

AR 2681

Sven Schlarb

Das PLANETS Testbed: Erforschung und Planung der digitalen Langzeitarchivierung

Zusammenfassung

Das PLANETS Testbed ist eines der zentralen Ergebnisse des PLANETS Projekts (<http://www.planets-project.eu>) und dient der Durchführung und Dokumentation von Experimenten im Bereich der digitalen Langzeitarchivierung. Dieser Artikel stellt das PLANETS Projekt und das PLANETS Testbed als eines der fünf Teilprojekte von PLANETS vor. Daran anschließend beschreibt er ein konkretes Anwendungsszenario im Kontext der digitalen Langzeitarchivierung.

The PLANETS Testbed: Exploration and planning of digital long term preservation

The PLANETS Testbed is one of the major outcomes of the PLANETS Project (<http://www.planets-project.eu>). It serves for execution and documentation of experiments in the field of digital long term preservation. This paper provides an overview on the PLANETS project and the PLANETS testbed as one of the five subprojects of PLANETS followed by a concrete case study in the context of digital long term preservation.

Le PLANETS Testbed: Exploration et implantation de la préservation numérique à long terme

Le PLANETS Testbed est un des résultats les plus importantes du Projet PLANETS. Le Testbed serve pour la réalisation et la documentation des expériences dans le domaine de conservation de données digitales. Cet article donne une idée générale du projet PLANETS et du Testbed en concrète, et explique en détail une expérience typique qu'on pourrait conduire dans une institution de préservation de données digitales.

Schlagworte: PLANETS Projekt, Testbed, Langzeitarchivierung

Keywords: PLANETS Project, Testbed, Longterm Preservation

mots clés: Projet PLANETS, Testbed, archivage numérique

1. Einführung

Die langfristige Archivierung und Bereitstellung von Informationsträgern und Kulturgütern ist traditionell eine Kernaufgabe der Bibliotheken, Archive und Museen. Dieser Aufgabenbereich hat sich in den letzten Jahrzehnten grundlegend verändert, da zu den klassischen Objekten digitale Objekte mit ganz neuen und spezifischen Anforderungen an die langfristige Erhaltung hinzugekommen sind. Die besondere Herausforderung bei der Verwaltung digitaler Objekte liegt darin, dass diese bedingt durch den technologischen Fortschritt in doppelter Hinsicht einem sehr schnellen Wandel unterliegen: Es ändern sich sowohl Medien (CD-ROM, Festplatte, etc.), die zur Speicherung der Daten verwendet werden, als auch Standards (Dateiformate), welche die Interpretierbarkeit der Daten gewährleisten.

Grundlegenden Änderungen der handelsüblichen Speichermedien begegnen wir im Prinzip durch das Kopieren der Daten vom veralteten (z.B. einer CD) auf den neuen Datenträger (z.B. eine Festplatte). Dies hört sich zunächst trivial an, wenn wir jedoch 3000 CDs migrieren und dabei sicherstellen wollen, dass die Daten unversehrt auf dem neuen Datenträger gespeichert werden, müssen zahlreiche Aspekte berücksichtigt werden, wie sie zum Beispiel Forscher der Universität Basel in einer Fallstudie aufgezeigt haben.¹

Änderungen bezüglich der Standards von Dateiformaten wird gegenwärtig mit zwei grundsätzlich verschiedenen Ansätzen begegnet: Einerseits werden Daten, die in einem nicht mehr hinreichend unterstützten Dateiformat gespeichert sind, durch *Migration* in ein aktuelles Format überführt, von dem angenommen wird, dass es auch in Zukunft von Institutionen und Software-Herstellern breite Unterstützung erhält. Bei diesem Ansatz wird das Originalobjekt verändert, weil die Daten im neuen Dateiformat für gewöhnlich in anderer Form oder nach vollkommen anderen Prinzipien organisiert sind. Andererseits wird bei der *Emulation* das digitale Objekt nicht verändert, sondern es wird durch eine Software-gestützte Nachbildung der veralteten technischen Umgebung möglichst originalgetreu mit den Mitteln der neuen Soft- und Hardwareumgebung dargestellt.

Beide zuletzt genannten Ansätze waren Gegenstand des EU-geförderten Forschungsprojekts PLANETS (Preservation and Long term Access through NETworked Services), das wir in diesem Artikel im nächsten Abschnitt vorstellen werden. Im zweiten Abschnitt gehen wir dann näher auf das "Testbed", eines der fünf Teilprojekte von PLANETS, ein. Im dritten Abschnitt beschreiben wir ein Experiment, das mit Hilfe des Testbeds durchgeführt und dokumentiert wurde und bewerten die Ergebnisse in Hinblick auf die zuvor definierten Zielsetzungen.

2. Das PLANETS Projekt

Das PLANETS Projekt war ein Forschungsprojekt, welches von der Europäischen Kommission im Rahmen des Sechsten Forschungsrahmenprogramms kofinanziert wurde. Es begann im Juni 2006 und endete vier Jahre später im Mai 2010. Das PLANETS Konsortium umfasste insgesamt 16 Partnerinstitutionen, darunter Nationalbibliotheken (u.a. die Königliche Bibliothek der Niederlande, die British

Library und die Österreichische Nationalbibliothek) und Nationalarchive (u.a. das Schweizer Bundesarchive) ebenso wie führende Forschungseinrichtungen (University of Glasgow, Technische Universität Wien) und Unternehmen (Microsoft Research, IBM, Tessella). Die Österreichische Nationalbibliothek leitete das Teilprojekt "Testbed" und koordinierte die Entwicklung und Einrichtung der Soft- und Hardwareumgebung zur Evaluierung von Werkzeugen der digitalen Langzeitarchivierung.

Um digitale Daten oder Objekte erfolgreich auf lange Zeit zu erhalten, benötigen wir einerseits die Mittel zur Dokumentation der Strategie, mit deren Hilfe Ziele, Grundsätze, institutionelle Vorgaben, und Methodik festgehalten werden. Andererseits braucht es zuverlässige Werkzeuge für die digitale Langzeitarchivierung, das heißt ausgereifte Software zur Datenanalyse und -manipulation. Für diese Bereiche hat das PLANETS Projekt mit der Entwicklung der "PLANETS Software Suite" Werkzeuge und Hilfsmittel zusammengestellt und geschaffen, die von der Planung bis zur Umsetzung alle Maßnahmen der digitalen Langzeitarchivierung unterstützen.

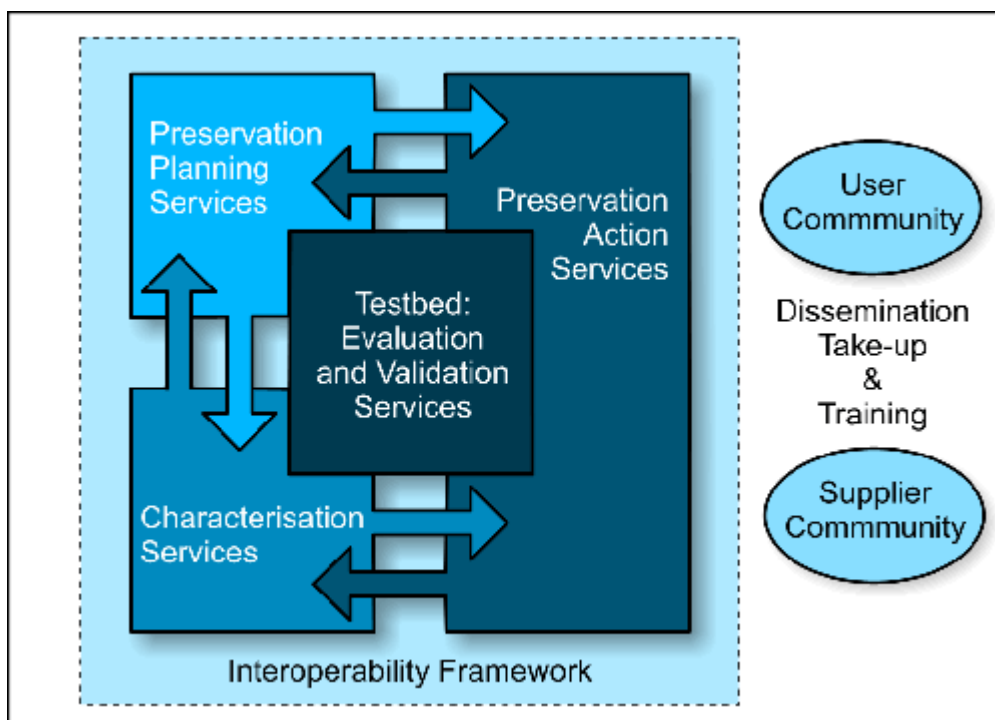


Abbildung 1: Grafische Darstellung der verschiedenen Teilprojekte von PLANETS

Grundlage der technischen Komponenten von PLANETS ist das *Interoperabilitätssystem (in PLANETS: "Interoperability Framework")*, welches eine einheitliche Software-Plattform bereitstellt und Standards für die Kommunikation zwischen den verschiedenen Komponenten definiert. Außerdem definiert das Interoperabilitätssystem die allgemeine Schnittstelle für Anwendungsprogramme der einzelnen Komponenten, das heißt, es werden beispielsweise die Grundfunktionen eines Werkzeugs der Datenmigration dergestalt festgelegt, dass dieser in einheitlicher Form als Dienst² von allen Komponenten genutzt werden kann.

Abbildung 2 zeigt als Beispiel für einen sichtbaren Teil des Interoperabilitätssystems einen Ausschnitt aus der Liste der verfügbaren Datenmigrationsdienste in der Dienst-Datenbank des Interoperabilitätssystems mit der Detailansicht des KakadCompressMigration-Dienstes.

Planets IF Service Registry

Home Registered Services Browse Local Endpoints Add External Endpoints

Migrate Refresh List Search

Name ↕	Registered ▲	Action	Service Details
Ghostscript Migrate	!		<p>KakaduCompressMigration Service Endpoint Information</p> <p>WSDL: http://testbed.planets-project.eu:80/pserv-pa-kakadu/KakaduCompressMigration?wsdl Status: OK</p> <p>KakaduCompressMigration 0.1</p> <p>Type eu.planets_project.services.migrate.Migrate</p> <p>Author Sven Schlarb <shsschlarb-planets@yahoo.de></p> <p>Provider</p> <p>Description Kakadu Version 6.2.1 JPEG2000 sample command lineapplication 'kdu_compress' which shows the potential of the JPEG 2000 Developers' Toolkit. The service uses the 'kdu_compress' command line tool. This command line application uses a simple TIFF file reader which can only read uncompressed TIFF files. This has nothing to do with Kakadu itself. In order to be able to read compressed TIFF files, it should be sufficient to re-compile the command line application with the symbol KDU_INCLUDE_TIFF defined, and link it against the public-domain LIBTIFF library. Furthermore, it is possible to compress variousinputfiles into one jpeg2000 file. But this service only supportsthe one tif input file to one jp2 output file migration. IMPORTANT: The license conditions for this web service apply as if youwere downloading and using the kakadu executables locally. KAKADU LICENSE NOTE: Copyright is owned by NewSouth Innovations Proprietary Ltd, commercial arm of the University of New South Wales, Sydney, Australia. You are free to play around with these executables and even to re-distribute them, so long as such use or re-distribution is accompanied this copyright notice and is not for commercial gain. Note: Binaries can only be used for non-commercial purposes. If in doubt please contact Dr. Taubman.Notice: Using this web service, you are acknowledging and accepting these terms and conditions.</p> <p>KakaduCompressMigration Migrate Service Already Registered</p> <p>There is already an entry for the KakaduCompressMigration Migrate service in the Service Registry. It is not advisable to add multiple registrations for the same service.</p>
Im4JavImageMagickMigrate Migrate	!		
Pdf2PsMigration Migrate	!		
PS2PDF Migrate	!		
KakaduDecodeMigration Migrate	✓		
SanselanMigrate Migrate	✓		
KakaduCompressMigration Migrate	✓		
SoX Migrate	✓		
JJ2000MigrateService Migrate	✓		
NetPbmMigration Migrate	✓		
XenaOOMigration Migrate	✓		
ImageMagickMigrate Migrate	✓		
Avidemux Migrate	✓		
FloppyImageHelperService Migrate	✓		
DiaMigrationService Migrate	✓		
Inkscape Migrate	✓		

Abbildung 2: Ein Teil der Liste der verfügbaren Datenmigrationsdienste in der Dienst-Datenbank des Interoperabilitätssystems mit der Detailansicht des KakadCompressMigration-Dienstes (PLANETS Interoperability Framework - Service Registry)

Die *Werkzeuge zur Planung der Langzeitarchivierungsmaßnahmen (in PLANETS: "Preservation Planning Services)* dienen der Entwicklung einer Strategie zur Planung der Langzeitarchivierungsmaßnahmen unter Berücksichtigung verschiedenster Aspekte, die bei der Langzeitarchivierung eine Rolle spielen können. Sämtliche Schritte des Planungsprozesses werden mit Hilfe des online-basierten Werkzeugs PLATO dokumentiert, das im Rahmen des PLANETS Projekts von der Technischen Universität Wien entwickelt wurde.³

Die *Werkzeuge zur Datenanalyse (in PLANETS: "Characterisation Services")* sind Software-Komponenten, die zur Datenanalyse und zur Bestimmung der

Eigenschaften von Dateien unterschiedlichster Dateiformate verwendet werden können. Zum einen Teil wurden hier an der Universität Köln die XML-basierten Sprachen XC*L (XCEL und XCDL)⁴ und dazu passende Software entwickelt, welche die Dateiformat-übergreifende Eigenschaftsanalyse anhand von Dateien unterschiedlichster Formate sowie den gezielten Vergleich dieser Dateien ermöglichen. Zum anderen Teil wurden bestehende Software-Lösungen, wie zum Beispiel JHove⁵ und Droid⁶ in das PLANETS System eingebunden, um mehrere Möglichkeiten der Datenanalyse unter einem Dach mit einheitlichen Schnittstellen anzubieten.

Die *Werkzeuge zur Datenmigration (in PLANETS: "Preservation Action Services)* sind Werkzeuge, die eine Datentransformation, zum Beispiel die Konvertierung einer Datei in ein anderes Dateiformat durchführen. Dienste der Datenmigration sind in den verschiedenen PLANETS-Komponenten meist eng mit den Werkzeugen der Datenanalyse verknüpft, mit deren Hilfe überprüft werden kann, ob eine Datenmigration definierte Qualitätskriterien erfüllt.

Das *Testbed*, eine web-basierte Applikation, erlaubt es registrierten NutzerInnen, Experimente innerhalb einer kontrollierten Testumgebung auszuführen, zu dokumentieren, zu evaluieren und für andere zu publizieren. Letzteres ist auch dazu gedacht, die Nachnutzung des Experiments für andere zu erleichtern und somit die organisationsübergreifende Zusammenarbeit zu fördern. Zentrale Aufgabe des Testbeds ist es, Leistungsmerkmale von Anwendungen und Werkzeugen der digitalen Langzeitarchivierung zu bestimmen, und die Eignung und Zuverlässigkeit für die Aufgabenstellungen der Langzeitarchivierung zu überprüfen.

Schließlich wurden im Teilprojekt *Öffentlichkeitsarbeit (in PLANETS: "Dissemination and Takeup")* die nachhaltige Weiterverwendung durch andere Institutionen und Firmen gefördert, sowie Schulungen und Workshops organisiert.

Die Ergebnisse und Entwicklungen, die im Rahmen des Projekts erarbeitet wurden, werden in die Open Planets Foundation, eine Non-for-Profit Organisation, eingehen und weitergeführt werden.

3. Das PLANETS-Testbed

Wie bereits erwähnt, werden die Software-Komponenten zur Datenmigration, -manipulation und -analyse durch das Interoperabilitätssystem in einer einheitlichen, PLANETS-weiten Umgebung als Dienste (engl. *web services*) bereit gestellt. Zum Einen können im Testbed die einzelnen Dienste in einem Testbed-Experiment unmittelbar getestet werden. Als Datenmaterial hat das Testbed dafür einen Korpus von über 5000 Dateien unterschiedlichster Dateitypen und -formate bereitgestellt, wobei verschiedene typische Eigenschaften der unterschiedlichen Dateiformate für die jeweiligen Korpuselemente erfasst sind, so dass die Auswirkung der Anwendung eines Dienstes mit Hinblick auf genau diese Merkmale mit bekannten Ausprägungen überprüft werden können. Zum Anderen ermöglicht das Testbed auch die Entwicklung komplexer Abläufe, wie zum Beispiel die Ausführung mehrerer Dienste der Datenmigration hintereinander in Kombination mit Diensten der Datenanalyse, so dass komplexe Szenarien analysiert und bewertet werden können.

Während die PLANETS-Komponente PLATO die Planung einer Strategie der Langzeitarchivierung ermöglicht, kann im Testbed die Skalierbarkeit dieser Strategie am Beispiel großer Datenmengen überprüft werden. Zu diesem Zweck kann der ausführbare PLATO-Langzeitarchivierungsplan im Testbed als Experiment durch Auswahl des entsprechenden Experiment-Typs geladen und anhand der genannten Korpusdaten evaluiert werden.

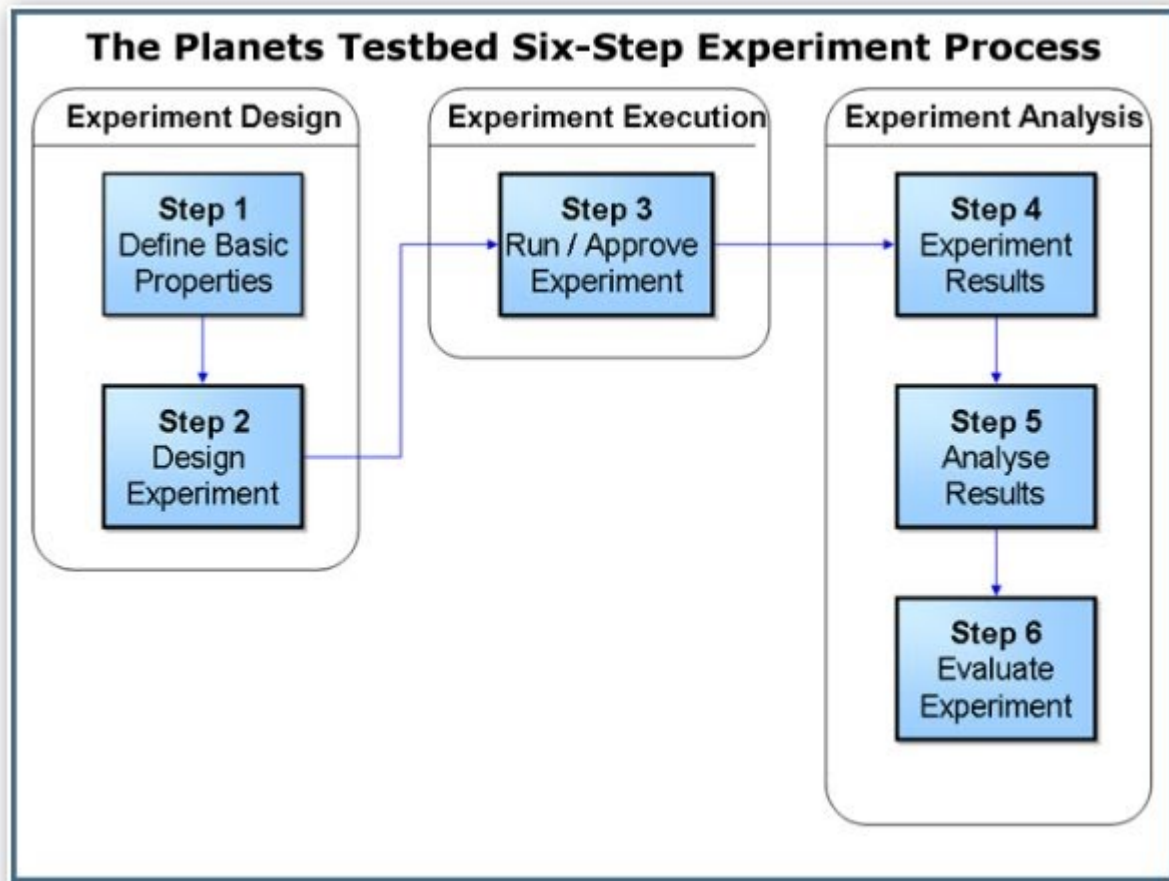


Abbildung 3: Die sechs Planungsschritte eines Testbed-Experiments

Um eine einheitliche Form der verschiedenen Experimente zu gewährleisten, gibt es im Testbed einen festgelegten Ablauf zur Experimentplanung, der in Abbildung 3 anschaulich dargestellt ist. Im ersten Schritt werden grundlegende Informationen zum Experiment, wie zum Beispiel Autor, Beweggründe zur Durchführung und Zielsetzung des Experiments, Verweise auf existierende Experimente, etc. angeführt. Im zweiten Schritt wird der Experiment-Typ (Migration, Skalierbarkeits-Test, etc.), die involvierten Datendienste und die Eingabedaten für das Experiment ausgewählt. Im dritten Schritt wird das Experiment ausgeführt, wobei ein technischer Administrator die Möglichkeit hat, Ressourcen-intensive Experimente auf einen späteren Zeitpunkt zu verschieben. Im vierten Schritt werden die Ergebnisse der Ausführung angezeigt, die anschließend im fünften Schritt mit Hilfe der Datenanalyse-Dienste analysiert werden können. Der letzte Schritt sechs dient zur Erfassung der gesamten Ergebnisse des Experiments.

Das Testbed macht die Evaluation von Software-Werkzeugen zu einem wissenschaftlichen Experiment, da es durch die strukturierte Erfassung der Informationen zu einem Experiment (Grunddaten, verwendete Dienste, Eingabedaten aus dem Korpus) die prinzipielle Reproduzierbarkeit eines Experiments gewährleistet. Auf diese Weise können Schlussfolgerungen, die aufgrund bestimmter Experimentergebnisse gezogen wurden, von anderen registrierten Testbed-BenutzerInnen überprüft und gegebenenfalls verifiziert beziehungsweise falsifiziert werden.

Im Folgenden wird anhand eines Beispielszenarios Planung und Ablauf eines typischen Testbed-Experiments vorgestellt.

4. Das Testbed-Experiment

Als Hintergrund für das Beispielszenario stellen wir uns vor, dass eine Institution die Datenmigration einer ihrer digitalen Sammlungen, deren Digitalisate als unkomprimierte TIFF Bilder vorliegen, von TIFF nach JPEG2000 erwägt, um den erforderlichen Speicherplatzbedarf zu verringern. Wir gehen also davon aus, dass die strategische Entscheidung für eine Migration nach JPEG2000 nach Abwägen der Vor- und Nachteile im Vergleich mit alternativen Migrationsszenarien bereits getroffen wurden. Um das Beispielszenario weiter zu vereinfachen, gehen wir davon aus, dass nur zwei Software-Alternativen, die in der detaillierten Beschreibung des Experiments noch genannt werden, in Frage kommen.

An dieser Stelle muss angemerkt werden, dass dieses Testbed-Experiment dazu dient, ein mögliches Anwendungsszenario vorzustellen, das heißt, die Ergebnisse sind nicht als Grundlage für Schlussfolgerungen bezüglich der Eignung der Dienste für die Langzeitarchivierung vorgesehen, sondern dienen lediglich der Veranschaulichung des Sachverhalts.

Dieses Experiment ist im Zusammenhang mit anderen Studien, die zur JPEG2000 Kodierung unter Verwendung unterschiedlicher Codecs⁷ und der Eignung dieses Bildformats für die Langzeitarchivierung⁸ durchgeführt wurden, als eine konkrete Untersuchung ausgewählter Werkzeuge am Beispiel tatsächlich vorhandener Bibliotheksdaten zu sehen.

4.1. Datenmaterial

Als Datenmaterial wird aus Bildmaterial eingescannter Bücher und Zeitschriften eine Stichprobe bestehend aus 750 Dateien im TIFF-Format, die insgesamt ein Datenvolumen von rund 17,4 Gigabyte ausmachen, zufällig zusammengestellt. Die Daten werden zentral im Testbed verfügbar gemacht und sind auch nach Durchführung des Experiments für andere registrierte BenutzerInnen zugänglich. Auf diese Weise ist die Nachvollziehbarkeit der Ergebnisse des Experiments oder die Anwendung eines alternativen Dienstes unter Verwendung der gleichen Eingabedaten gewährleistet.

4.2. Entwurf des Experiments

Wie bereits erwähnt, werden Werkzeuge und Anwendungen der Langzeitarchivierung in PLANETS als Dienste (engl. *web services*) bereitgestellt. Diese Dienste sind in verschiedene Kategorien gegliedert, je nachdem, ob sie der Format-Migration von einem Ausgangsformat in ein Zielformat (in PLANETS: *Preservation Action Services*), der Merkmals-Extraktion (in PLANETS: *Characterisation Services*) oder der Planung der Langzeitarchivierung (PLANETS - Kategorie: *Preservation Planning*) dienen.

Abbildung 4 zeigt den Ablaufplan des Testbed Experiments, der für das Beispielszenario entwickelt wurde. Ein Kasten mit grauem Hintergrund und abgerundeten Ecken stellt einen Dienst dar. *Gimp*⁹, *Jasper*¹⁰, *KakaduCompress* und *KakaduDecode*¹¹ sind Format-Migrations-Dienste für Bildformate. *XCLCompare*¹² ist ein Dienst, der im Rahmen von PLANETS entwickelt wurde und der selbst wiederum auf einem Dienst der Merkmals-Extraktion basiert. In diesem Experiment wird der Dienst dazu verwendet, spezifische Merkmale digitaler Objekte, die in verschiedenen Formaten vorliegen, zu vergleichen. Die Kästen mit weißem Hintergrund stellen einzelne digitale Objekte eines bestimmten Formats, in diesem Fall der Bildformate TIFF¹³, PNM¹⁴, und JP2¹⁵, dar. Pfeile mit einer Pfeilspitze bedeuten die Migration von einem Quell- in ein Zielformat, und Pfeile mit Doppelspitze bedeuten den Merkmalsvergleich der digitalen Objekte. Die Linien ohne Pfeilspitze deuten darauf hin, dass die entsprechenden Dienste hinsichtlich bestimmter Leistungsmerkmale verglichen werden.

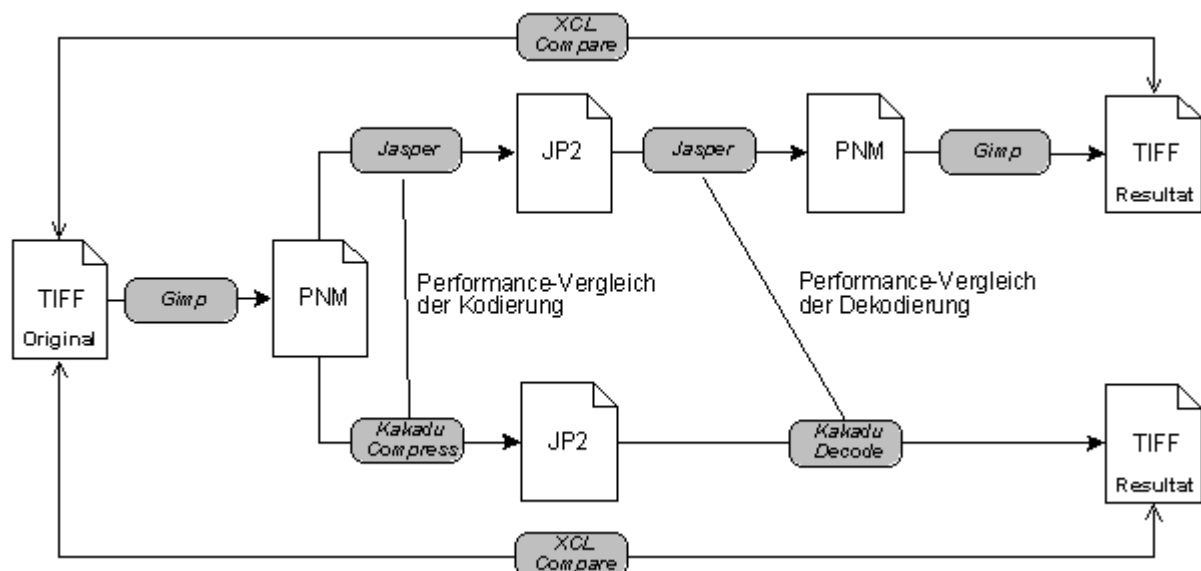


Abbildung 4: Beispielszenario des hier vorgestellten PLANETS Testbed-Experiments

4.3. Zielsetzung des Testbed-Experiments

Der Fokus dieses Experiments liegt auf dem Vergleich verschiedener Werkzeuge zur Kodierung von Bildern im JPEG2000-Format.

Die folgenden Ergebnisse sollen mit Hilfe dieses Experiments ermittelt werden:

- Wie zuverlässig arbeiten die Software-Alternativen (gemessen an der Fehlerquote der Migration)?
- Wie gut ist die Leistung der Software-Alternativen gemessen am Durchsatz (Verarbeitungszeit pro Datenmenge) für JPEG2000-Kodierung und -dekodierung?
- Erreichte Reduktion des Datenvolumens durch die JPEG2000-Kompression.
- Vergleichswerte der Übereinstimmung von Original- und Zielobjekt eines Workflows (jeweils für Kakadu und JasPer) durch Normdatenvergleich.

4.4 Ergebnisse des Testbed-Experiments

In Tabelle 1 sind die Ergebnisse des Experiments zusammengefasst. Der obere Tabellenteil zeigt die Ergebnisse für den Kakadu-Workflow-Pfad und der untere Tabellenteil für den JasPer-Workflow-Pfad aus Abbildung 4. Diese Ergebnisse wurden auf dem zentralen Testbed-Server mit 2.5Ghz Intel Quad Core Xeon E5420, 16GB RAM, 64-bit Ubuntu 8.04 Ubuntu Server Betriebssystem und einer für 8-Gigabyte Arbeitsspeichernutzung konfigurierten Java Virtual Machine ermittelt. Beide Dienste wurden mit den Standard-Service-Parametern aufgerufen.

	<i>tiff</i>	<i>tiff</i> → <i>ppm</i>	<i>ppm</i> → <i>jp2</i>	<i>jp2</i> → <i>ppm</i>	<i>ppm</i> → <i>tiff</i>	<i>tiff</i>	<i>tiff</i> → <i>xcdl</i>	<i>tiff</i> → <i>xcdl</i>
Kakadu-Workflow-Pfad	JHove	Gimp	Kakadu	Kakadu	Sanselan	JHove	XCDL	XCDL
Durchschnittlicher Durchsatz des Dienstes in Millisekunden pro MB	193,24	1051,41	241,17	843,30	537,68	986,75	1665,83	14493,37
Durchschnittlicher Durchsatz der Migrations-Software in Millisekunden pro MB	-	651,13	136,98	413,89	-	-	-	-
Erzeugtes Datenvolumen in GB	-	17,53	5,24	17,39	0,33	-	50,76	0,95
Anzahl der Dateien	-	748	742	745	736	-	732	737
Durchschnittliche Dateigröße in Megabyte	-	24,00	7,24	23,90	0,46	-	71,01	1,32
Kleinste Dateigröße in Megabyte	-	1,18	0,39	1,18	0,02	-	0,00	0,08
Größte Dateigröße in Megabyte	-	153,13	31,59	153,13	1,63	-	147,98	4,82
	<i>tiff</i>	<i>tiff</i> → <i>pnm</i>	<i>pnm</i> → <i>jp2</i>	<i>jp2</i> → <i>pnm</i>	<i>pnm</i> → <i>tiff</i>	<i>tiff</i>	<i>tiff</i> → <i>xcdl</i>	<i>tiff</i> → <i>xcdl</i>
JasPer-Workflow-Pfad	JHove	Gimp	Jasper	JasPer	Gimp	JHove	XCDL	XCDL
Durchschnittlicher Durchsatz des Dienstes in Millisekunden pro Megabyte	197,58	987,08	451,74	903,52	332,55	69,84	1594,55	1844,03
Durchschnittlicher Durchsatz der Migrations-Software in Millisekunden pro Megabyte	-	610,28	346,64	645,40	191,78	-	-	-
Erzeugtes Datenvolumen in Gigabyte	-	17,62	8,06	17,62	17,62	-	52,57	46,72
Anzahl der Dateien	-	750,00	750,00	750,00	750,00	-	750,00	683,00
Durchschnittliche Dateigröße in Megabyte	-	24,06	11,00	24,06	24,06	-	71,78	70,05
Kleinste Dateigröße in Megabyte	-	1,18	0,57	1,18	1,18	-	0,00	3,58
Größte Dateigröße in Megabyte	-	153,13	58,93	153,13	153,13	-	147,98	182,75

Tabelle 1: Überblick über die Leistungseigenschaften der Dienste gemessen am Durchsatz (Verarbeitungszeit pro Dateimenge, hier in Millisekunden pro Megabyte angegeben) für die jeweiligen Formatmigrationschritte von Kakadu (oberer Teil der Tabelle) und JasPer (unterer Teil der Tabelle). Außerdem bietet die Tabelle Einblick in Datenvolumen und Dateigrößen.

Auf dieser Grundlage interpretieren wir einige der Werte aus der Tabelle wie folgt:

- Bei der Verarbeitung größerer Dateien hat Kakadu bei der JP2-Kodierung einen klaren Geschwindigkeitsvorteil gegenüber JasPer.
- Die Dekodierung von JP2 nach PNM/PPM dauert ungefähr zweimal so lange wie die Kodierung von PNM/PPM nach JP2 für den JasPer Dienst und ungefähr dreieinhalb mal so lange für den Kakadu Dienst.
- Mit den gewählten Service-Parametern sind die JP2-Objekte, die durch den Kakadu Dienst kodiert wurden, ungefähr 35% kleiner als diejenigen des JasPer-Dienstes.

Bezüglich der Zuverlässigkeit eines Workflow-Pfads wird jeweils bei der Original- und Zieldatei mit Hilfe des PLANETS-Dienstes *XCLCompare*¹⁶ ein Normdatenvergleich durchgeführt. Für den Normdatenvergleich wird eine vom Dateiformat unabhängige Beschreibung der Datei erstellt, die es ermöglicht, auch Dateien des gleichen Dateityps (zum Beispiel Bilder) unterschiedlicher Dateiformate zu vergleichen. Ergebnis dieses Normdatenvergleichs ist ein Kennwert (mittlere quadratische Abweichung - engl. Root mean squared error RMSE), der einen Hinweis auf den Unterschied zwischen Original- und Zieldatei gibt, wobei der kleinste Wert 0 bedeutet, dass die Normdaten absolut identisch sind, wogegen je größer dieser Zahlenwert ist, desto größer ist auch die angenommene Abweichung.

Abbildung 5 und 6 zeigen Diagramme die ~750 Dateimigrationen auf der X-Achse und die jeweilige mittlere quadratische Abweichung (RMSE) auf der Y-Achse geordnet nach der mittleren quadratischen Abweichung (RMSE) als Hinweis auf die Abweichung der Normdaten zwischen Original-Datei und Zieldatei jeweils für den Kakadu- und den JasPer-Dienst.

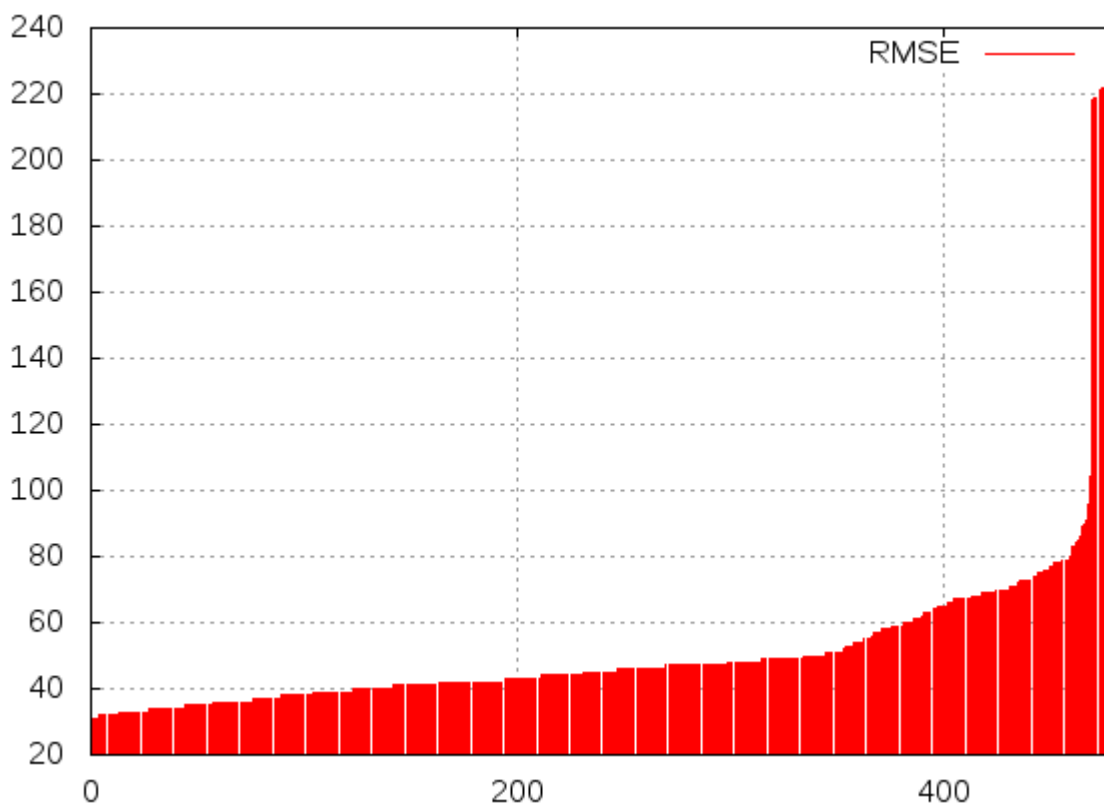


Abbildung 5: ~750 Dateimigrationen auf der X-Achse und die jeweilige mittlere quadratische Abweichung (RMSE) auf der Y-Achse geordnet nach der mittleren quadratischen Abweichung (RMSE) als Hinweis auf die Abweichung der Normdaten zwischen Original-Datei und Zieldatei für den **Kakadu**-Dienst

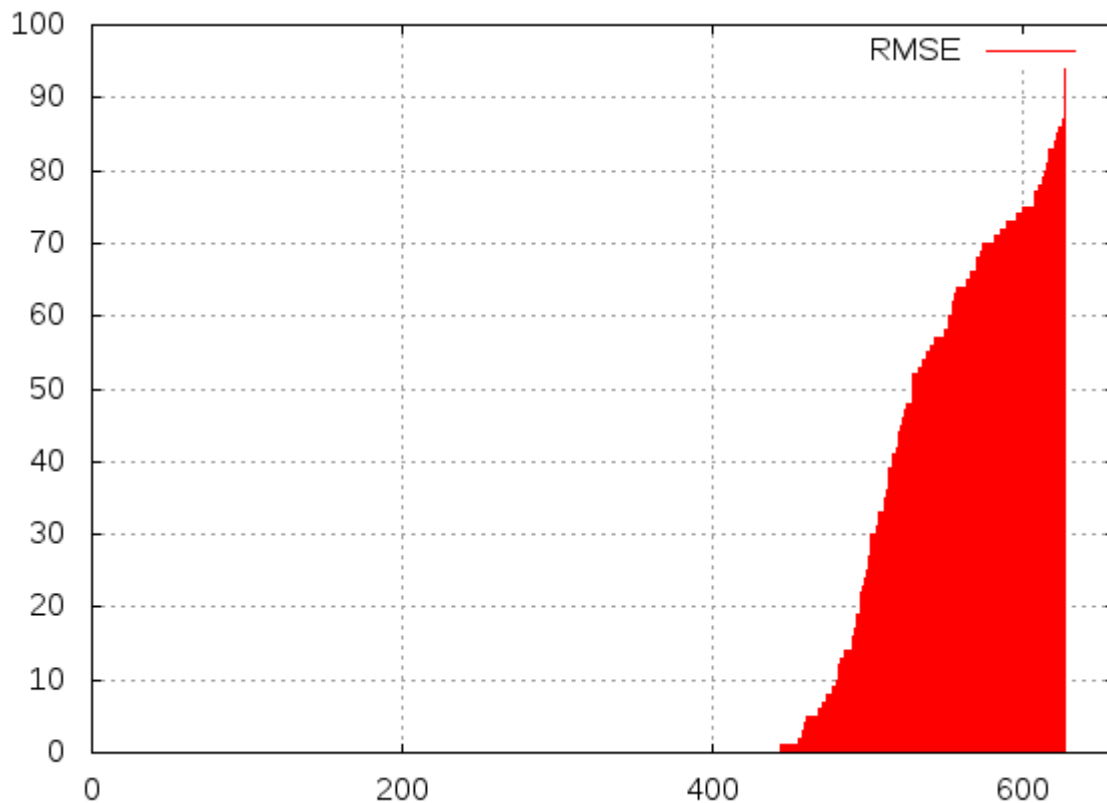


Abbildung 6: ~750 Dateimigrationen auf der X-Achse und die jeweilige mittlere quadratische Abweichung (RMSE) auf der Y-Achse geordnet nach der mittleren quadratischen Abweichung (RMSE) als Hinweis auf die Abweichung der Normdaten zwischen Original-Datei und Zieldatei für den **JasPer**-Dienst

Der Normdatenvergleich gibt lediglich einen groben Hinweis auf das Ergebnis der Migration und leistet keine Beurteilung des Erfolgs auf Basis einer Bild-Qualitätsanalyse.

Schlussbemerkungen

Mit diesem Testbed-Experiment haben wir gezeigt, in welcher Weise das Testbed für empirische Studien im Bereich der digitalen Langzeitarchivierung genutzt werden kann. Das Testbed erwies sich dabei als stabile Anwendung für die strukturierte Planung, Durchführung und Dokumentation dieses Experiments. Insbesondere war es mit Hilfe des Testbed möglich auch größere Datenmengen für das Testen der Skalierbarkeit des gewählten Migrationsszenarios heranzuziehen.

Das Testbed ist als öffentlich zugängliche Webanwendung erreichbar unter

<https://testbed.planets-project.eu>

Zugangsdaten können über den PLANETS Helpdesk angefragt werden:

helpdesk@planets-project.eu

Und die Ergebnisse des PLANETS Projekts wurden zum Ende des Projekts an die Open Planets Foundation weitergereicht, die für die Wartung und Erweiterung der "PLANETS Software Suite" zuständig ist:

www.openplanetsfoundation.org

Anschrift des Autors:

Dr. Sven Schlarb
Abteilung Forschung und Entwicklung
Österreichische Nationalbibliothek
Josefsplatz 1
A-1015 Wien
Österreich
E-Mail: sven.schlarb@onb.ac.at

¹ Bienz, Daniela, Gschwind, Rudolf, Pozza, Mario F., Gantner Ludwig: "Mass CD/DVD Migration: A Novartis Case Study". In: Archiving 2010. Preservation Strategies and Imaging Technologies for Cultural Heritage Institutions and Memory Organisations. Final Program and Proceedings, Springfield 2010.

² "Dienst" meint in diesem Zusammenhang eine Software-Komponente, die über eine Webservice-Schnittstelle genutzt werden kann.

³ Becker, Christoph, Kulovits, Hannes, Rauber, Andreas, Hofman Hans: Plato: a service oriented decision support system for preservation planning, In: JCDL '08: Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries. New York 2008, S. 367-370. Siehe <http://doi.acm.org/10.1145/1378889>

⁴ XCL ist die „XCL eXtensible Characterisation Language“, die mit einer zugehörigen XCL-Software-Suite im Rahmen des PLANETS-Projekts entwickelt wird. XCLCompare ist ein Service, der ein Werkzeug zur Dateiformat-unabhängigen Merkmals-Extraktion (Extractor) sowie ein Werkzeug für den Vergleich Dateiformat-unabhängiger Beschreibungen verwendet. Siehe http://planetarium.hki.uni-koeln.de/planets_cms/about-xcl.

⁵ <http://hul.harvard.edu/jhove/>

⁶ <http://droid.sourceforge.net/>

⁷ Buonora, Paolo, Liberati, Franco: A Format for Digital Preservation of Images, 2008.
URL: <http://www.dlib.org/dlib/july08/buonora/07buonora.html>;

⁸ Ebrahimi, Farzad, Chamik, Matthieu, Winkler, Stefan: JPEG vs. JPEG2000: An Objective Comparison of Image Encoding Quality, 2004.
URL <http://stefan.winklerbros.net/Publications/adip2004.pdf>.

⁹ "The GIMP" (GNU image manipulation program) version 2.6, siehe <http://www.gimp.org>.

¹⁰ Jasper Transcoder Version 1.900.1, siehe <http://www.ece.uvic.ca/~mdadams/jasper>.

¹¹ Kakadu Version 6.2.1 JPEG2000. Die Kommandozeilenanwendung 'kdu_compress' (die Basis-Software des Dienstes KakaduCompress) und 'kdu_expand' (die Basis-Software des Dienstes KakaduDecode) dienen als Beispielimplementierung um die Möglichkeiten der JPEG2000 Bibliothek für Entwickler aufzuzeigen. Siehe <http://www.kakadusoftware.com>.

¹² s. Anm. 4

¹³ TIFF steht für "Tag Image File Format".

¹⁴ PNM steht für "Portable any Map". PNM umfasst die spezifischeren Formate PBM, PGM, and PPM.

¹⁵ JP2 ist ein Dateiformat des JPEG2000 Bildformat-Standards.

¹⁶ s. Anm. 4