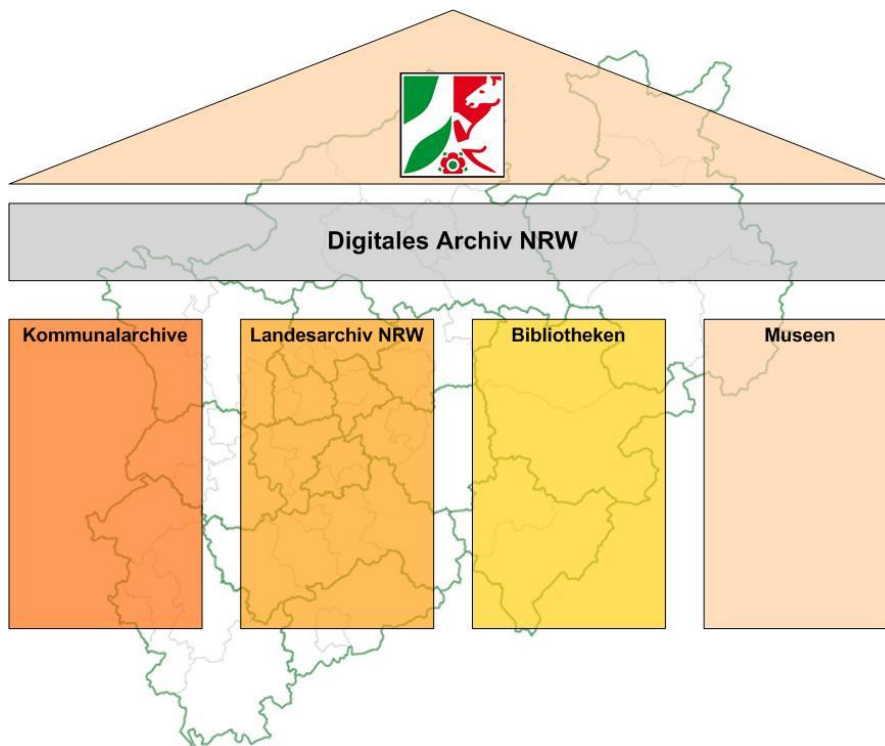


Abbildung 1: Lösungsverbund des Digitalen Archivs NRW

Seit März 2013 hat sich das Digitale Archiv NRW zu einem Lösungsverbund unterschiedlicher Anwendungen und Workflows zur langfristigen Erhaltung digitalen Kulturgutes im Land Nordrhein-Westfalen entwickelt. Unter einem einheitlichen Dach wird zum einen die „DA-NRW Software Suite“ (DNS)<sup>1</sup> Kulturinstitutionen angeboten. Zum anderen ist vorgesehen, die in den vergangenen Jahren von staatlichen und kommunalen Archiven gemeinsam mit den Dienstleistern HP und SER entwickelte System „Digital Preservation Solution“ (DiPS) im Lösungsverbund DA NRW anzubieten.

---

<sup>1</sup> Siehe DA NRW Basissystem – Leistungsbeschreibung,  
DA NRW DNS-System Betrieb - Leistungsbeschreibung



Abbildung 2: Gesamtstruktur des Digitalen Archivs NRW (Vgl. DA NRW – Organisationskonzept)

DiPS ist zunächst als Spartenlösung konzipiert worden. Das System orientiert sich konsequent an den Anforderungen der Archive an die elektronische Langzeitarchivierung. Bei der Entwicklung von DiPS wurden die rechtlichen Gesichtspunkte, welche im Gesetz über die Sicherung und Nutzung öffentlichen Archivguts (ArchivG NRW<sup>2, 3</sup>) geregelt sind, berücksichtigt.

## 1.1 Leistungsübersicht

DiPS.kommunal stellt eine mandantenfähige OAIS-konforme Verbundlösung zur elektronischen Langzeitarchivierung dar. Es ist in der Lage, strukturierte und unstrukturierte born-digitals zu übernehmen und langfristig zu erhalten. Es versetzt kommunale Archive in die Lage, ihre gesetzlichen Fachaufgaben zu erfüllen.

Zu den Kernkomponenten gehören:

- revisionssichere Speicherung
- sichere Übertragungswege und Datenverarbeitung
- Schulungs- und Beratungsangebot (inkl. Handbücher)
- Programm zur Strukturierung von Dateiablagen (PIT.plus)
- Übernahmemöglichkeit von externen Metadaten zu den Objekten
- XDomea 2.2.0 Standard-Schnittstelle zur Übernahme von Unterlagen aus aktenbildenden Systemen
- Integrationsmöglichkeit von Erschließungssoftwares

Optional und aufwandsbezogen sind Erweiterungen zur Anbindung weiterer DMS- und Fachverfahren zu implementieren, die sich über die vorhandenen Importschnittstellen noch nicht abbilden lassen. Da es hierbei ganz wesentlich auf die Quellsysteme, ihre Konfigurierbarkeit und ihre Exportmöglichkeiten ankommt, sind z.T. individuelle Lösungen umzusetzen.

<sup>2</sup> [https://recht.nrw.de/lmi/owa/br\\_vbl\\_detail\\_text?anw\\_nr=6&vd\\_id=12067](https://recht.nrw.de/lmi/owa/br_vbl_detail_text?anw_nr=6&vd_id=12067)

<sup>3</sup> Siehe DA NRW Organisationskonzept – Landesbereich

Die Prozesse in DiPS.kommunal orientieren sich an den aus dem analogen Bereich bekannten archivischen Arbeitsweisen. Dabei wird eine möglichst intuitive Handhabung angestrebt und Potentiale der elektronischen Arbeitsweise genutzt (z.B. durch die Nachnutzung von Metadaten der Registraturbildner für die archivische Erschließung).

## 1.2 Datenschutzklassifizierung

DiPS.kommunal ist freigegeben für Daten mit hohem Schutzbedarf. Hierdurch werden im Regelfall alle in der Kommunalverwaltung anfallenden Daten abgedeckt und können in das elektronische Langzeitarchiv übernommen werden. Die gesamte Architektur von DiPS.kommunal wurde strikt daraufhin entwickelt und bereits in der Umsetzungsphase durch stetige Sicherheitsanalysen der einzelnen Module vom entsprechenden Fachbereich geprüft.

## 2 DiPS.kommunal - Funktionen und Komponenten

Das System DiPS.kommunal ist modular aufgebaut und besteht aus folgenden Komponenten<sup>4</sup>:

- Pre-Ingest / PIT.plus
- Ingest
- Archivspeicher
- Benutzungsmodul / Archivsoftware
- Preservation Management

Der modulare Aufbau gewährleistet einen zukünftig flexiblen und herstellerunabhängigen Betrieb. Damit entspricht das System strikt den grundlegenden Standards der elektronischen Langzeitarchivierung nach ISO-Standard 14721:2012<sup>5</sup> (OAIS-Modell<sup>6, 7</sup>).

Eine Interoperabilität kann grundsätzlich über die Annahme von Daten aus weiteren Systemen hergestellt werden. Hierunter sind sowohl Quellsysteme der Verwaltung als auch Nachlässe und Sammlungen zu subsumieren.

---

<sup>4</sup> [Siehe DA NRW Referenzarchitektur – Technische Referenzarchitektur](#)

<sup>5</sup> [http://www.iso.org/iso/home/store/catalogue\\_tc/catalogue\\_detail.htm?csnumber=57284](http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=57284)

<sup>6</sup> <http://de.wikipedia.org/wiki/OAIS>

Das Modell des Archivinformationssystems „Open Archival Information System“ bzw. „Offenes Archiv-Informationssystem“ (OAIS) wurde von der NASA, gemeinsam mit der Raumfahrtorganisation ESA und Weltraumforschungszentren in Großbritannien, Kanada, Frankreich, Deutschland, Brasilien, Japan und Russland, entwickelt.

<sup>7</sup> Siehe DA NRW Projektleitfaden – Ziele

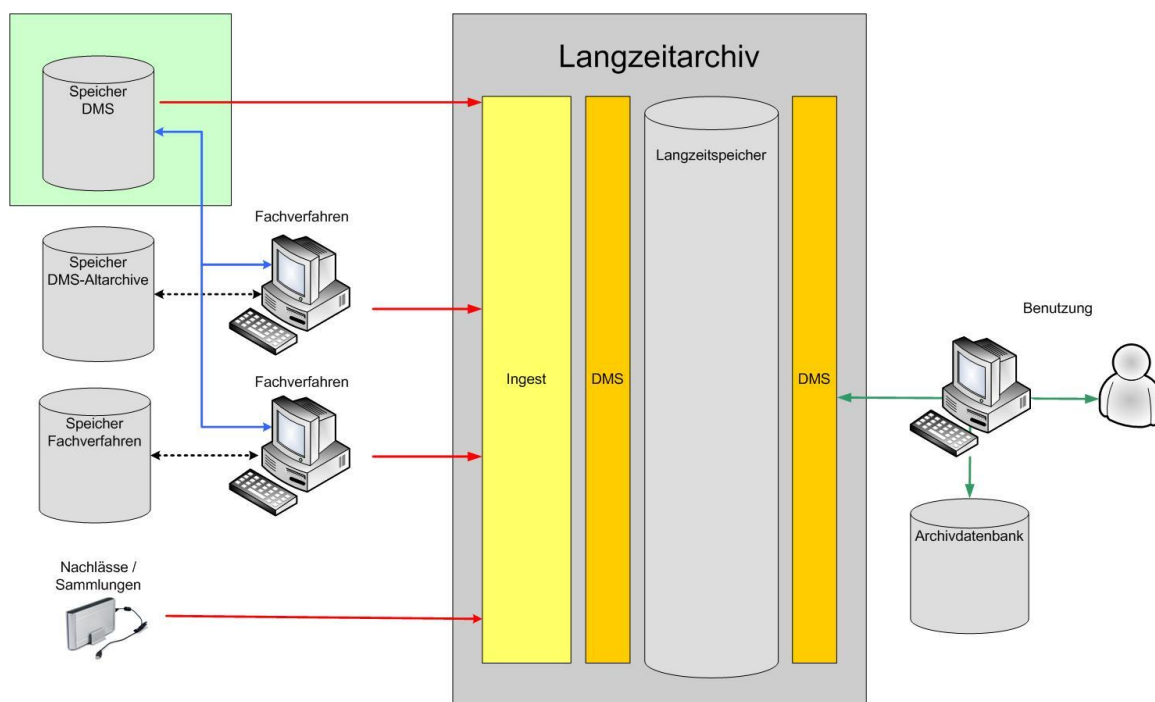


Abbildung 3: OAIS-konformer Aufbau und Funktionsweise von DiPS.kommunal nach ISO-Standard 14721:2012

## 2.1 Pre-Ingest / PIT.plus

Im Pre-Ingest-Prozess, welcher der Übernahme des digitalen Archivgutes in das elektronische Langzeitarchiv vorgeschaltet ist, werden mit Hilfe der Software PIT.plus (Pre-Ingest-Toolset) die Übernahmepakete, sogenannte Submission Information Packages (SIPs) gebildet. Die Bedienung der Software ist einfach und selbsterklärend. Der gesamte Funktionsumfang vom PIT.plus, sowie alle weiteren DiPS.kommunal Funktionalitäten, werden vor Betrieb innerhalb einer praktischen Anwenderschulung erläutert.

Im Pre-Ingest-Prozess gilt die Grundannahme, dass eine Inhaltliche Einheit eine Verzeichnungseinheit im archivischen Sinne ist und aus mehreren Repräsentationen bestehen kann (s.u.). Ein Submission Information Package (SIP) wird im Laufe des Ingest-Prozesses zu einem Archival Information Package (AIP). Ein AIP kann grundsätzlich in mehreren Repräsentationen vorhanden sein. Die Beziehung zwischen Inhaltlicher Einheit (IE) bzw. Verzeichnungseinheit zu den AIPs bzw. Repräsentationen ist eine 1:n Beziehung. Zusätzlich werden neben dem eigentlichen SIP auch Metadaten zur Übernahme generiert. Die Primärdaten bleiben dabei zur Wahrung der Integrität vollständig unverändert. Um sicherzustellen, dass die im Pre-Ingest-Prozess gebildeten Übernahmeeinheiten unverändert an den nächsten Prozess (Ingest) übergeben werden, wird für die gesamte Akzession zur Überprüfung ein Hashwert erzeugt.

### 2.1.1 Begriffsdefinitionen

Begriff	Erläuterung
IE (Inhaltliche Einheit)	Logische, inhaltliche (semantische) Einheit
Verzeichnungseinheit	Fachliche Beschreibung einer IE in der Erschließungssoftware
Repräsentation	Physische Manifestation / technische Ausprägung einer IE
SIP (Submission Information Package)	Im Pre-Ingest-Prozess oder bei der abgebenden Stelle erzeugtes Informationspaket, Vorprodukt für den Ingest
AIP (Archival Information Package)	Physisches Archivpaket, in Form eines unkomprimierten TAR-Containers, der alle Primär- und zum Verständnis notwendigen Metadaten enthält; Ergebnis des Ingest-Prozesses
DIP (Dissemination Information Package)	Physisches Benutzungspaket, das im Zuge der Access-Prozesse erzeugt und ausgeliefert wird
Object	Physisches Unterelement eines SIP oder eines AIP (i.d.R. eine Datei)
Item	Stück: Logische, inhaltliche Beschreibungsebene für je ein Objekt
Lieferung	Eine archivische Akzession aus 1-n IE, die durch eine Transaktions-ID (i.d.R. eine GUID) identifizierbar ist

Alle im Gesamtsystem verwendeten Objekte sind durch geeignete Identifier eindeutig gekennzeichnet. Die Identifier sind Teil der Metadaten und der Datenbankinhalte von DiPS.kommunal.



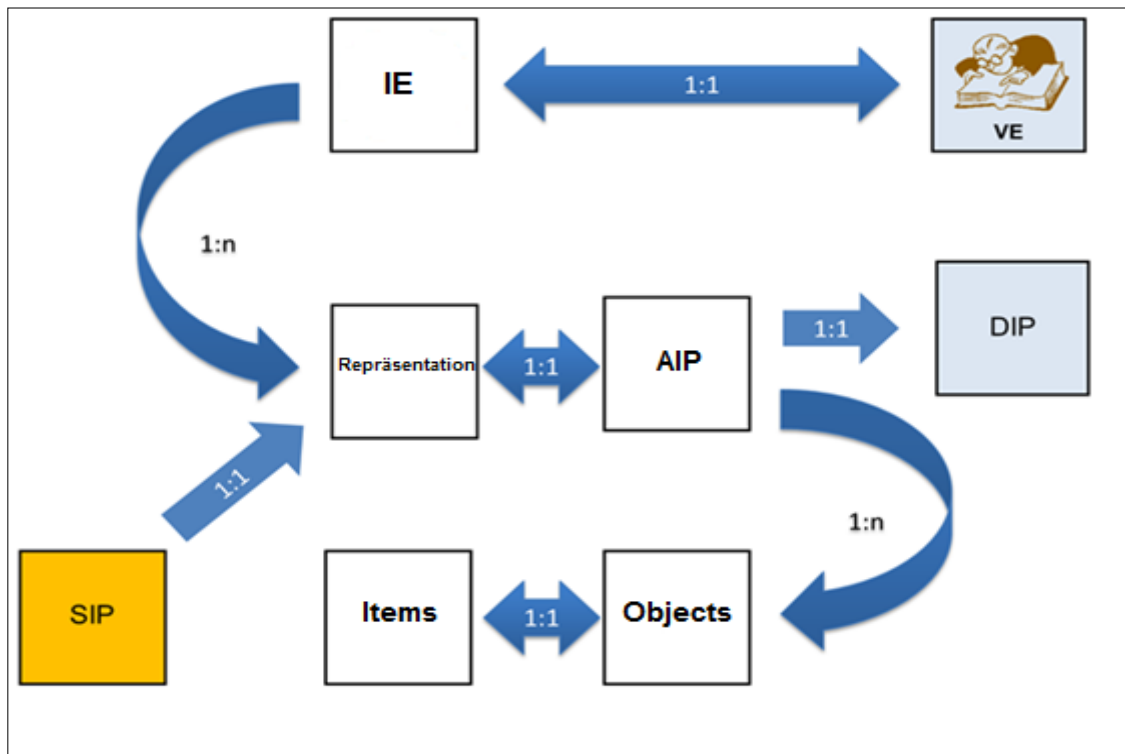


Abbildung 4: Relationen in DiPS.kommunal

**Daraus ergeben sich folgende logische und numerische Verhältnisse der verwendeten Objekte:**

### **Inhaltliche Einheit (IE)**

1 IE = 1 Verzeichnungseinheit in der Erschließungssoftware  
1 IE = 1 bis n Repräsentationen

### **Repräsentation**

1 Repräsentation = 1 AIP

### **AIP**

1 SIP wird zu 1 AIP  
1 AIP = 2 bis n Objects  
1 AIP wird zu 1 DIP

### **Objects**

1 Object = 1 Item

DiPS.kommunal ermöglicht die Langzeitarchivierung bereits inhaltlich strukturierter Daten (z.B. elektronische Akten), wie sie aus Dokumentenmanagement-Systemen oder vergleichbaren Fachverfahren der Verwaltung entstehen. Es werden dafür die Nachrichtentypen 503 und 506 des bundesweit verbreiteten XDomea 2.2.0 Standards unterstützt. Eine elektronische Akte (eAkte) entspricht hierbei in der Regel einer IE; 1-n eAkten werden zu einer Lieferung (Akzession) formiert, die mit den entsprechenden Metadaten aus dem Quellsystem angereichert und dann in den Ingest eingeschleust wird.

### 2.1.2 Repräsentationen

Jedes von DiPS.kommunal erzeugte AIP enthält genau eine Repräsentation einer IE. Diese Repräsentation beinhaltet alle Primär- und Metadaten, die für den Erhalt und das langfristige Verständnis dieser IE benötigt werden.

Jede Veränderung an Primär- oder Metadaten führt zur Erstellung einer neuen Repräsentation und somit auch zur Bildung eines neuen physischen Archivpakets. Dies führt dazu, dass in DiPS.kommunal mehrere AIPs zur gleichen inhaltlichen (semantischen) Einheit vorliegen. Die Differenzierung (= Repräsentationenverwaltung) erfolgt über die eindeutigen IDs der Inhaltlichen Einheit (IEID) sowie der Archivpakete (AIPID) und innerhalb der AIP-XML der Nachfolge-Repräsentationen.

Zusätzlich besteht bereits im Pre-Ingest die Möglichkeit ergänzende Nutzungsrepräsentationen in einem getrennten Speicherbereich abzulegen. Zum Beispiel zu hochauflösenden Video- oder Audio-Rohdaten können Repräsentationen in einem komprimierten Format (z.B. mp4/mp3) abgelegt werden. Diese Repräsentationen werden Nutzungsrepräsentationen genannt und werden ausschließlich für die Benutzung dem Anwender zur Verfügung gestellt.

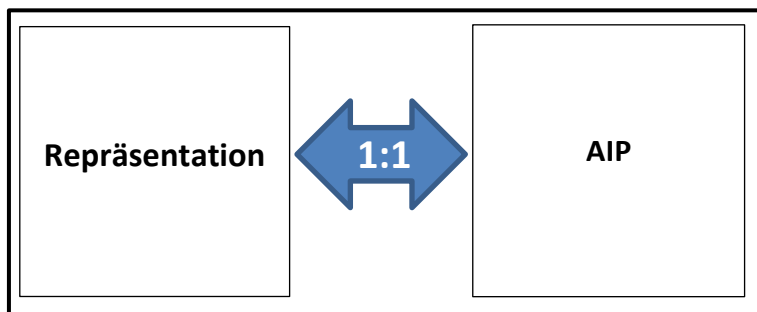


Abbildung 5: Für jede Repräsentation wird ein eigenes AIP angelegt

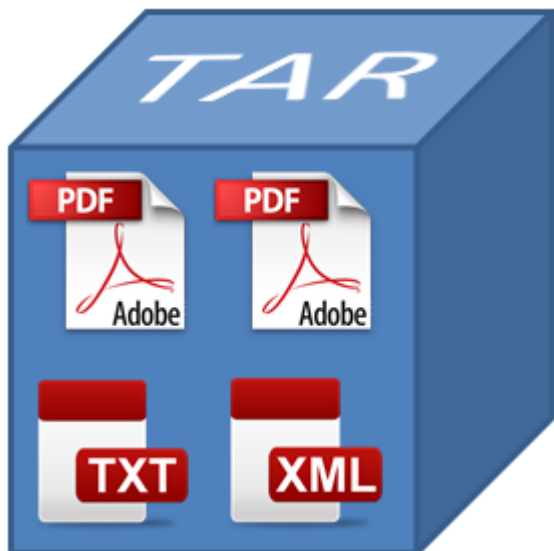
### 2.1.3 Struktureller Aufbau des AIP

Jedes von DiPS.kommunal erzeugte AIP besteht aus einem unkomprimierten TAR-Container als Hülle und 1-n Items sowie einer XML-Datei, die einem offenliegenden, archivübergreifend genutzten XML-Schema folgt.

Die Metadatendatei enthält einen kompletten PREMIS-Datensatz<sup>8</sup>, administrative und inhaltlich beschreibende Metadaten inkl. objektspezifischer Strukturmetadaten.

---

<sup>8</sup> PREMIS steht für „PREservation Metadata: Implementation Strategies“ vgl. <http://www.loc.gov/standards/premis>; der sog. „Rights-Bereich“ wird nicht befüllt, weil sich diese Informationen in der Erschließungssoftware befinden und sich z.B. durch Gesetzesänderungen verändern können. Diese Trennung sorgt für eine höhere Stabilität und geringeren Pflegeaufwand im elektronischen Langzeitarchiv.



Die im AIP enthaltenen Primärdateien werden im Ingest-Prozess einer Formaterkennung und einer Formatvalidierung unterzogen und für eine Archivierung zugelassen.

Die Datenintegrität wird systemseitig für den gesamten Pre-Ingest-Bereich geprüft und nachgewiesen. Die XML-Datei wird im Zuge des Ingests inhaltlich überprüft und technisch validiert.

Abbildung 6: Schematische Darstellung eines AIPs in DiPS.kommunal innerhalb eines TAR-Containers

## 2.1.4 Lieferung

Die Archivierung von Daten erfolgt in DiPS.kommunal lieferungs- bzw. akzessionsweise. Eine Lieferung besteht aus 1-n Primärobjekten im Ingest und 2-n Archivpaketen bei Abschluss der Eingangsbearbeitung. Hierbei wird ein Archivpaket mit den Prüfsummen der Gesamtlieferung innerhalb des Ingests angelegt. Der Umfang der Lieferung ist vom Archivar bei Anlage des Übernahmeprojekts zu bestimmen. Während der Übernahme kann der Umfang der Lieferung nur unter genau definierten Bedingungen verändert werden. Alle Prozessschritte in DiPS.kommunal werden lieferungsweise durchlaufen, d.h., ein neuer Prozessschritt beginnt erst dann, wenn der vorherige Schritt für die gesamte Lieferung erfolgreich abgeschlossen werden konnte.

## 2.1.5 Übergabe der Lieferung

Nachdem im Pre-Ingest die Lieferung erstellt wurde, muss die durch den PIT.plus erzeugte ZIP-Datei vom lokalen Arbeitsplatzrechner zum serverseitig ausgeführten Ingest-Prozess übertragen werden. Jeder DiPS.kommunal Anwender erhält einen Online-Zugang zu dem System. Der Zugang erfolgt über eine sichere Zwei-Faktor-Authentifizierung (Token und Passwort).

Nach dem Login hat der Nutzer die Möglichkeit die zuvor erstellte ZIP-Datei in den Ingest-Prozess hochzuladen. Die weitere Verarbeitung erfolgt völlig automatisiert.

## 2.2 Ingest

Die Übernahme der Übergabepakete und der dazu gehörigen Metadaten erfolgt im Ingest-Prozess automatisiert.

Der Ingest besteht aus folgenden Schritten:

- Integritätsprüfung
- Formaterkennung
- Formatvalidierung
- Ggf. Konvertierung
- Ggf. Formatvalidierung
- Metadatenergänzung
- Export der AIPs
- Ggf. Rückmeldung an das Produktivsystem
- Export Metadaten für Erschließungssoftware

## 2.2.1 Integritätsprüfung

Die übernommenen SIPs und die dazugehörigen Metadaten werden auf Vollständigkeit und auf Unversehrtheit überprüft. Wird ein Fehler festgestellt, so wird eine entsprechende Meldung erzeugt und die Lieferung abgelehnt. Darüber hinaus erfolgt eine weitere Überprüfung, ob das SIP nicht bereits in der Vergangenheit abgeliefert worden ist. Dadurch wird verhindert, dass ein SIP doppelt verarbeitet und in das elektronische Langzeitarchiv abgelegt wird.

## 2.2.2 Formaterkennung und -validierung

Die Formaterkennung und Formatvalidierung werden eigenständig unter Einbindung der freien Software DROID (Pronom-Datenbank) und JHOVE vollzogen. Die Ergebnisse aus den Anwendungen liefern anschließend die notwendigen Informationen, um im Ingest-Prozess die Daten entsprechend weiterzuverarbeiten.

DiPS.kommunal unterscheidet im Ingest-Prozess zwischen Archiv- und Konversionsformaten.

Archivformate sind Formate, die in das elektronische Langzeitarchiv übernommen werden. Die Definition dieser Formate erfolgt durch das Preservation Management. Die Definitionen werden stetig vom Systemanbieter geprüft und den DiPS.kommunal Benutzern dynamisch zur Verfügung gestellt.

Konversionsformate sind Formate, die in ein Archivformat konvertiert werden können. Dies geschieht bei DiPS.kommunal völlig automatisiert, aber dennoch hat der Archivar im Ingest-Prozess die Option die Datenverarbeitung manuell zu steuern. Das ursprüngliche Format wird zum Erhalt der Daten als Repräsentation 0 ebenfalls in das elektronische Langzeitarchiv übernommen.

Wird ein Format identifiziert, welches weder ein Archiv- noch ein Konversionsformat ist, so wird in dem Verarbeitungsprozess eine Meldung ausgegeben und keine Übernahme durchgeführt. Dadurch wird das elektronische Langzeitarchiv vor Formaten geschützt, die nicht im Vorfeld spezifiziert worden sind. Soll das Format dennoch in das Langzeitarchiv übernommen werden, besteht die Möglichkeit, die Konfiguration entsprechend anzupassen. Aufgrund der heutigen Datenvielfalt kann es vorkommen, dass Dateien nicht einwandfrei

validiert werden konnten. In diesem Fall besteht in DiPS.kommunal die Möglichkeit, die Informationen in dem vorliegenden Format durch eine Kontrollinstanz zu Sichten und ggf. anschließend manuell zur weiteren Verarbeitung und Übernahme in das Langzeitarchiv freizugeben.

Das hat den Vorteil, dass relevante Informationen aufgenommen werden können, wenn es erforderlich ist (d.h. die Informationen existieren nur in diesem Format und dem aktuellen Zustand). Alle Schritte werden entsprechend protokolliert und zur zukünftigen Nachvollziehbarkeit in den Metadaten des betreffenden AIP abgespeichert.

### **2.2.3 Metadatenergänzung**

Die Metadaten können im Ingest-Prozess durch den Bearbeiter ergänzt oder ggf. angepasst werden. Es besteht die Möglichkeit über eine AIP-Vorlage die Informationen an alle AIPs einer Lieferung zu vererben oder die benötigten Metadaten für jedes AIP einzeln zu vergeben. Dabei sind relevante Metadatenfelder als Pflichtfelder definiert worden.

### **2.2.4 Bildung der AIPs und Export der Metadaten für Erschließungssoftware**

Nach der Freigabe der Lieferung durch den Bearbeiter werden die AIPs entsprechend der in Abbildung 6 dargestellten Struktur gebildet. Es wird eine XML-Datei mit allen erschließungsrelevanten Metadaten und den technischen Referenzierungen generiert, welche anschließend in die entsprechend im Einsatz befindliche Erschließungssoftware zur Erstellung eines Findhilfsmittels importiert werden kann. Das Archiv kann in gewohnter Weise eine archivische Erschließung für das elektronische Archivgut anschließen sowie übergreifend über analoge und elektronische Bestände recherchieren.

## **2.3 Archivspeicher**

Der Archivspeicher dient der Langzeitarchivierung der durch DiPS.kommunal generierten AIPs. Der Archivspeicher wird redundant in zwei hoch modernen Rechenzentren betrieben. Es erfolgt eine eigenständige Replikation der Daten, so dass auf beiden geographisch voneinander getrennten Speichersystemen zu jedem Zeitpunkt ein einheitlicher Zustand der abgelegten AIPs besteht.

## **2.4 Preservation Management**

In den Metadaten, die zu jedem AIP erhoben und gespeichert werden, steht für die technischen Metadaten der PREMIS-Datensatz zur Verfügung. Darüber hinaus können innerhalb der Metadaten Angaben zu der jeweiligen Erhaltungsgruppe, zu der die Primärdaten des AIP gehören, gemacht werden. Diese Informationen sind innerhalb des Preservation Management die Grundlage der digitalen Bestandserhaltung.

## **2.5 Benutzung**

Jede Software soll dazu genutzt werden, wozu sie erstellt worden ist. DiPS.kommunal trennt daher bewusst die Aufgabenbereiche des elektronischen Langzeitarchivs und der

Erschließungssoftware. Das bedeutet, dass Erschließungsinformationen nicht mit in die Metadaten des AIPs geschrieben werden. Der große Vorteil bei diesem Verfahren ist, dass bei einer Änderung oder Ergänzung der Erschließungsinformationen das AIP unverändert bestehen bleiben kann. Quellenkritisch bietet die Lösung dem Benutzer eine eindeutige Trennung von Registraturbildnerinformation (im AIP) und archivischen Informationen (in der Erschließungssoftware). Die inhaltliche Suche nach Verzeichnungseinheiten geschieht in der Erschließungssoftware. Aus dieser Archivsoftware wird die entsprechende Inhaltliche Einheit aus dem Langzeitarchiv angefordert. Dieses generiert „Benutzungskopien“ sog. DIPs, während die eigentlichen AIPs unangetastet bleiben. Die Anzeige des Archivguts aus dem elektronischen Langzeitarchiv (dem digitalen Magazin) erfolgt browsergestützt über den DOXiS4-webCube. Erschließungssoftware und Langzeitarchiv sind über eindeutige Identifikatoren (GUIDs) miteinander verknüpft, sodass der Aufruf der AIPs zu den Verzeichnungseinheiten über den Erschließungsdatensatz auch nach Migrationen oder anderen technischen Erhaltungsmaßnahmen im Langzeitarchiv noch funktioniert.

Im Langzeitarchiv selbst ist nur eine rudimentäre inhaltliche Recherche über die Metadaten des Registraturbildners (Aktentitel, Aktenzeichen, Akzessionsnummer, Transaktions-ID etc.) möglich.

## 3 Software

Die nachfolgenden Unterpunkte beschreiben die Software, die im Rahmen von DiPS.kommunal eingesetzt wird. Dazu gehören die Betriebssysteme und die notwendigen Laufzeitumgebungen sowie weitere Module, die im Ingest-Prozess genutzt werden.

Die Nutzung der einzelnen Komponenten ermöglicht die Aufnahme und Verarbeitung von SIPs, welche nach erfolgreicher Bearbeitung als AIPs in das elektronische Langzeitarchiv gespeichert werden.

### 3.1 Betriebssysteme

Der Einsatz von DiPS.kommunal ist aufgrund der Modularität und der daraus resultierenden Herstellerunabhängigkeit bewusst nicht von einem bestimmten Betriebssystem abhängig. Diese Flexibilität wurde bei DiPS.kommunal strikt auf Basis des OAIS-Modells eingehalten.

Derzeitig sind die Betriebssysteme Windows Server 2012 R2 und Linux in aktuellen Versionen und Servicepacks in Betrieb.

### 3.2 Java (JDK)

Java ist eine objektorientierte Programmiersprache und dient als Grundlage für die Ausführung der technischen Module in der elektronische Langzeitarchivkomponente Ingest (Siehe 3.3). Die Installation des Java Development Kits (JDK) ist eine notwendige Systemvoraussetzung für das Ausführen der Komponenten.

Das JDK ist für die gängigsten Betriebssysteme, wie z.B. Mac OS X, Linux, Windows, Solaris, sowohl in der 32 als auch in der 64 Bit Version erhältlich. Dies hat abermals den Vorteil, dass die technischen Komponenten innerhalb DiPS.kommunal Lösung Plattformunabhängig betrieben werden können.

### 3.3 Ingest-Komponente (technisch)

Die technische Ingest-Komponente wird als Dienst auf dem Server ausgeführt und überwacht, protokolliert und koordiniert bestimmte Prozessschritte, die notwendig sind, um gelieferte SIPs in AIPs zu wandeln.

Sie überwacht im Vorfeld festgelegte Ressourcen und initiiert weitere Schritte, sobald z.B. eine Lieferung durch den PIT.plus erstellt und im Eingangsordner abgelegt worden ist. Anschließend werden die in Kapitel 2.2 beschriebenen Schritte automatisiert gesteuert durchlaufen.

Aufgrund des modularen Aufbaus besteht die Möglichkeit, neben den bereits verwendeten Anwendungen JHOVE und DROID, bei Bedarf weitere Module einzubinden.

Innerhalb des Verarbeitungsprozesses, der durch eine Lieferung angestoßen wird, erfolgt eine Kommunikation mit den weiteren DiPS.kommunal-Komponenten „Content Service Bus“ (CSB), dem Workflow-Service sowie dem Rendition-Server. Die Funktionsbeschreibung der genannten Komponenten erfolgt in den nachfolgenden Kapiteln.

Die Verarbeitung wird durch die im Vorfeld fachlich definierten Abläufe (Workflows) gesteuert. Dadurch wird gewährleistet, dass sämtliche Aktionen, entsprechend den festgelegten Vorgaben ausgeführt werden.

### 3.4 DROID

DROID (Digital Record Object IDentification) ist eine Software, die von dem Nationalen Archiv Englands entwickelt worden ist und im Rahmen des technischen Ingest-Prozesses eingesetzt wird. Auf Basis einer Signatur<sup>9</sup>, die regelmäßig durch das Digital Preservation Department des National Archive UKs aktualisiert und in dem PRONOM-Onlinedienst<sup>10</sup> bereitgestellt wird, ist die Software DROID in der Lage, eine Vielzahl von Formaten zu erkennen. Dies garantiert den DiPS.kommunal-Archiven, ohne dass manuell Einstellungen vorgenommen werden müssen, eine ständig aktuelle Formaterkennung.

Dabei ist jedem Dateiformat eine bestimmte PUID zugeordnet, so dass das Format einer zu verarbeitenden Datei eindeutig identifiziert werden kann. Die Ergebnisse, die DROID im Rahmen der Formaterkennung ausgibt, dient der anschließenden Weiterverarbeitung.

---

<sup>9</sup> <http://www.nationalarchives.gov.uk/aboutapps/pronom/release-notes.xml>

<sup>10</sup> <http://www.nationalarchives.gov.uk/help/pronom/faq.htm>

### **3.5 Modul Fileextension**

Damit die Genauigkeit bei der Formaterkennung in DiPS.kommunal noch weiter erhöht wird, kommt das Fileextension-Modul im Bereich der Formaterkennung zum Einsatz. Fileextension nutzt als Information die Endung der jeweiligen zu verarbeitenden Datei. Dadurch kann das bereits im Vorfeld ermittelte Ergebnis durch DROID zum einen bekräftigt und zum anderen evtl. korrigiert werden, falls DROID verschiedene Formate in einer Datei erkennen sollte. Dies kann bspw. eintreten, falls verschiedene Dateien ineinander verschachtelt sein sollten.

### **3.6 JHOVE**

JHOVE (Java Harvard Object Validation Environment) ist ein Modul, welches zur Formatvalidierung eingesetzt wird. Es wurde entwickelt um Dateiinhalte anhand des vorab erkannten Dateiformates auf ihre Validität hin zu überprüfen.

Das System liefert den Benutzern aussagekräftige Meldungen, wenn bspw. eine Datei korrekt erkannt worden ist, aber ihr inhaltlicher Aufbau nicht dem spezifizierten Standard entspricht. Die betreffende Kontrollinstanz (DiPS.kommunal-Rolle „Ingestmanager“) kann dann entscheiden, ob die betroffene Datei nach manueller Sichtung in der Form in das elektronische Langzeitarchiv aufgenommen oder abgelehnt werden soll.

### **3.7 Pre-Ingest-Toolset (PIT.plus)**

Das Pre-Ingest-Toolset (PIT.plus) ist eine Anwendungssoftware, welche die Übernahme von strukturierten (z.B. elektronische Akten) sowie unstrukturierten Daten (Dateien) in das Langzeitarchiv erlaubt.

Der PIT.plus ist eine Client-Anwendung und dient der Erstellung von Lieferungen in das Langzeitarchiv. Er wird lokal auf dem Arbeitsplatzrechner des Benutzers ausgeführt. Systemvoraussetzung hierfür ist lediglich ein Windows-Betriebssystem inkl. einer .NET Laufzeitumgebung.

Bei der Erstellung der Lieferung hat der Anwender im PIT.plus die Möglichkeit zu der Erhaltungsrepräsentation (ERP) auch die entsprechende Nutzungsrepräsentation (NRP) mit abzuliefern. Durch den PIT.plus wird automatisiert verifiziert, dass die 1:1 Zuordnung zwischen der ERP und der NRP erhalten bleibt.

Damit eine Lieferung abschließend mit dem PIT.plus erzeugt werden kann, müssen fest definierte Metadaten angegeben werden. Zu jeder erzeugten Lieferung wird vom System eine Prüfsumme generiert, die im weiteren Verarbeitungsprozess zur Validierung genutzt wird. Dadurch wird sichergestellt, dass alle Dateien der Lieferung vollständig und unversehrt an den nachgelagerten Ingest übergeben werden.

Das Pre-Ingest-Toolset ist ein Produkt der Firma SER, das im Auftrag der Stadt Köln und des Landschaftsverbandes Westfalen-Lippe durch Hewlett Packard (HP) entwickelt wurde.



## **3.8 DOXIS 4**

DOXIS 4 ist ein Enterprise-Content-Managementsystem der Firma SER und beinhaltet folgende Komponenten.

### **3.8.1 Webcube (Ingest)**

Der Webcube ist eine browserbasierte und -unabhängige Webanwendung und dient der finalen Bearbeitung der Lieferung im Ingest-Prozess vor der eigentlichen Langzeitarchivierung. Der Webcube wurde um den Funktionsumfang des DiPS.kommunal Ingest-Prozesses erweitert. Der Aufruf erfolgt zur Wahrung des Datenschutzes über eine verschlüsselte HTTPS-Verbindung. Im Webcube besteht die Möglichkeit weitere AIP-Metadaten einzugeben oder bereits vorhandene zu bearbeiten.

Über die Webcube-Suchfunktion können die Anwender innerhalb der bereits archivierten Daten recherchieren. Der eigentliche Zugriff auf die AIPs im Langzeitarchiv erfolgt über eine Erschließungs- bzw. Archivverwaltungssoftware.

Die Webcube-Rolle des Ingestmanagers verfügt neben den Rechten und Möglichkeiten des Archivars noch über weitere Funktionalitäten. Er ist in der Lage auf Systemmeldungen, die im Rahmen der Überprüfung der Lieferung auf ihre Vollständigkeit, Unversehrtheit und Validität durchgeführt wurden, zu reagieren. Um die Flexibilität zu erhalten, besteht in DiPS die Möglichkeit, Daten zu archivieren, obwohl diese durch JHOVE als nicht valide klassifiziert worden sind.

Dateien, die keinem definierten Archivformat entsprechen und ein Konversionsformat darstellen, werden durch den in DiPS.kommunal eingebundenen Rendition-Server<sup>11</sup> selbstständig in ein spezifiziertes archivfähiges Zielformat gewandelt. Bevor die AIPs in das elektronische Langzeitarchiv gespeichert werden, kann der Benutzer die gewandelten Dateien sich über den AIP-Viewer anzeigen lassen. Der AIP-Viewer ist Bestandteil des Webcubes und erlaubt die komfortable Anzeige der erzeugten AIPs.

Darüber hinaus werden aus dem Webcube heraus die Metadaten der bereits archivierten Daten für die jeweilig verwendete Erschließungssoftware bzw. Archivverwaltungssoftware aufbereitet exportiert.

### **3.8.2 Content Service Bus (CSB)**

Der CSB stellt eine zentrale Komponente der Infrastruktur des elektronischen Langzeitarchivs DiPS.kommunal dar. Der CSB wird als Dienst im Hintergrund betrieben.

Der Dienst bindet alle weiteren Module an, die im Rahmen des Ingests benötigt werden. Dazu gehören die Anbindung der Datenbank (Oracle), des Webservers (Apache Tomcat), des Workflow-Services sowie des Archivspeichers.

---

<sup>11</sup> Siehe Kapitel 3.9

### **3.8.3 Workflow-Service**

Der Workflow-Service wird eingesetzt, um die im Vorfeld definierten Abläufe automatisiert zu verarbeiten. Dies umfasst z.B. die Bearbeitungsschritte Formaterkennung und Formatvalidierung.

Durch den Einsatz des Workflow-Services sind die Abläufe der Prozesse im Ingest fest definiert und werden protokolliert. Dadurch wird sichergestellt, dass immer eine definierte Reihenfolge von Prozessen durchgeführt wird. Dies garantiert eine dauerhafte und gleichbleibende Qualität der Übernahme von Informationen in das elektronische Langzeitarchiv.

### **3.9 Rendition-Server**

Der Rendition-Server stammt von der Firma SER und ist die Wandlungseinheit für festgelegte Formate und ist als separates Modul innerhalb der DiPS.kommunal-Infrastruktur eingebunden. Es werden im Ingest-Prozess proprietäre Dateiformate, wie z.B. .doc, .xls, .ppt u.v.m, in langzeittaugliche Formate umgewandelt. Bei der Wandlung werden neue Repräsentationen erzeugt und in den PREMIS-Daten nachgehalten, wobei die Ursprungsrepräsentation erhalten bleibt.

Durch Konfigurationen besteht in DiPS.kommunal die Möglichkeit zu definieren, welche Formate langzeitarchivfähig sind und welche Dateiformate in ein Archivformat, z.B. PDF/A-1b, konvertiert werden sollen.

### **3.10 Erschließungssoftware / Archivverwaltungssoftware (Optional ACTApro)**

DiPS.kommunal ist ein offenes System und somit unabhängig von der zu verwendenden Erschließungssoftware. Einzige Voraussetzung ist, dass die anzubindende Software XML-Exporte verarbeiten und über deren Benutzeroberfläche die IEs im Langzeitarchiv wieder aufrufen kann. Über die Erschließungssoftware werden i.d.R. die fachlichen Recherchen ausgeführt und die Zugangsrechte (Schutz- und Sperrfristen) verwaltet.

Projektbezogen kann mit DiPS.kommunal auch der Betrieb einer Erschließungssoftware angeboten werden. Zurzeit ist die Anbindung der Erschließungssoftware ACTApro realisiert. Eine Anbindung weiterer Erschließungssoftwareprodukte ist in Planung.

### **3.11 Überwachung der eingesetzten Komponenten**

In DiPS.kommunal werden alle eingesetzten Komponenten laufend überwacht. Im Falle einer Störung werden die zuständigen Mitarbeiter automatisiert benachrichtigt und können auf das eingetretene Ereignis schnell reagieren.

## **4 Qualitätssicherung**

Die DiPS.kommunal-Qualitätssicherung sieht eine dreistufige Infrastruktur vor. Es wird eine Entwicklungs-, Test- und Produktivumgebung eingesetzt.

Als erstes werden Neuentwicklungen und Erweiterungen innerhalb der Entwicklungsumgebung implementiert oder bestehende Module erweitert.

Im nächsten Schritt werden die Neuerungen innerhalb der Testumgebung spezifischen Fachtests unterzogen.

Sind die Testergebnisse erfolgreich verlaufen und die festgelegten Kriterien erfüllt worden, werden die Neuerungen, für alle DiPS.kommunal Benutzer verfügbar, in die Produktivumgebung übertragen.

Dieser Ablauf gewährleistet eine dauerhafte Qualitätssicherung und eine sichere und stabile Nutzung von DiPS.kommunal.

### **4.1 Archivspeicher**

Als Archivspeicher können verschiedene Systeme angebunden werden. Es wird ausschließlich revisionssicherer Speicher eingesetzt, sodass eine langfristige Aufbewahrung und eine hohe Verfügbarkeit gewährleistet sind. Aus Sicherheitsgründen werden die Archivspeicher jeweils an geografisch getrennten Rechenzentren betrieben. Die Replikation zwischen den Rechenzentren erfolgt völlig automatisiert.

### **4.2 Infrastruktur**

Der Aufbau der Infrastruktur gestaltet sich nach der folgenden Abbildung. Dort sind alle elementaren Bestandteile ausgeführt, die das elektronische Langzeitarchiv DiPS.kommunal zur Verfügung stellt.

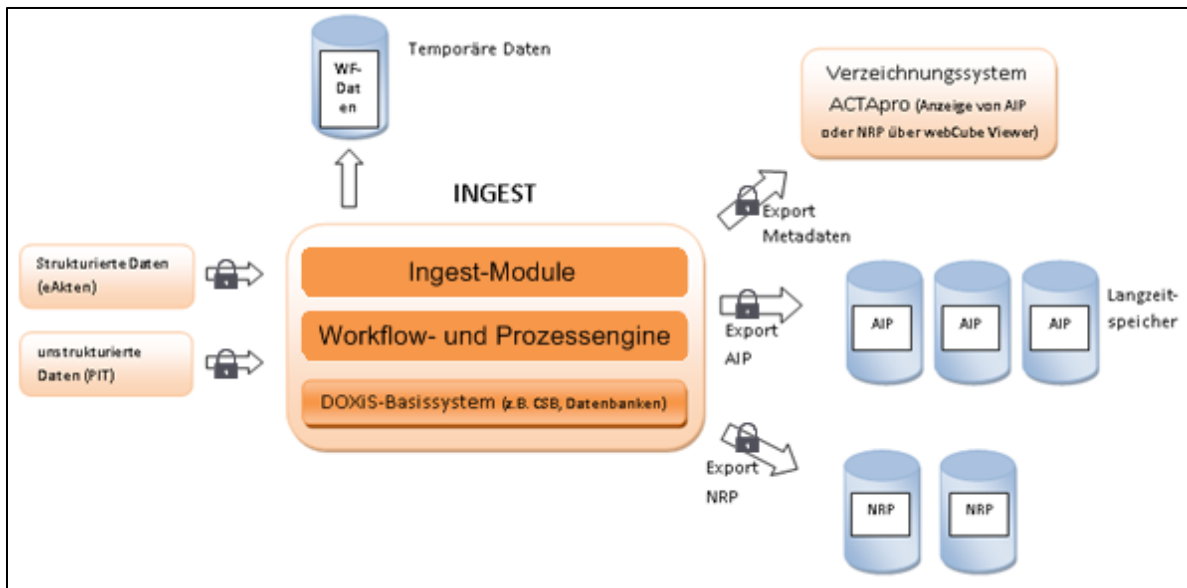


Abbildung 7 - Infrastruktur DiPS.kommunal

#### 4.2.1 Kommunikation innerhalb des Services

Aufgrund des hohen Schutzbedarfs der Archivalien findet zur Abwehr von Datendiebstahl und Manipulation die Kommunikation in DiPS.kommunal zwischen den sicherheitsrelevanten Kanälen ausschließlich auf einer 2048 Bit zertifikatsbasierten SSL Verschlüsselung statt.

#### 4.2.2 Ausfallsicherheit der Infrastruktur

Alle bei DiPS.kommunal eingesetzten Ingest-Komponenten (Webcube, Datenbanken, CSB etc.) werden redundant betrieben. Dies ermöglicht eine hohe Verfügbarkeit (z.B. auch im Falle von Wartungsarbeiten an Hard- oder Software).

#### 4.2.3 Sicherungsstrategie der Server

Es erfolgt eine tägliche Datensicherung aller Server.

### 4.3 Betrieb der technischen Infrastruktur

Die in DiPS.kommunal eingesetzte Hardware wird in zwei geografisch getrennten und redundant betriebenen Rechenzentren betrieben.

#### 4.3.1 Rechenzentren

Die genutzten hochmodernen Rechenzentren befinden sich an geographisch auseinander liegenden Standorten und werden vom Knotenbetreiber betrieben. Alle für den Betrieb der Rechentechnik erforderlichen technischen Anlagen der Gebäudeinfrastruktur werden ebenfalls redundant eingesetzt.

Die Rechenzentren verfügen über sichere Strom-, Notstromversorgungen und Klimatisierungen. Im Rahmen der verwendeten Sicherheitstechnik ist ein Einbruchschutz (Einbruchwiderstandsklasse 4), Brandschutz sowie ein System für die Zutrittskontrolle aktiv.

### **4.3.2 Wartung der Hard- und Software**

Die Wartung der Hard- und Software erfolgt gemäß den vereinbarten Verträgen. Die Wartung beinhaltet den Austausch von defekten Hardwarekomponenten, die Behebung von Fehlern und die Aktualisierung der eingesetzten Software.

### **4.3.3 Überwachung der Hardware**

Um frühestmöglich auf mögliche Hardwarefehler zu reagieren, werden alle eingesetzten Komponenten, gemäß den abgeschlossenen Verträgen, kontinuierlich durch den Betreiber überwacht.

### **4.3.4 DiPS.kommunal Produktiv und Testumgebung**

Nach dem Login werden die Links zu den Anwendungen angezeigt.

Grundsätzlich erhalten die DiPS.kommunal-Archive zwei Bereiche.

#### **1. Produktivumgebung:**

Innerhalb der Produktivumgebung werden die zu archivierenden Daten nach Abschluss der Bearbeitung final auf die speziellen Archivspeicherlangzeitarchiviert.

#### **2. Testumgebung:**

Innerhalb der Testumgebung haben die DiPS.kommunal-Archive die Möglichkeit intern zu schulen, Dateiformate und Dateigrößen etc. zu prüfen. Die AIPs werden nach Abschluss der Bearbeitung auf einem einfachen und nicht für die Langzeitarchivierung geeigneten Speicher geschrieben.

## Anhang 1

### Archivformate:

Name	Archivformat	Bemerkung
Comma-Separated Values – CSV	X	
Extensible Markup Language 1.0 – XML	X	
Waveform Audio File Format	X	
Broadcast WAVE	X	
MP3	X	
JPEG File Interchange Format (JPG/JPEG)	X	
DPX	X	Unkomprimiertes Filmformat (Ausbaustufe)
Portable Document Format (Archival) - PDF/A 1b	X	
SIARD	X	
Plain Text File – TXT	X	
Portable Document Format (Archival) - PDF/A 1.0 (a)	X	Weitere Ausbaustufe
Portable Document Format (Archival) - PDF/A 2.0 (a, b, u)	X	

## Konversionsformate:

Name	Konversionsformat	Bemerkung
Portable Network Graphic 1.0 bis 1.2 – PNG	X	
Portable Document Format 1.0 bis 1.7 – PDF	X	
BMP - Bitmap graphics	X	
GIF - 8-bit color bitmap graphics	X	
PNG - Portable Network Graphics	X	
TIF - Tagged Image File	X	Ausbaustufe: Verlustfreie Wandlung
DOC - Word 97-2003 Dokument	X	
DOCX - Word 2007-Dokument	X	
MSG - Outlook - Message Format	X	
PPS - PowerPoint Presentation file	X	
PPT - PowerPoint 97-2003 Presentation	X	
PPTX - Powerpoint 2007 Presentation	X	
RTF - Rich Text Format	X	
VSD - Drawing file (Visio) Version 2000/2002/5.0	X	
XLS - Excel 97-2003 Dokument	X	
XLS - Excel 2007 Dokument	X	
HyperText Markup Language document (HTM/HTML)	X	
XSL - Extended Style Sheet Language file	X	Auf Anfrage

OpenOffice Calc	X	Auf Anfrage
OpenOffice Draw	X	Auf Anfrage
OpenOffice Impress	X	Auf Anfrage
OpenOffice Writer	X	Auf Anfrage