

Guideline zur Langzeitarchivierung

Work-Package-Cluster:	Cluster J: Dauerhafte Sicherung der Daten (aus nicht-technischer & technischer Sicht)	
Leitung des Clusters:	Adelheid Mayer	Universitätsbibliothek Wien adelheid.mayer@univie.ac.at
Datum:	10.05.2016	
Version	1.1	
AutorInnen:	Michael Birkner	AK Bibliothek Wien michael.birkner@akwien.at
	Gerhard Gonter	Universität Wien gerhard.gonter@univie.ac.at
	Karin Lackner	Universitätsbibliothek Graz karin.lackner@uni-graz.at
	Bettina Kann	Österreichische Nationalbibliothek bettina.kann@onb.ac.at
	Michael Kranewitter	Universitätsbibliothek Linz michael.kranewitter@jku.at
	Adelheid Mayer	Universitätsbibliothek Wien adelheid.mayer@univie.ac.at
	Andreas Parschalk	Universitäts- und Landesbibliothek Tirol andreas.parschalk@uibk.ac.at



Kurzbeschreibung (Deutsch):	Guideline zur Hilfestellung für die Langzeitarchivierung von Daten und Objekten im Kontext des Publikations- und Forschungswesens.
Description (English):	Guideline for Digital Preservation of data in an academic setting.
Schlagwörter (Deutsch):	Guideline - Langzeitarchivierung – Hochschulen
Keywords (English):	Guideline - Digital Preservation – Universities



Inhalt

Einleitung	5
Beweggründe	6
Datenerhebung	7
Formate.....	9
Volumen	11
Metadaten I	12
System-Auswahl	13
Metadaten II	19
Organisatorisches.....	21
Workflow-Design.....	22
Rechtliche Fragen.....	23

Einleitung

Diese Guideline soll Hilfestellung für die Langzeitarchivierung von Daten und Objekten im Kontext des Publikations- und Forschungswesens bieten. Sie ist ausdrücklich nicht für den Kontext der Compliance anwendbar. Sie soll dazu ermächtigen, die richtigen Fragen zur Auswahl der für die eigene Institution geeigneten Langzeitarchivierungs-Lösung zu stellen und bei der Entscheidung für eine System-Lösung behilflich sein.

Langzeitarchivierungssysteme werden hier als Systeme verstanden, die im Workflow hinter einem Repository stehen, in dem digitale Objekte und ihre Metadaten gespeichert und angezeigt werden sowie recherchierbar sind.

Allfällige Begriffserklärungen finden Sie im Glossar des Clusters C (Aufbau eines Wissensnetzwerks: Erarbeitung eines Referenzmodells für den Aufbau von Repositorien / Themenbereich Terminologie und Standards).

Warum will ich Langzeitarchivierung (LZA) betreiben?

An erster Stelle sollten Sie sich über die Beweggründe Langzeitarchivierung betreiben zu wollen, im Klaren werden:

- Gibt es externe (z.B. gesetzliche/rechtliche/vertragliche) oder interne (z.B. Auftrag des Rektorats) Vorgaben zur Archivierung (Pflichtexemplare, evtl. Verwaltungsdaten, Nachlässe, Vorlässe, Verträge mit Fördergebern)?
- Herrscht bei allen Stakeholdern übereinstimmendes Bewusstsein über die Notwendigkeit von LZA?

„Langzeitarchivierung ist gekennzeichnet durch einen nicht näher bestimmten Zeitraum, der in einigen Fällen Jahre, in anderen Jahrzehnte oder Jahrhunderte betragen kann. In diesem Zeitrahmen verändern sich nicht nur die Dateiformate, sondern auch die Anwendungsprogramme, Betriebssysteme und die Speichermedien.“¹

¹ ITWissen - Das große Online-Lexikon für Informationstechnologie: Langzeitarchivierung:
<http://www.itwissen.info/definition/lexikon/Langzeitarchivierung-long-time-backup.html> 26.04.2016

Was will ich langzeitarchivieren?

Welchen Content (Inhalt, Medientypen, Daten etc.) habe ich?

Verschaffen Sie sich einen Überblick, welche digitalen Inhalte an Ihrer Institution vorhanden sind.

- Publikationen
- Hochschulschriften (wissenschaftliche Abschlussarbeiten)
- Digitalisate alter Handschriften, Bücher, Zeitschriften, Umschlagbilder etc.
- Künstlerische/audiovisuelle Werke
- Digitale Materialien der Architektur und des Industrial Design (Entwürfe mit CAD)
- Angekaufte elektronische Ressourcen: E-Books, E-Journals, Datenbanken
- Web-Seiten
- Forschungsdaten
- Sonstiges

Forschungsdaten sind Daten, "die im Zuge wissenschaftlicher Vorhaben z. B. durch Digitalisierung, Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen."²

Welcher Content (Inhalt, Medientypen, Daten etc.) wird in Zukunft auf mich zukommen?

Versuchen Sie herauszufinden, welche zusätzlichen Inhalte künftig an Ihrer Institution gesammelt werden sollen.

- Forschungsdaten
- ...

² Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen: Forschungsdaten: <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/> 26.04.2016

Für welchen Content suche ich ein System zur Langzeitarchivierung?

Überlegen Sie, welche der vorhandenen Inhalte tatsächlich langzeitarchiviert werden sollen.

- Für welchen Content gibt es andere Lösungen? (z.B. eRessourcen)

Lizenzpflichtige Datenbanken/e-journals, welche nicht in der Institution archiviert werden sollen, können über Serviceprovider archiviert werden. Z.B. JSTOR.

Formate

Welche Formate habe ich?

Erstellen Sie eine Liste mit den Formaten, in denen der an Ihrer Institution vorhandene Content vorliegt. Beachten Sie dabei, welche dieser Formate proprietär sind.

- Textdokumente (z.B. DOC, ODF, PDF, TXT etc.)
- Strukturierter Text (z.B. HTML, JSON, TEX, XML etc.)
- Tabellenarbeitsblätter (z.B. CSV, ODS, XLS, SAS, Stata, SPSS etc.)
- Datenbanken (z.B. MS Access, MySQL, Oracle etc.)
- Grafiken/Bilder (z.B. JPEG, SVG, PNG, GIF, TIFF etc.)
- 2D- und 3D-Entwürfe mit CAD (z.B. STL, IFC, DXF, DWF, VDA-FS, IGES, STEP etc.)
- Audio (z.B. MP3, WAV, AIFF, OGG etc.)
- Video/Film (z.B. MPEG, AVI, WMV, MP4 etc.)
- Quellcode (z.B. CSS, JavaScript, Java etc.)
- Konfigurationsdaten (z.B. INI, CONF etc.)
- Software Applikationen
- Sonstiges

Sind diese Formate für die Langzeitarchivierung geeignet, d.h. entsprechen sie den Standards für die Langzeitarchivierung?

Für die digitale Langzeitarchivierung sind grundsätzlich geeignete Dateiformate zu wählen. Sie müssen langfristig eine originalgetreue Reproduktion der Archivdaten sowie ausgewählter Merkmale des ursprünglichen Dokumentmediums (z. B. Papierformat, Farben, Logos, Seitenzahl, Wasserzeichen, Unterschrift) ermöglichen.³

³ Siehe Empfehlungen des Bundesamts für Sicherheit in der Informationstechnik (BSI): *Auswahl geeigneter Datenformate für die Archivierung von Dokumenten*: https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/Inhalt/_content/m/mo4/mo4170.html 26.04.2016. Langzeiterhaltung digitaler Daten in Museen : Tipps zur dauerhaften Bewahrung digitaler Daten. 07: Dateiformatauswahl: http://files.dnb.de/nesstor/sheets/07_dateiformatauswahl.pdf 26.04.2016.

Will ich nur bestimmte Formate für die (zukünftige) Langzeitarchivierung zulassen und welche?

Bitte beachten Sie: Je weniger unterschiedliche Formate verwendet werden, desto einfacher ist die LZA; offene Formate sind proprietären vorzuziehen; manche Dateitypen sind anfällig für Schadsoftware aller Art.

Volumen

Welche Gesamtgröße (bezogen auf den Speicherplatz) haben die derzeitigen Daten, die archiviert werden sollen?

Schätzen Sie anhand ihrer bisherigen Aufzeichnungen ab, welcher Speicherbedarf derzeit benötigt wird.

Wie groß wird der geschätzte jährliche Zuwachs sein?

Hilfe bei der Abschätzung des jährlichen Zuwachses: Wieviel wird jährlich an der Institution publiziert bzw. an Forschungsdaten erhoben, wieviel Prozent davon sollen archiviert werden?

Je nach Gesamtgröße und geschätztem Zuwachs muss das System in einem angemessenen Zeitraum diesem Datengesamtvolumen performant gewachsen sein.

Wo werden die Daten derzeit abgespeichert?

Verschaffen Sie sich einen Überblick, wo der Content Ihrer Institution gespeichert wird.

- Zentral auf einem Server der Institution
- Bei einem externen Datenzentrum
- Bei einem Cloud-Service
- Lokal auf einem persönlichen Rechner
- Auf einer externen Festplatte/auf einem USB-Laufwerk
- Auf CDs/DVDs
- Auf Magnetbändern
- Direkt am Gerät/Instrument
- Sonstiges

Metadaten I

Für die Migration von Daten in ein neues System ist es notwendig zu wissen, welche Metadaten bereits vorhanden sind und aus welchen Informationen neue Metadaten generiert werden könnten.

Werden die Daten in der Regel beschrieben, d.h. liegen bereits Metadaten vor?

- Ja, unter Verwendung geeigneter Standards.
- Ja, individuell und einheitlich.
- Ja, individuell und nicht einheitlich.
- Nein.
- Ich weiß es nicht.
- Sonstiges.

Gibt es bereits eine für die Institution allgemein gültige Konvention der Dateibenennung?

Für die Ableitung von Metadaten aus Filebenennungen ist eine Konvention von Vorteil. In einem LZA-System ist dringend empfohlen, dass Dateinamen eindeutig (unique) sind.

System-Auswahl

Gibt es an meiner Institution bereits ein Institutionelles Repository (IR)?

- Will ich dieses Repository weiterführen?
- Kann ich es um die Funktionalität der LZA erweitern?
- Wie kann ich es in den Workflow einer davon unabhängigen LZA Lösung einbinden?
- Will ich dieses IR durch eines mit LZA Funktionalität ersetzen?
- Wie kann ich die Daten in das neue System migrieren?

Grundsätzliche Anforderungen an ein Langzeitarchivierungs-System

Die wichtigsten Anforderungen an ein Langzeitarchivierungs-System sind im ISO-Standard des OAIS-Modells beschrieben.

Weist das System verbindlich OAIS-Konformität auf?

Das Open Archival Information System ist ein Referenzmodell für ein dynamisches, erweiterungsfähiges Archivinformationssystem. Als ISO-Standard 14721:2012 veröffentlicht, ist es der wichtigste Standard für digitale Langzeitarchivierung.⁴

⁴ Deutsche Übersetzung durch das nestor-Kompetenznetzwerk. Siehe nestor-Arbeitsgruppe OAIS-Übersetzung/Terminologie: Referenzmodell für ein Offenes Archiv-Informationssystem. Deutsche Übersetzung 2.0 (nestor-materialien 16):. http://files.d-nb.de/nestor/materialien/nestor_mat_16-2.pdf
26.04.2016

Datensicherheit

Unter Datensicherheit fallen drei Aspekte: physikalische Sicherung (Bitstream Preservation), Schutz vor unberechtigten Zugriffen und Manipulationen, Verwendbarkeit veralteter Datenformate.

- Welche Unterstützung bietet das System zur Erkennung veralteter Datenformate und für Migrationsszenarien?
- Gibt es ein Monitoring / Alert-System zur Datenerhaltung? Hat das LZA-System einen Mechanismus zur Prüfung der Datenintegrität?
- Wie wird die Sicherheit meiner Daten gewährleistet? (z.B. redundante Datenhaltung, standortübergreifende Replikation, Hashvalidierung)
- Werden die Daten vor Manipulation durch Dritte geschützt? (z. B. Prüfung auf Schadsoftware, etc.)
- Werden die Archival Information Packages (AIP) in einem computerlesbaren, offenen und standardisierten Format (z.B. METS) beschrieben?

Dadurch wird sowohl die Recovery-Möglichkeit, als auch die Search/Retrieve Funktionalität erleichtert.

Sollen Daten nur gespeichert oder auch zugänglich gemacht werden?

Überlegen Sie, wer in welcher Form Zugriff auf die digitalen Inhalte haben soll.

- Nur Langzeitarchivierung
- Langzeitarchivierung und Webpräsentation (mit Viewer)
- Öffentlich zugänglich oder geschlossener Benutzerkreis

Wenn das Langzeitarchivierungs-System keine Präsentationskomponenten aufweist, müssen entsprechende Schnittstellen für die Anbindung eines externen Viewers vorhanden sein.

Software-Modelle / Welches Implementierungs-Modell kommt für mich in Frage?

Es gibt unterschiedliche Software-Modelle:

- Proprietär oder Open Source
- Onsite oder Software as a Service (SaaS)
- Kauf oder Eigenentwicklung
- Mischformen

Entscheiden Sie anhand der nachfolgenden Fragen, welches der vier Modelle für Ihre Institution am besten geeignet ist. Vorweg einige innerbetriebliche Fragen, die Möglichkeit der Auswahl mitbestimmen.

Innerbetriebliches

- Gibt es eine institutionelle Strategie (Vorgabe) für die Verwendung bestimmter Lizenz- oder Implementierungs-Modelle?
- Existieren gesetzliche oder institutionelle Vorschriften, die beim Einsatz der Technik berücksichtigt werden müssen (z. B. vorgeschriebene Betriebssysteme, Bestimmungen zu Off-site Datenhaltung, etc.)?
- Ist Mandantenfähigkeit (mehrere Mandanten/Kunden auf einem Server/System) nötig?
- Habe ich ausreichend technisches Knowhow sowie personelle und zeitliche Ressourcen im Betrieb?
 - Für Entwicklung und eigenen Betrieb von Software:
 - SystemadministratorInnen
 - Entwicklungs-Abteilung
 - SystemarchitektInnen
 - Für Installation und Betrieb von extern entwickelten Open Source Produkten:
 - SystemadministratorInnen
 - Eventuell EntwicklerInnen
 - Für Betrieb von kommerzieller Software:
 - SystemadministratorInnen
 - Für Betrieb von Software as a Service:
 - SystemadministratorInnen
 - Für eigene Programmierung zur Nutzung von Schnittstellen:
 - EntwicklerInnen

Wichtige Kriterien für die Auswahl des Implementierungs-Modells

Je nachdem, für welches Modell Sie sich entscheiden, beachten Sie bitte folgende Punkte.

Kommerzielle Produkte

- Kann ich das Produkt „Out-of-the-Box“ einsetzen oder benötige ich zusätzliche Entwicklung?
- Gibt es eine Community für das Produkt?
- Wie garantiert die Firma die Weiterentwicklung des Produkts?
- Welches Lizenz-Modell wird angeboten?
- Wie ist der Support gestaltet? Können Änderungsvorschläge eingebracht werden?
- Wie zugänglich bzw. intransparent arbeitet das System? Wie kann/muss man eingreifen?
- Welche Konfigurationsmöglichkeiten werden geboten?
- Was besagen die Vertragsdetails hinsichtlich Service-Level-Agreement (SLA), Antwortzeiten bei Supportanfragen, etc.?
- Gibt es ein Exit-Szenario?
- Wird der Quellcode der Software an mich übergeben, falls das Unternehmen aufgelöst werden sollte (Konkurs, Übernahme, etc.), und kann ich den Quellcode ohne Einschränkungen weiterentwickeln?
- Wie vertrauenswürdig ist die Firma?
- Gibt es Referenz-Anwendungen, die ich ansehen bzw. kontaktieren kann?

LZA as a Service

- Welche Bandbreite wird benötigt, um vernünftige Übertragungsgeschwindigkeiten gewährleisten zu können?
- Welche Garantien in Bezug auf Ausfallsicherheit, Sicherheit vor Fremdzugriff und Datenintegrität kann mir der Dienstleister bieten?
- Langfristigkeit der Partnerschaft: Ist der Dienstleister eine öffentliche Institution oder ein Privatunternehmen?
- Was passiert, falls das Unternehmen aufgelöst wird? (Kann der SaaS-Betrieb an eine andere Firma weitergeben werden oder kann ich ihn selbst fortführen?)
- Gibt es Erfahrungswerte mit dem Dienstleister an einer anderen Institution?

Offene Modelle

- Kann ich das Produkt „Out-of-the-box“ einsetzen oder benötige ich zusätzliche Entwicklung?
- Gibt es eine Community, die gemeinsam an der Weiterentwicklung arbeitet?
- Welches Lizenz-Modell wird geboten?
- Wie hoch ist der Aufwand, das Open Source Modell an meine Bedürfnisse anzupassen?
- Gibt es kommerzielle Anbieter, die Entwicklungsarbeit anbieten?
- Gibt es Widersprüche zu den internen Vorgaben bezüglich Programmiersprachen, bestehenden Systemen, Betriebssystem, etc.?

Weitere Auswahlkriterien

Die folgenden Kriterien treffen auf alle Implementierungs-Modelle (kommerzielle Installation, Service, Open Source) zu.

Allgemein

- Gibt es Referenz-Implementierungen aus vergleichbaren Institutionen und mit vergleichbaren Datenmengen? Kann ich diese sehen?
- Welche Sprache wird für die Kommunikation mit den Herstellern benutzt und welche Kanäle sind dafür vorgesehen?
- Gibt es ein Exit-Szenario? (Wie komme ich an alle meine Daten heran?)
- Gibt es eine ausreichende technische Dokumentation für das LZA-System bzw. die Hardware?
- Wie zugänglich bzw. intransparent arbeitet das System?
- Kann/muss man eingreifen?
- Welche Konfigurationsmöglichkeiten werden geboten?

Technische Fragen

- Wie hoch kann das LZA-System bzw. die Hardware skaliert werden (Datenmenge, Anzahl gleichzeitiger NutzerInnen bzw. Prozesse, etc.)?
- Gibt es ein Monitoring-System, das AdministratorInnen über kritische Abläufe oder Fehler benachrichtigt?
- Welche Hardwareressourcen benötigt das System?
- Welche Datenbanken und Betriebssysteme unterstützt das LZA-System?
- Welche Daten habe ich und wie groß muss der Speicherplatz dafür bemessen werden? Ist der Speicherplatz inhouse oder extern? Speicherplatz für Replikation?

- Lassen sich die archivierten Inhalte auf unterschiedliche Speicherklassen (Storage Tiers) aufteilen?
- Muss zusätzliche lokale Entwicklungsarbeit geleistet werden (z. B. Skripte zur Datenmigration, etc.)
- Welche Schnittstellen bietet das LZA-System? Entsprechen diese Schnittstellen gängigen Standards?
- Welche Bandbreite wird benötigt, um vernünftige Übertragungsgeschwindigkeiten gewährleisten zu können?
- Codierung (UTF8, ISO, ...)

Usability

- Wie bedienungsfreundlich ist die Benutzeroberfläche?
- Ist die Benutzerführung effizient gestaltet?
- Gibt es ein ausführliches Benutzerhandbuch?
- Kann die Benutzeroberfläche gemäß den eigenen Anforderungen angepasst werden?

Metadaten II

Müssen im Rahmen der LZA zusätzliche Metadaten generiert werden? ⁵

Metadaten werden üblicherweise diesen Kategorien zugeteilt:

- Technische Metadaten: Auflösung, Dateiformat, Speichermedium etc. z.B. Exif, GeoTIFF
- Inhaltliche Metadaten: Angabe des Inhalts, Schlagworte, Titel etc. z.B. MODS, LOM
- Strukturelle Metadaten in Zusammenhang mit anderen Objekten : "Teil von", "Version von" etc. z.B. ZB RDF bei der Verwendung von standardisierten Prädikaten⁶, METS
- Administrative Metadaten: Hersteller, Nutzungsbedingungen, Rechte etc. z.B. METS

Im Folgenden finden Sie einen Überblick zu Metadatenstandards.

Empfehlungen zur Verwendung von Metadatenstandards⁷:

- Der Premis-Standard soll berücksichtigt werden.
- Informationen zu Metadaten: siehe auch Ergebnisse des e-Infrastructures Austria Clusters „Metadatenkomplex aus nicht-technischer und technischer Sicht“.

Für Forschungsdaten:

- CERA-2 (Simulationsdaten der Klimaentwicklung)
- FGDC, ISO 19115:2003 (Geodaten, Georeferenzierung von Daten)
- QuakeML (seismologische Daten)
- CDISC (Daten klinischer Studien)
- SDMX (statistische Daten)
- TEI (Geisteswissenschaften)
- DDI (sozialwissenschaftliche Daten)
- ...⁸

⁵ Siehe nestor Informationsblätter : Langzeiterhaltung digitaler Daten in Museen, Tipps zur dauerhaften Bewahrung digitaler Daten. 06: Metadaten: http://files.dnb.de/nestor/sheets/06_metadaten.pdf 26.06.2016

⁶ Siehe z.B. <http://purl.org/dc/terms/> 26.04.2016.

⁷ Siehe auch nestor Handbuch : eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 2.0.

Kapitel 6: Metadatenstandards im Bereich der digitalen LZA: [http://nestor.sub.uni-](http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_360.pdf)

[goettingen.de/handbuch/artikel/nestor_handbuch_artikel_360.pdf](http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_360.pdf) 26.04.2016. Stefan Hein: Metadaten für die Langzeitarchivierung. 15. September 2011: <http://files.dnb.de/nestor/presentationen/Gesamt/hein.pdf>

26.04.2016. Wikipedia: Liste von Metadatenformaten:

http://de.wikipedia.org/wiki/Liste_von_Metadatenformaten 26.04.2016.

⁸ Siehe auch Uwe Jensen: Metadaten für Forschungsdaten: Welche Standards gibt es? 2012:

<http://www.opus-bayern.de/bib->

[info/volltexte/2012/1283/pdf/Metadatenstandards_Welche_gibt_es_Btag2012_Uwe_Jensen.pdf](http://www.opus-bayern.de/bib-info/volltexte/2012/1283/pdf/Metadatenstandards_Welche_gibt_es_Btag2012_Uwe_Jensen.pdf) 26.04.2016

Für Publikationen:

- Dublin Core
- METS
- MODS
- MAB
- MARC
- PREMIS
- LMER
- TEI
- XDOMEA
- EAD (Beschreibung von Archivgut)
- EAC (Beschreibung von Provenienzen)
- ...

Für künstlerische/audiovisuelle Werke :

- VRA Core (Visuelle Objekte: Bilder, Filme, Gegenstände...)
- NISO (technische Zusatzinformationen für Bilddateien)
- Exif (technische Zusatzinformationen in Bilddateien von Digitalkameras)
- CDWA, CDWA Lite (Beschreibung von Objekten der Kunst und Kultur)
- IPTC-IIM (Textinformationen in Bilddateien)
- RDF (Beschreibung von Webressourcen)

Wenn digitale Objekte bereits Metadaten enthalten (z.B. EXIF-Daten in Digitalfotografien) -> wie erfolgt die automatische Extraktion?

Beachten Sie, dass die automatische Extraktion von Metadaten in der Regel in die Phase der Datenaufbereitung vor dem Ingest fällt.

Organisatorisches

Wie bei jedem Projekt gilt es, das organisatorische Umfeld zu beachten!
(Projektmanagement!!!)

Allgemein

- Gibt es ein Risiko-Management (Ausfallsicherheit, Recovery-Plan, etc.)?
 - Wer ist verantwortlich für die Wiederherstellung der Daten im Fehlerfall?
 - Werden manuelle Überprüfungen der archivierten Daten durchgeführt?
- Wie wird das Projektmanagement aufgestellt?
- Wer erstellt den Anforderungskatalog (Requirements-Engineering)?
- Welche Schulungen sind für wen notwendig?

Personal

- Welche Abteilungen und Personen sind am Archivierungsworkflow beteiligt?
- Welche Kompetenzen sind dabei nötig (Vertrautheit mit dem Archivierungssystem, Fileformaten, Konvertierung/Normalisierung, Qualitätskontrolle, Metadaten)?
- Wer ist für die LZA verantwortlich?
- Habe ich ausreichend Personal und Expertise in folgenden Bereichen?
 - Metadaten
 - Bestandsmanagement und/oder Preservation Planning
 - Prozessmanagement
 - Projektmanagement
 - Technik

Finanzierung

- Habe ich ausreichende Finanzierungszusagen, auch für die laufenden Kosten?
- Kann ich (alternativ) Kooperationen eingehen? Welche Kooperationsmodelle gibt es?
- Wieviel Budget steht mir für die Technik der LZA zur Verfügung?
- Werden Kosten intern zw. Abteilung, Instituten, etc. verrechnet?

Workflow-Design

Überlegen Sie sich im Vorhinein, wie der Arbeitsablauf zu folgenden Punkten gestaltet werden soll.

- Wie gelangen die Daten ins Archiv?
 - Ausschließlich automatisiert?
 - Können Daten manuell in das System eingebracht werden?
- Wer hat in welchem Umfang Zugriff auf welche Daten (intern vs. extern, AIP, DIP, Beschränkungen, etc.)?
- Wie werden Objekte aus dem LZA ausgeschieden? Wie werden Löschungen dokumentiert?
- Wo und wie werden deskriptive Metadaten für die archivierten Objekte vergeben?
- Wie werden die archivierten Objekte referenziert? Wird ein oder werden mehrere System(e) zur Vergabe von Persistent Identifiern verwendet?
- Wann sollen die Dissemination Information Packages (DIP)⁹ im Workflow generiert werden?
- Wie, in welchen Intervallen und von wem werden Datenintegritätstests (Fixity checks) durchgeführt bzw. die Ergebnisse kontrolliert?
- Muss innerhalb der Institution ein Support für das System eingerichtet werden (z.B. internes Ticket-System)?

⁹ Das DIP ist die Ansicht eines Datenpakets, das an BenutzerInnen ausgeliefert wird. Siehe auch OAIS-Model.

Rechtliche Fragen

Eine gute juristische Unterstützung ist bei der Behandlung von Rechtsfragen empfehlenswert!

- Hat die Institution das Recht, die Daten zu speichern?
- Wer hält das geistige Eigentum an den gespeicherten Daten?
- Gibt es Verträge mit den Urhebern? (z.B. Publikationsverträge). Wie wird das dokumentiert?
- Wird eine LZA-Policy veröffentlicht, die klar ersichtlich die Verwendung der Daten erläutert?
- Gibt es institutionsinterne Vorgaben und Regeln? (z.B. Verankerung der Sammlung von forschungsbezogenen Daten in der Satzung der Universität)
- Wie wird die „Ownership“ im System abgebildet?
- Werden bei Veröffentlichung der Daten Lizenzen verwendet (z.B. CC-Lizenzen)?
- Gibt es Daten, die speziell gesichert archiviert werden müssen? (Sensible Daten/Datenschutz)
- Gibt es rechtliche Vorgaben zur Zugangsbeschränkung der archivierten Daten (Datenschutz, Urheberrecht)?
- Gibt es Vorgaben zu einzuhaltenden Standards und Normen (z. B. ISO 14721:2003 OAIS, BCP/Disaster Recovery, ...)?
- Gibt es Ablauffristen dieser Vorgaben?
- Bei proprietären Produkten: Erhält die Firma das Recht an den Daten (z. B. Metadaten)? Gibt es Escrow-Vereinbarungen¹⁰ bei einer Insolvenz?

¹⁰ Software-Escrow ist die Hinterlegung der Quelltexte sowie der Software-Dokumentation bei einem unabhängigen Unternehmen oder Notar.

e-Infrastructures Austria

Sustainable data storage and the provision of data for use by third parties are the central roles of science. e-Infrastructures Austria is a federally funded program for the coordinated expansion and continued development of data repositories across Austria, and is made possible by a grant from the Austrian Ministry of Science, Research and Commerce (BMWFW). This program enables the safe archival and lasting availability of electronic publications, multimedia objects and other digital data from the research and teaching fields. Concurrently, topics relating to research data management and digital archiving workflows will be addressed.

The working area is organized in twelve Work-Package-Clusters:

Cluster A	Monitoring of Document Repositories within the Partner Network <i>Patrick Danowski (IST Austria)</i>
Cluster B	Planning and Implementation of a „National Survey“ for Research Data <i>Christian Gumpenberger (University of Vienna)</i>
Cluster C	Designing a Knowledge Network: Development of a reference structure for the construction of Repositories <i>Paolo Budroni (University of Vienna)</i>
Cluster D	Infrastructure <i>Raman Ganguly (Vienna University Computer Center)</i>
Cluster E	Legal and Ethical Issues <i>Seyavash Amini (Counsellor-at-law, University of Vienna)</i>
Cluster F	Open Access <i>Andreas Ferus (academy of fine arts vienna)</i>
Cluster G	Visual Data modeling <i>Martin Gasteiner (University of Vienna)</i>
Cluster H	Life Cycle Management <i>Andreas Rauber (Technical University Vienna)</i>
Cluster I	Metadata <i>Susanne Blumesberger (University of Vienna)</i>
Cluster J	Permanent backup of the data <i>Adelheid Mayer (University of Vienna)</i>
Cluster K	Data from scientific and artistic-scientific research processes <i>Bernhard Haslhofer (Austrian Institute of Technology)</i>
Cluster L	Cross-project issues (technical and non-technical) <i>Andreas Jeitler (University of Klagenfurt)</i>