

Leitlinien für die Fortführung und die Langzeitspeicherung von Geoinformationen

Arbeitskreis Architektur
der GDI-DE

Version	1.0
Datum	26.10.2021
Status	Final
Autoren	Falk Würriehausen, Iris Kohnen, Mathis Eckhoff, Anja Schupp, Tillmann Lübker, Eszter Kiss
Herausgeber	AK Architektur
Ablage	GDI-DE Wiki
Haftungsausschluss	Dieser Bericht fasst die Ergebnisse der Architekturmaßnahme A1.13 zur Langzeitspeicherung von Geoinformationen in der GDI-DE zusammen. Aussagen daraus können nicht im Rahmen gesetzlicher Verpflichtungen und deren Umsetzung geltend gemacht werden.

Änderungshistorie

Version	Datum	Änderung	Autor
0.1	30.10.2020	Initiale Version	Falk Würriehausen
0.2	30.11.2020	Ergänzte und kommentierte Version	Iris Kohnen, Anja Schupp, Tillmann Lübker, Sandra Rein, Maren-Lisa Kyas, Mathis Eckhoff, Eszter Kiss
0.3	29.01.2021	Anpassungen zu Definition und Gliederung des Dokumentes	Falk Würriehausen
0.4	01.07.2021	Erhebungsbogen und fachliche Ergänzung	Falk Würriehausen
0.5	01.10.2021	Inhaltliche Ergänzung und Fazit	Falk Würriehausen
0.9	05.10.2021	Vorlage und Abstimmung im Arbeitskreis Architektur	AK Architektur
0.91	11.10.2021	Kommentierung und Ergänzung HH	Iris Kohnen, Mathis Eckhoff
0.92	21.10.2021	Bearbeitung	Falk Würriehausen, Iris Kohnen, Mathis Eckhoff, Anja Schupp, Eszter Kiss
1.0	26.10.2021	Finalisierung	Falk Würriehausen

Inhaltsverzeichnis

1. Einleitung	4
2. Begriffe der Langzeitspeicherung und Langzeitarchivierung	5
3. Abgrenzung der Langzeitspeicherung und GDI-DE	6
4. Konzept für die Langzeitspeicherung	7
4.1 <i>Struktur der Speicherung (Wo?)</i>	7
4.2 <i>Umsetzungsanforderung an das System (Was?)</i>	8
4.3 <i>Zeit der Speicherung (Wann?)</i>	10
5. Digitale Geodaten – Bestandsaufnahme und Beschreibung	10
5.1 <i>Bestandsaufnahme</i>	10
5.2 <i>Art der Geodaten</i>	11
5.3 <i>Größe der Geodaten</i>	11
5.4 <i>Datenmenge</i>	11
5.5 <i>Datenformate</i>	11
5.6 <i>Datenschutz</i>	11
6. Anwendungsszenarien	12
6.1 <i>Bereitstellung im Übergangsarchiv</i>	12
6.2 <i>Registrierung und Verwaltung</i>	13
6.3 <i>Bestandserhaltung und Abgabe</i>	14
7. Fazit	15
8. Quellen	16
Anlage 1:	17
Erhebungsbogen	17

1. Einleitung

Im Zielbereich B „Wirtschaftlichkeit und Effizienz“ der **Nationalen Geoinformationsstrategie (NGIS)** ist die versionierte Langzeitspeicherung (LZS) und -archivierung von Geoinformationen verankert. Mit diesem Werkzeug soll die Recherche und Nachvollziehbarkeit zeitlicher Entwicklungen ermöglicht werden. Bisher werden in der Geodateninfrastruktur Deutschland (GDI-DE) allerdings nur „in Verwendung stehende Geodaten“ geführt und gepflegt. Aufgrund neuer Anforderungen und Interessen, wie beispielsweise der Ermittlung und Nachvollziehbarkeit von ausgewählten zeitlichen, gesellschaftlichen oder umweltbedingten Veränderungen, besteht jedoch der Bedarf, neben den aktuell in Verwendung befindlichen Geodaten der GDI, auch die nicht mehr regelmäßig verwendeten Geodaten sicher vorzuhalten und entsprechend nachhaltig verfügbar zu machen. Diese Anforderungen soll eine LZS erfüllen.

Aus diesem Grund hat die **GDI-DE** dieses Thema nunmehr auf seine Agenda genommen. Um eine sinnvoll strukturierte und nachhaltige LZS von versionierten Geobasis- und Geofachdaten aufzubauen und wirtschaftlich zu pflegen, sollen neue Erhebungs- und Fortführungsprozesse konzeptionell erarbeitet werden. Hierfür soll auch auf entsprechende, bereits vorhandene, Konzepte zurückgegriffen werden. Zur Verfügung stehen hier zum einen das Konzept „Langzeitspeicherung digitaler Geodaten“ aus Hamburg sowie weitere relevante Vorgaben und Anforderungen aus Bund und den Ländern, z. B. dem Konzept der AdV „Leitlinien zur bundesweit einheitlichen Archivierung von Geobasisdaten“, das „Organisationskonzept elektronische Verwaltungsarbeit – Baustein E-Langzeitspeicherung (BMI)“ und die neue „ISO-Norm zur Langzeitarchivierung von Geodaten und Metadaten“.

Mit einer neuen Maßnahme im Maßnahmenkatalog des AK Architektur der GDI-DE sollen die Auswirkungen der LZS auf die Architektur der GDI-DE untersucht, Notwendigkeiten zur Anpassung der Architekturdokumente geprüft und technische Möglichkeiten von Langzeitdatenspeicherung und Anwendungsübertragung der LZS-Methoden auf die GDI-DE erarbeitet werden. Mögliche Risiken, welche sich als problematisch bei der Realisierung des Konzepts erweisen können sind zum einen, wenn sich ein Dissens über die Definition von Langzeitspeicherung und dauerhafter Verfügbarkeit ergeben. Diese müssen noch vor Beginn jeglicher Umsetzung und Schaffung von Infrastrukturteilen in der GDI fest definiert sein, da die Anforderungen und Strukturformen entsprechend anzupassen sind. Zum anderen muss die Abgrenzung zur Archivierung klar festgelegt sein. Nach dieser Definition ergibt sich auch die Abgrenzung des Verantwortungsbereichs gegenüber den Bundes- und Landesarchiven. Auch ist diese Aufgabe der Langzeitspeicherung ein neues und umfangreiches Aufgabenfeld mit vielen Anforderungen. Dieses neue Aufgabenfeld erfordert neue Ressourcen, sowohl materiell (Server, Programme, etc.) als auch personell. Die aufgeführten Risiken sind dabei ebenfalls zu betrachten und zu berücksichtigen, um eine erfolgreiche Lösung für die Langzeitdatenspeicherung der Geodaten zu ermöglichen.

2. Begriffe der Langzeitspeicherung und Langzeitarchivierung

Unter dem Begriff der **Langzeitspeicherung** wird die Auswahl, revisions sichere Aufbewahrung, Erhaltung und Wiedernutzbarmachung von älteren, nicht mehr regelmäßig verwendeten elektronischen Dokumenten/Geodaten verstanden. Die Daten der Langzeitspeicherung werden für den Zeitraum der Aufbewahrungsfrist und mit einer definierten Übergabe von Datenpaketen an die Archivverwaltungen langzeitgespeichert. Damit grenzt sich die Langzeitspeicherung von der **Langzeitarchivierung** ab, mit der eine dauerhafte Aufbewahrung und Erhaltung von Unterlagen durch Archive nach Ablauf der Aufbewahrungsfrist zur Bewahrung historischer Überlieferungen gemeint ist. Zur Archivierung gehören die Festlegung der archivwürdigen Datensätze bzw. Eigenschaften eines Datensatzes und die Wiederherstellung für die Nutzbarmachung. Die Regelungen und Anforderungen werden in verschiedenen Gesetzen (z. B. Bundesarchivgesetz) und Verordnungen (z. B. Signaturverordnung) definiert. Die Festlegung der Eigenschaften eines Datensatzes werden u. a. in den AdV-KLA-Leitlinien vorgeschlagen. Für textliche Dokumente ist zudem die Langzeitspeicherung bereits in Form der E-Akte vorgesehen (siehe Abbildung 1).

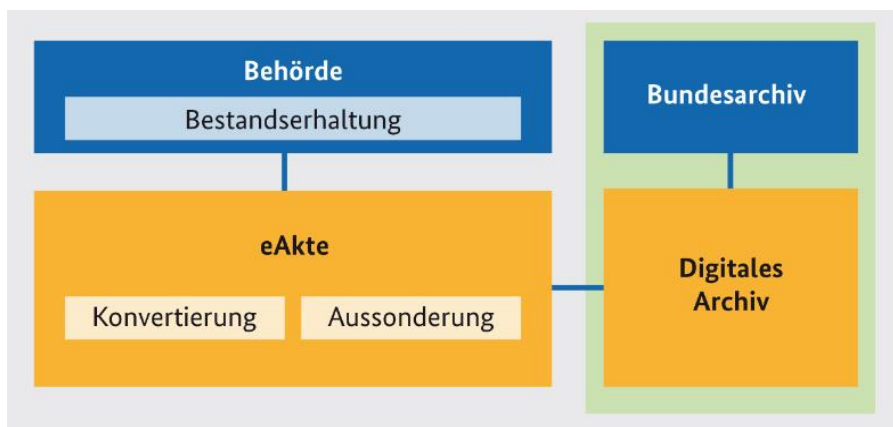


Abbildung 1: Langzeitspeicherung durch eAkte [BMI, 2014]

Um dabei alle wichtigen Einflussfaktoren zu berücksichtigen und die **Bestandserhaltung** sicherzustellen, ist eine entsprechende Konzeption zu entwickeln und umzusetzen. In der praktischen Arbeit umfasst dies zum Beispiel die Überwachung der Lesbarkeit von Formaten, deren Validierung und **Konvertierung** oder auch die **Migration** auf andere Speichermedien. Im Folgenden wird dies daher als ein technischer Prozess verstanden, um elektronischen Dokumente/Geodaten zu sichern und in unterschiedlichen Formaten wiederherstellen zu können. Unter **Aussonderung** wird das Anbieten von elektronischen Dokumenten/Geodaten, welches die Verwaltung für ihre laufende Aufgabenerfüllung nicht mehr benötigt (i. d. R. nach Ablauf der Aufbewahrungsfrist), gegenüber dem zuständigen Archiv verstanden, einschließlich der Übergabe der archivwürdigen und der Vernichtung der nicht archivwürdigen Unterlagen. Die Festlegung der archivwürdigen Datensätze bzw. Eigenschaften eines Datensatzes sollte von der geodatenhaltenden Stelle und dem Archiv gemeinsam geschehen.

Ein weiterer wichtiger Begriff zur Abgrenzung der Langzeitspeicherung von der Archivierung ist die **Aufbewahrungsfrist**. Diese bezeichnet den Zeitraum, nach dessen Ablauf die Dokumente/Geodaten als Archivgut behandelt werden und entsprechend den Regelungen des Landes- und Bundesarchivgesetzes aus dem eigenen Datenbestand gelöscht werden. Die Aufbewahrungsfrist wird von der geodatenhaltenden Stelle und dem Archiv gemeinsam festgelegt und beträgt maximal 30 Jahre. Aufbewahrungsfristen, die aus datenschutz- oder sonstigen rechtlichen Gründen behördenübergreifend verbindlich geregelt sind, sowie die einschlägigen Rechts- und Verwaltungsvorschriften sind zu beachten. Die Aufbewahrungsfristen beziehen sich dabei immer auf die ältere Version des elektronischen Dokumentes/der Geodaten. Sie beginnt mit Ablauf des Kalenderjahres, in dem die Bearbeitung abgeschlossen worden ist, und endet mit dem Ablauf eines Kalenderjahres.

Durch **Versionierung** erhalten elektronische Dokumente/Geodaten nach Änderung eine neue Versionsnummer. Im Falle kleinerer Änderungen wird lediglich die Ziffer nach dem Punkt der Versionsnummer, nach größeren Änderungen die Hauptkennziffer geändert. Bei europäischen und BKG-Produkten wird meist die Jahreszahl statt einer Versionsnummer verwendet, z.B. EBM 2020 oder CLC5-2018.

Im gesamten Kontext der Langzeitspeicherung wird es also immer darum gehen, über die Zeit der Aufbewahrungsfristen, seitens eines IT-Systems bzw. der Infrastruktur den Bestand an Daten langfristig zu sichern bzw. zu erhalten und die **nachhaltige Verfügbarkeit** der Bestandsdaten in der Langzeitspeicherung zu gewährleisten. Zudem sollten Datenverluste oder Formatinkompatibilitäten über die Zeit vermieden und im Sinne der nachhaltigen Verfügbarkeit als Anforderung der Langzeitspeicherung berücksichtigt werden.

3. Abgrenzung der Langzeitspeicherung und GDI-DE

In den letzten Jahren wurde durch **verschiedenste Umsetzungsvorgehensweisen** versucht, eine möglichst umfangreiche Bereitstellung und Mehrfachnutzung von Geodaten zu realisieren. Für die Verwaltung und Haltung der aktuellen Geodaten ist prinzipiell die datenhaltende Stelle selbst verantwortlich. Über unterschiedliche Dienste realisiert sie aufgrund gesetzlicher Regelungen (z.B. Geodatenzugangsgesetz - GeoZG) die Anbindung an die Geodateninfrastruktur Deutschland (GDI-DE). Insgesamt sollten und wurden von den jeweiligen Geodatenhaltern primär die in aktueller Verwendung befindlichen Geodaten einem breiten Nutzerkreis zugänglich gemacht. Es werden überwiegend Daten geführt und bereitgestellt, die den neuesten Informationsgehalt besitzen.

In bestimmten Fällen, z.B. weil auch die **historischen Stände** regelmäßig einem breiten Interessentenkreis zur Verfügung stehen sollen, werden zusätzlich auch ältere Versionen der aktuellen Datensätze über Dienste der GDI bereitgestellt. Ein Beispiel wäre die Langzeitdatenspeicherung der Wasserstände und Abflüsse der Oberflächengewässer Mecklenburg-Vorpommern in der sogenannten LOWO. Diese Geodaten werden jährlich

aktualisiert und die „veralteten“ Daten weiterhin geführt. Sowohl die aktuellen als auch die älteren Geodaten sind über einzelne Dienste abrufbar bzw. zugänglich. Dadurch sind eine langfristige Analyse sowie Vergleich von Geodaten aus verschiedenen Epochen möglich. Andere Beispiele sind im Bereich der Meteorologie zu finden. Diese Datenhalter führen ihre Wetteraufzeichnung über längere Zeiträume, um aus ihnen Vorhersagen und Tendenzen für zukünftige Entwicklungen abzuleiten. Dieser Fall wird im weiteren Dokument als „aktuell in Verwendung befindliche Geodaten“ betrachtet.

Wenn Daten, egal welcher Art, bisher ihren aktuellen Nutzen verloren hatten, wurden sie entweder für die **historische Aufbewahrung**, also Archivierung, ins Archiv überführt oder gelöscht. Für diese Aufbewahrung sind in Deutschland die Bundes- und Landesarchive verantwortlich. Bei Geodaten gibt es jedoch häufig den Fall, dass Daten zwar nicht einigermaßen regelmäßig und von einem breiten Nutzerkreis verwendet werden, jedoch trotzdem (z.B. für Langzeitstudien oder Vergleiche) sicher aufbewahrt werden sollen. Zu diesem Zweck zunächst die Archivverwaltung kontaktieren zu müssen, ist derzeit kein praktikables Szenario. Sowohl der zeitliche als auch der Arbeitsaufwand wären bei der Häufigkeit der Anfragen für die Archivverwaltungen in Bund und den Ländern enorm, sofern keine automatisierte Bereitstellung existiert. Auch wäre die ständige Bereitstellung von Daten in der GDI keine sinnvolle Option, wenn Daten nur z.B. einmal in 3 Jahren benötigt werden. In dem skizzierten Fall ist daher die Einrichtung einer Zwischenstelle oder auch eines Übergangsarchivs notwendig. Diese Zwischenstelle übernimmt die Langzeitspeicherung. Dafür gibt es allerdings noch keine Empfehlungen, diese Lücke soll durch das vorliegende Konzept mit den bisherigen Erkenntnissen gefüllt werden.

4. Konzept für die Langzeitspeicherung

4.1 Struktur der Speicherung (Wo?)

Für die Verwaltung und Haltung der aktuellen Geodaten ist prinzipiell die datenhaltende Stelle selbst verantwortlich. Damit diese entscheiden kann „in welchem System welche Daten sinnvoll gespeichert werden“, muss Klarheit darüber bestehen wie sich die drei „Systeme“ grundsätzlich voneinander abgrenzen. Dies wurde in den vorhergehenden Abschnitten beschrieben.

In einer Übersicht zeigt die folgende **Abbildung 2** die möglichen Lebenszyklen eines Geodatenbestandes in den Systemen.

Dabei fallen unter das grün unterlegte System „aktuell verwendete Daten“ nicht nur die „Daten der GDI“, sondern auch die Originaldatenverwaltung (also Daten, die nicht über GDI-Schnittstellen zugänglich zu machen). Auch werden hier die Fälle zugeordnet, in denen historische Zeitscheiben von Daten zur ständigen Mehrfachnutzung über die GDI-Schnittstellen einem Nutzer zugänglich gemacht werden. Der Vollständigkeit halber sei an dieser Stelle erwähnt, dass es auch den Fall gibt, bei dem Daten in der „aktuellen Datenverwaltung“ nicht mehr von Interesse sind und dann gelöscht werden.

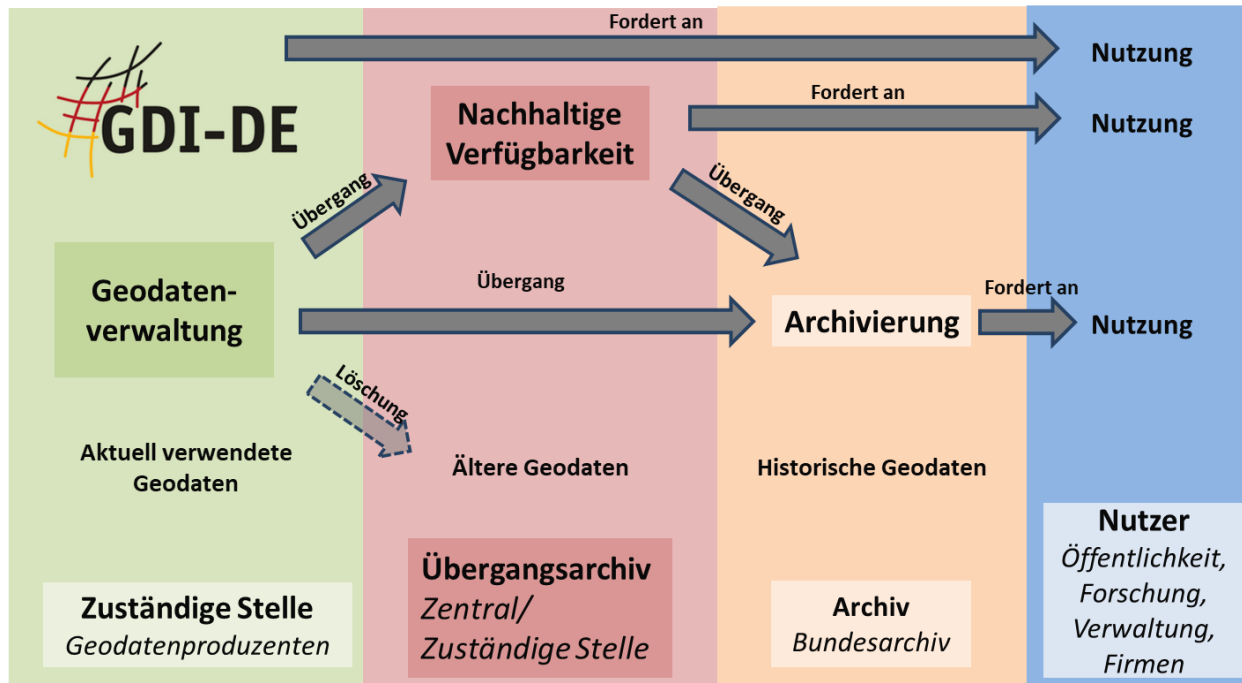


Abbildung 2: Geodatenverwaltung und Lebenszyklus

Für die Realisierung eines **Übergangsarchives** gibt es prinzipiell zwei verschiedene Möglichkeiten der Umsetzung. Zum einen könnten die Geodatenproduzenten (geodatenhaltende Stellen) unmittelbar verantwortlich für die Übergangsarchivierung sein und entsprechend ihre „älteren“ Daten selbst auf einem File-Server verwalten. Eine echte Nachhaltigkeit sicherzustellen ist dann jedoch schwierig. Eventuell ist aber nur eine „Low-Level-Variante“ gewollt. Die andere Möglichkeit wäre die Einrichtung einer zentralen Stelle entsprechend den Archiven, nur dass dort ausschließlich Geodaten geführt und gespeichert werden. Ein Vorteil wäre beispielsweise die einfach zu realisierende einheitliche Speicherung sowie Verwaltung. Dem gegenüber stehen der hohe zusätzliche Aufwand sowie die Kosten, für die eine zusätzliche Finanzierung notwendig wäre. Insgesamt bieten beide Varianten verschiedene Vor- und Nachteile die entsprechend abgewogen werden müssten, bevor es zu einer Entscheidung diesbezüglich kommt. Die „Low-Level-Variante“ wird in den weiteren Betrachtungen nicht mehr als LZS bezeichnet.

4.2 Umsetzungsanforderung an das System (Was?)

Als weiterer wichtiger Punkt neben der Zuständigkeits- und Strukturfrage sind die Fragen der zu unterstützenden **Schnittstellen bzw. Umsetzungsanforderungen** an die Langzeitspeicherung. Dazu gehören die Nachhaltigkeit, die Migration, die Sicherheit sowie die Zugänglichkeit der Daten. Bezüglich der Nachhaltigkeit und Migration sind Aspekte wie Formate oder die Interpretationsfähigkeit sowie eine nachhaltige Speicherform zu gewährleisten. Dafür sind einheitliche Standards zu definieren, in denen die Geodaten konvertiert werden müssen, bevor sie in die Übergangsarchivierung gehen. Dadurch wird gewährleistet, dass die Daten auch nach einigen Jahren genutzt werden können. Für den Sicherheitsaspekt sind zentrale Fragen wie die Originalität und die Zugriffsrechte auf die Daten zu beantworten. Auch sind Absicherungen vor

Verlusten (Brand, Stromausfall, etc.) zu treffen. Als letzten der vier Punkte ist die Zugänglichkeit zu realisieren.

Daneben gibt es **fachliche Anforderung** einer Langzeitdatenspeicherung, wie physikalische Sicherung, Bestandserhaltung, Herkunft, Revisionssicherheit oder Vollständigkeit (temporär oder geometrisch). Für die Suche und Interpretation der Langzeitspeicherung wird eine Beschreibung mit Metadaten sowie eine GDI-Einbindung empfohlen. Nachfolgend wird ein Versuch einer Gegenüberstellung der unterschiedlichen Anforderungen und Systeme (z.B. File-Server, Archiv und LZS) für eine nachhaltige Verfügbarkeit unternommen.

Anforderung	File-Server	LZS	Archiv
• Sicherung (physikalisch)	✓	✓	✓
• Bestandserhaltung (physikalisch)	✗	✓	✓
• Herkunft (Amtlich / Original)	?	✓	✓
• Vollständig (temporär, geometrisch)	✓	✓	?
• Revisionssicher	✗	✓	✓
• Formatkontrolle	✗	✓	✓
• Metadaten (Suche, Interpretation)	?	✓	✓
• GDI-Einbindung	?	✓	?
• Raumbezug	?	✓	?
• Regelbasierte Löschung	✗	✓	%

Abbildung 3: Gegenüberstellung der Anforderungen und möglicher Systeme für eine nachhaltige Verfügbarkeit [LZS Hamburg, 2019]

Aus der Abbildung 3 zeigt sich deutlich, dass die bisherige Praxis zur Speicherung bzw. Aufbewahrung auf File-Servern keine vollumfängliche Lösung für die Langzeitspeicherung bietet. Die Anforderungen zur Revisionsicherheit, Formatkontrolle oder Betrachtung und Recherchemöglichkeiten von Geodaten über Metadaten, kann eine Geodatenlangzeitspeicherung über File-Server nicht erfüllen. Die Langzeitspeicherung in den Archiven ist zwar revisionssicher und bestandserhaltend, allerdings fehlen oft Verwaltungssysteme für Geodaten und Geodatendienste, die eine nachhaltige Nutzung durch GDI-Einbindung (z.B. in einem WebGIS oder Geoportal) ermöglichen.

4.3 Zeit der Speicherung (Wann?)

Des Weiteren muss für jeden Datensatz festgelegt werden, **über welchen Zeitraum** die Geodaten im Übergangsarchiv gespeichert werden sollen und ab wann sie entsprechend ins Archiv abgegeben werden. Auch könnte dabei die Möglichkeit der direkten Archivierung eingerichtet werden – für Daten, für die keine absehbare Verwendung in naher Zukunft besteht. Dafür ist vorab von einem Gremium aus datenhaltender Stelle und dem zuständigen Archiv zu entscheiden, nach welchen Kriterien in vorzuhaltende und zu archivierende Daten unterschieden wird. Ein entscheidender Unterschied zwischen den zwei Datenspeicherorten wäre dann, dass archivierte Daten nur mit einer gewissen Vorlaufzeit abgefragt/angefordert werden könnten, während die vorzuhaltenden (langzeitgespeicherten) Daten möglichst direkt in die eigenen Nutzungsverfahren eingebunden bzw. insgesamt abgerufen werden könnten.

Im Grunde muss die **geodatenhaltende Stelle** zusammen mit dem Archiv die Aufbewahrungsfrist bestimmen. Wichtig ist allerdings, dass der Zeitraum der Aufbewahrung vor Übergang in das Systems der Langzeitspeicherung bestimmt und die Aufbewahrungsfrist (5, 10 oder 30 Jahre) definiert sein muss. Diese und weitere Angaben zu den Daten (siehe Kapitel 5) sind im Vorfeld der Datenabgabe im Rahmen der Bestandsaufnahme und Beschreibung zu erheben.

5. Digitale Geodaten – Bestandsaufnahme und Beschreibung

5.1 Bestandsaufnahme

In der Anlage 1 ist ein im Rahmen der Maßnahme erarbeiteter Erhebungsbogen zur Bestandsaufnahme und Beschreibung digitaler Geodaten enthalten.

Zu erfassen sind die Art der Geodaten (Raster, Vektordaten, etc), die Formate und Beschreibungen des Datensatzes, die Datenmenge, die Versionen sowie der Speicherbedarf. Weitere relevante Informationen zum Datenschutz, den Nutzungsbedingungen oder auch Festlegungen zur Aufbewahrungsfrist sind ebenfalls im Vorfeld der Datenabgabe zu erheben. Auf einzelne Punkte für die LZS wird im Folgenden näher eingegangen.

Abbildung 4: Erhebungsbogen zur Bestandsaufnahme und Beschreibung digitaler Geodaten (Anlage 1)



Anlage 1:

Erhebungsbogen

„Fortführung und Langzeitspeicherung von Geodaten“

Geodatenbestand (z. B. Verwaltungsgebiete 1:250 000, Digitale Orthophotos 20cm, etc.):

Kurzbezeichnung (z. B. VG250, DOP20, etc.): _____ Stand/Versionen des Datensatzes (z. B. 31.12, 2018, etc.): _____

Art der Geodaten:
 Rasterdaten (z. B. Pläne, Digitale Orthophotos, Satellitenbilder, etc.)
 Vektordaten (z. B. Verwaltungsgemeinschaften, Digitale Orthophotos, etc.)

Datenformat (z. B. JPG, TIFF, TIF+TFW): _____

Datenmenge insgesamt (z. B. 50 Datensätze): _____ Speicherbedarf insgesamt (z. B. 100 GB gepackt): _____

Zukünftiger Speicherbedarf je Fortführungszyklus (z. B. ca. 100 MB pro Jahr, etc.): _____

Vektordaten (z. B. Verwaltungsgebiete, Digitale Orthophotos, etc.)

Datenformat (z. B. DXF, SHP, NAS, GML, etc.): _____ Datenbank (z. B. Oracle, MS SQL Server, PostgreSQL, etc.): _____

Datenmenge insgesamt (z. B. 50 Datensätze): _____ Speicherbedarf insgesamt (z. B. 100 GB gepackt): _____

Zukünftiger Speicherbedarf je Fortführungszyklus (z. B. ca. 100 MB pro Jahr, etc.): _____

Dokumente / Dossiers und deren Datenformat
 (Dokumente, die die Interpretation des Datensatzes erleichtern, z. B. Spezifikationen, Datenmodellbeschreibungen, Produktbeschreibungen, Signaturkataloge, Texte, Tabellen, Sachattribute, Grafiken, etc.)

Dokument / Dossier (z. B. Kachelübersicht, etc.): _____ Datenformat (z. B. XLS, PDF, DOC): _____

Datenmenge insgesamt (z. B. 50 Dateien): _____ Speicherbedarf insgesamt (z. B. 530 KB): _____

5.2 Art der Geodaten

Bei der **Art der Geodaten** unterscheidet man grundsätzlich zwischen Rasterdaten, Vektordaten und beschreibenden Dokumenten. Um entscheiden zu können, wie man vorhandene Geodaten in der Langzeitspeicherung verfügbar macht, betrachtet man zunächst die bestehende Datenstruktur. Vektordaten werden in der Regel in Datenbanken, Rasterdaten und Dokumente im Dateisystem verwaltet.

5.3 Größe der Geodaten

Relevant für die Datenspeicherung sind auch die **Größe der Dateien** oder Datenbanken, da diese im System gespeichert und dauerhaft verfügbar gemacht werden soll. Dabei kann es auch zweckmäßig sein die Dateien vor der Langzeitspeicherung zu komprimieren, um eine bessere Ausnutzung vorhanden Speichers zu ermöglichen.

5.4 Datenmenge

Die **Datenmengenabschätzung** liegt in der Verantwortung der Datenverantwortlichen und ist eine essenzielle Vorarbeit der Speicherung in die LZS (vgl. Erhebungsbogen Anlage 1).

5.5 Datenformate

Bezüglich den **Datenformaten** hat die „Architektur der GDI-DE - Technik“ konkrete Vorgaben, in welchen Formaten digitale Geodaten und Metadaten beschrieben und über welche Dienste (Darstellungsdienste, Downloaddienste, etc.) diese Daten bereitzustellen sind [AK Architektur, 2019].

Sofern Geodaten in **proprietären Datenformaten** vorliegen, sind diese zunächst zu erheben (vgl. Erhebungsbogen Anlage 1) und im zweiten Schritt in ein Standardformat [AK Architektur, 2019] zu überführen. Dies sichert in der Langzeitspeicherung die nachhaltige Verfügbarkeit bereits zu Beginn der Übernahme. Zudem entfällt eine Migration zwischen unterschiedlichen Versionen proprietärer Datenformate und Softwareversionen über die Zeit.

5.6 Datenschutz

Die digitalen Geodaten können auch sensibel eingestufte Daten nach dem Bundesdatenschutzgesetz, wie z.B. Eigentümer- oder Adressangaben in ALKIS-Daten oder DOPs mit Auflösungen unterhalb der datenschutzrelevanten Grenze (länderspezifisch), von zzt. 20cm, enthalten. Entsprechende **Vorschriften zum Datenschutz** sind daher bei der Speicherung und Recherche der langzeitgespeicherten digitalen Geodaten zu berücksichtigen. Die Angabe Datenschutzrelevanter Inhalte in den Daten ist zu erheben (vgl. Erhebungsbogen Anlage 1).

6. Anwendungsszenarien

Anhand konkreter **Anwendungsszenarien** werden in diesem Abschnitt drei praktische Aspekte näher betrachtet und dabei konkrete Vorschläge für eine Realisierung gegeben:

1. Überlegungen zur Definition, wie und mit welchen Zeitabschnitten digitale Geodaten im System der Langzeitspeicherung eingebunden werden sollen (fiktives Beispiel der **Bereitstellung im Übergangsarchiv**)
2. Die **Registrierung und Verwaltung** digitaler Geodaten, die an verschiedenen Speicherorten vorgehalten werden können. Für Daten, die nicht oder nur als Duplikat an das Archiv abgegeben werden, ist eine unbegrenzte Speicherung so zu gewährleisten, dass auch nach Jahrzehnten die Daten noch gelesen und (eingeschränkt) verwendet werden können.
3. Die **Bestanderhaltung und Abgabe** der als archivwürdig eingestuft digitalen Geodaten an die digitale Datenhaltung des Staatsarchives. Die Übergabe der Daten erfordert eine besondere, standardisierte Form. Diese formatgerechte Aufbereitung der digitalen Geodaten ist dabei mit der Zusammenstellung in einer Containerstruktur verbunden. Die erforderlichen Datensatzbeschreibungen sind bekannt.

6.1 Bereitstellung im Übergangsarchiv

Vor der **Bereitstellung** im Übergangsarchiv ist für jeden digitalen Geodatenbestand vorab individuell zu definieren, wie und mit welchen Zeitabschnitten dieser im System der Langzeitspeicherung einzubinden ist. Am folgenden (fiktiven, aber realitätsnahen) Beispiel des Datenbestandes ALKIS lässt sich veranschaulichen, welche verschiedenen Anforderungen die Systeme aus Sicht eines Nutzers erfüllen müssen und welche Überlegungen ein Geodatenproduzent grundsätzlich anstellen sollte.

Fiktive Ausgangssituation: Der **Geodatenproduzent** als zuständige Stelle verfügt über einen umfangreichen Datenbestand "ALKIS-komplett". Dieser wird täglich aktualisiert. Eine spezielle Nutzung, z.B. für die "Liegenschaftsauskunft für Nutzer mit berechtigtem Interesse", erfolgt aus der Originaldatenbank und somit außerhalb der GDI-DE. Ein reduzierter Teil-Inhalt ("ALKIS-Behörden") wird einem eingeschränkten Nutzerkreis über die GDI-DE zur Verfügung gestellt. Ein weiterer reduzierter Teil-Inhalt ("ALKIS-öffentlich") wird als Open Data bereitgestellt. Automatisierte Prozesse stellen sicher, dass die Dienste zu "ALKIS-Behörden" und zu "ALKIS-öffentlich" täglich aus der "ALKIS-komplett"-Datenbank aktualisiert werden. So werden alle Anforderungen nach aktuellen Daten gut bedient.

Überlegungen bezüglich Übergangsarchiv: Für die Speicherung im Übergangsarchiv fallen Kosten an. Daher sollen keine redundanten Daten gespeichert werden, was aber bei einer 1:1 Übernahme der drei o.g. Produkte der Fall wäre. Also soll 2x/ Jahr "ALKIS-komplett" in das Übergangsarchiv eingelagert werden. So ist man für Anfragen mit verschiedenem Fokus vorbereitet und dünnt ggf. vor Abgabe an den Nutzer wieder aus.

Überlegung zur Archivierung: Angenommen, es besteht die Verpflichtung, dass ALKIS-Daten nach der Aufbewahrungsfrist alle 5 Jahre ins Staatsarchiv abzugeben sind. Dann wird dieser Schritt der **Datenabgabe an das Archiv** bei jeder 10. Einlagerung von ALKIS ins Übergangsarchiv vermerkt und vorbereitet. Die Erfahrung hat gezeigt, dass immer mal wieder Nutzer für eine Recherche historische ALKIS-Daten nachfragen, allerdings mit unterschiedlichen Fragestellungen. Zur Historisierung seiner Daten (2 Zeitscheiben/ Jahr haben sich als sinnvoll erwiesen) möchte der Geodatenproduzent das Übergangsarchiv nutzen. So kann er gleichzeitig sicher sein, dass eine dauerhafte Rückkonvertierung systemimmanent sichergestellt ist und der Zugriff für einen Nutzer relativ einfach möglich ist.

6.2 Registrierung und Verwaltung

Bereits bei der Anbindung von Geodaten an die Geodateninfrastruktur Deutschland ist es notwendig, einen Registrierungsprozess in der GDI-DE Registry durchzuführen. Dies beinhaltet die Registrierung der Behörde in einem Organisation-Register in der GDI-DE Registry, als auch die **Registrierung eines Namensraumes** für jeden aktuell bereitgestellten Geodatensatz. Für jeden dieser Datensätze existiert zudem ein passender Metadatensatz der u.a. Informationen zu Lizenz- und Nutzungsbedingungen der Daten enthält.

Wie bereits bei der Struktur der Speicherung in Kapitel 4.1 beschrieben, müssen für aktuelle Geodaten die datenhaltenden Stellen an die GDI-DE angebunden werden. Auch muss eine Möglichkeit, am besten standardisiert, bestimmt werden, über welche Schnittstellen die Geodaten zugänglich gemacht werden und wie bestimmte Geodaten gefunden werden können. Dafür wäre die **Einrichtung eines Verzeichnisregisters (Registry)** von Vorteil. Das Prinzip ist dabei genau wie bei der Registry der GDI-DE und dient der Dokumentation der Daten.



Abbildung 5: Namensräume in der GDI-DE Registry (Vereinfachte Darstellung)

Die **Metadaten der Geodaten** müssen ebenfalls langzeitgespeichert sowie abrufbar sein. Daher muss auch die Zugangsmöglichkeit zu den Metadaten geschaffen werden. Das OAIS-Modell zur standardisierten Archivierung bietet dabei eine gute Orientierung für eine mögliche Umsetzung eines Modells zur Speicherung und Verwaltung der Daten.

6.3 Bestandserhaltung und Abgabe

Des Weiteren ist nicht auszuschließen, dass im Laufe der Zeit Daten bereitgestellt werden, die **bereits langzeitgespeicherten Dateien** sehr ähnlich sind. Die Ähnlichkeit bei der Fortführung in der Langzeitspeicherung lässt sich aber u.U. nicht direkt aus dem Dateinamen ableiten. Damit im System der LZS dennoch Instrumente bereitgestellt werden können, nach solchen ähnlichen Daten zu suchen, um eine Entscheidung zugunsten der aufbewahrungswürdigen Daten zu treffen, sind sehr flexible Suchmöglichkeiten zu implementieren. Diese müssen eine kombinierte Suche über:

- ähnliche Dateinamen
- vergleichbare Dateigrößen
- letztes Änderungsdatum bzw. Datum der LZS-Anlieferung
- die Formate
- die LZS-Metadaten
- speziellen Identifikationsdaten einer Datei

zulassen. Wenn das Suchergebnis die Entscheidung unterstützt, dass es sich um bereits vorhandene Daten handelt, werden neu angelieferte Daten mit dem Hinweis auf schon gespeicherte Daten gar nicht mehr in die LZS übernommen. So lässt sich auch der gesamte LZS-Datenbestand in großen Abständen auf ähnliche Inhalte überprüft und entbehrliche Daten zur Löschung vorgeschlagen.

Der **Datenumfang der Abgabe** wird vom Betreiber der LZS, den geodatenhaltenden Stellen und den Vorschriften der Archiv-Gesetze („archivwürdige Geodaten“) bestimmt und kann in zwei Hauptgruppen aufgeteilt werden:

Es gibt die als **archivwürdig eingestuften digitalen Geodaten**. Danach ist der Datenbestand nach vereinbarten Aussonderungszeitplänen grundsätzlich bis zur Abgabe an das Archiv beim Datenproduzenten oder in einem Übergangsarchiv zu speichern, das können mehrere Jahre sein. Nach der Abgabe an das Archiv kann der Datenbestand in der LZS gelöscht werden, muss es aber nicht. Diese Entscheidung liegt beim Datenproduzenten und ist wesentlich von der Kosten-Nutzen-Frage und der Verwendungsmöglichkeit aus dem Digitalen Archiv heraus abhängig.

Darüber hinaus gibt es die als **nicht-archivwürdig eingestuften digitalen Geodaten**, deren Speicherung aber im direkten Interesse des Datenproduzenten liegen kann, um ggf. auch in späteren Jahren noch Auswertungen, Differenzanalysen oder Statistiken darauf aufzubauen.

Darüber hinaus könnten weitere vorbereitende Arbeiten notwendig sein, z.B. die Übernahme der **Metadaten** und Dossiers oder die **Pflege von Fachattributen**, wie der Jahrgang oder die Kachelnummer, zur besseren Verwaltung und Nutzung der Daten in der Langzeitspeicherung.

7. Fazit

In der **Geodateninfrastruktur Deutschland (GDI-DE)** werden Daten zu verschiedenen Geodatenbeständen über standardisierte Schnittstellen in Form von Webdiensten zugänglich gemacht. Beschrieben sind die Daten und Dienste im dazugehörigen Metadatenkatalog. Auf diesem Wege werden diejenigen Daten bereitgestellt, die einem relativ breiten Nutzerkreis ständig angeboten werden sollen. Dies sind überwiegend aktuelle Daten, aber zunehmend auch Daten als Zeitreihen.

Das **Langzeitarchiv** sichert die dauerhafte Aufbewahrung und Erhaltung von Daten und Unterlagen nach Ablauf der Aufbewahrungsfrist zur Bewahrung historischer Überlieferungen. Die **Langzeitspeicherung** ist ein Zwischenschritt: Dieses **Übergangsarchiv** bietet bei Bedarf die Möglichkeit, Daten nachhaltig vorzuhalten, die nicht (mehr) ständig bereitgestellt werden sollen, die aber auch nicht gelöscht werden sollen und noch nicht im Langzeitarchiv archiviert werden.

Aus **technischer und organisatorischer Sicht**, ist ein gemeinsames Auffinden von Daten zwischen der LZS und der GDI zukünftig sicherzustellen. Ebenso sollen die Schnittstellen der Metadatenkataloge genutzt werden, um soweit möglich die Metadatenbeschreibung bzw. die Dossiers in der Langzeitspeicherung zu befüllen und dauerhaft zugänglich zu machen.

Für die **zukünftige Gestaltung der Langzeitspeicherung** lässt sich festhalten, dass eine Vielzahl an unterschiedlichen Anforderungen zu berücksichtigen sind. Zum einen die technische Anforderung an die Art der abzuliefernden Geodaten um eine Standardisierung zu schaffen. Zum anderen muss eine Auffindbarkeit der Daten über Schnittstellen zur GDI-DE geschaffen werden sowie eine Dokumentation und Registrierung der Daten, um sie in der Langzeitspeicherung wieder zu finden. Auch müssen die zeitliche Befristung und Zuständigkeit, also wie lange und wo die Daten zu führen sind, definiert werden. Prinzipiell sind die einzelnen Komponenten (Technik, Infrastruktur und Rechtsgrundlage) gegeben und müssen nur entsprechend angepasst bzw. umgesetzt werden.

Zukünftig sollten datenhaltende Stellen vermehrt vorsehen, **zu ausgewählten Datenbeständen** je einen Dienst in der GDI zu erstellen, der historische Datensätze zu definierten Zeitabständen zugänglich macht (historischer Dienst). Ein solcher Dienst kann inhaltlich auch weiter als 10 Jahre zurückreichen. Die in der GDI über „historische Dienste“ zugänglichen Datenthemen sollten dabei aber nicht redundant in der LZS gespeichert werden.

Darüber hinaus sollten **weitere Anwendungsszenarien** der Langzeitspeicherung entwickelt werden, die eine verbesserte Einbindung und Nutzung beispielsweise für Daten der Wissenschaft ermöglicht. Hier kann zukünftig die Beteiligung der GDI-DE in der Nationalen

Forschungsdateninfrastruktur für die Erdsystemwissenschaften [NFDI4Earth, 2021] einen wichtigen Beitrag leisten.

8. Quellen

[AK Architektur, 2019] Architektur der GDI-DE – Technik, Version 3.4.1, AK Architektur, 2019 abgerufen am 01.10.2021, <https://www.gdi-de.org/GDI-DE/Arbeitskreise/Architektur>

[BMI, 2014] Verwaltung Innovativ, Organisationskonzept elektronische Verwaltungsarbeit, Baustein E-Langzeitspeicherung, BMI, 2014, abgerufen am 01.10.2021, https://www.verwaltung-innovativ.de/SharedDocs/Publikationen/Organisation/e_langzeitspeicherung.html

[Bundesarchiv, 2015] Leitlinien zur bundesweit einheitlichen Archivierung von Geobasisdaten, KLA – Konferenz der Konferenz der Leiterinnen und Leiter der Archivverwaltungen des Bundes und der Länder, AdV - Arbeitsgemeinschaft der Vermessungsverwaltung der Länder, 2015, abgerufen am 01.10.2021, <https://www.bundesarchiv.de/DE/Content/Downloads/KLA/leitlinien-geobasisdaten.html>

[LZS Hamburg, 2019] Langzeitspeicherung digitaler Geodaten, Hamburg, 2019, abgerufen am 01.10.2021, <https://www.hamburg.de/langzeitspeicherung-geodaten/12399872/lzs-seminar/>

[LZS Hamburg, 2021] Konzept Langzeitspeicherung digitaler Geodaten, Hamburg, 2021, abgerufen am 01.10.2021 <https://www.hamburg.de/langzeitspeicherung-geodaten/13106180/lzs-entwicklung/>

[NFDI4Earth, 2021] Nationale Forschungsdateninfrastruktur für Erdsystemwissenschaften, 2021, abgerufen am 01.10.2021, <https://www.nfdi4earth.de/>

Anlage 1:

Erhebungsbogen

„Fortführung und Langzeitspeicherung von Geodaten“

Geodatenbestand (z. B. Verwaltungsgebiete 1:250 000, Digitale Orthophotos 20cm, etc.):

--

Kurzbezeichnung (z. B. VG250, DOP20, etc.):

Stand/Versionen des Datensatzes (z. B. 31.12, 2018, etc.):

--	--

Art der Geodaten:

Rasterdaten (z.B. Pläne, Digitale Orthophotos, Satellitenbilder, etc.)

Datenformat (z. B. JPG, TIF, TIF+TFW):

--

Datenmenge insgesamt (z. B. 50 Datensätze):

Speicherbedarf insgesamt (z.B. 100 GB gepackt):

--	--

Zukünftiger Speicherbedarf je Fortführungszyklus (z.B. ca. 100 MB pro Jahr, etc.)

--

Vektordaten (z. B. Verwaltungsgebiete, Digitale Orthophotos, etc.)

Datenformat (z. B. DXF, SHP, NAS, GML etc.):

Datenbank (z. B. Oracle, MS SQL Server, PostgreSQL, etc.):

--	--

Datenmenge insgesamt (z. B. 50 Datensätze):

Speicherbedarf insgesamt (z.B. 100 GB gepackt):

--	--

Zukünftiger Speicherbedarf je Fortführungszyklus (z.B. ca. 100 MB pro Jahr, etc.):

--

Dokumente / Dossiers und deren Datenformat

(Dokumente, die die Interpretation des Datensatzes erleichtern, z.B. Spezifikationen, Datenmodellbeschreibungen, Produktbeschreibungen, Signaturenkataloge, Texte, Tabellen, Sachattribute, Grafiken, etc.)

Dokument / Dossier (z. B. Kachelübersicht, etc.):

Datenformat (z. B. XLS, PDF, DOC):

Datenmenge insgesamt (z. B. 50 Dateien):

Speicherbedarf insgesamt (z. B. 530 KB):

--	--

Geben Sie eine kurze inhaltliche Beschreibung des Datensatzes (z. B. Daten zu Verwaltungsgebieten der Bundesrepublik Deutschland von 1997 bis 2019, etc.):

Wie ist der Raumbezug / die Georeferenzierung gegeben (z. B. Koordinatenreferenzsystem ETRS89/UTM32 etc.).

Eingesetzte Software mit Versionsnummer zur Erfassung und/oder Verwaltung dieser Geodaten (z. B. ESRI ArcGIS/ArcInfo, QGIS, AutoCAD, etc.).

Erfassen oder pflegen Sie Metadaten zu den Geodaten?
(z. B. Erstellungsdatum, Änderungsdatum, Ansprechpartner, Nutzungsrechte, Copyright, etc.):

Eingesetzte Software mit Versionsnummer zur Erfassung und/oder Verwaltung der Metadaten (z. B. MS-Word, MS-Excel, GeoNetwork, TerraCatalog, etc.).

Gibt es eine gesetzliche Grundlage?

Ja
 Nein
 Teilweise

Falls ja, welche? (z. B. BGeoRG, GeoZG)

Gibt es eine INSPIRE Relevanz?

Ja
 Nein
 Teilweise

Falls ja, zu welchem Thema? (z. B. Bodennutzung)

Wer ist Datenproduzent/ Datenurheber? (z. B. „Bundesamt für Kartographie und Geodäsie (BKG)“)

Wie ist die räumliche Ausdehnung (EX_GeographicExtent) des Datensatzes? (z. B. Bundesrepublik Deutschland, Land Hessen, etc.)

Wie ist die zeitliche Ausdehnung (EX_TemporalExtent) des Datensatzes? (z. B. jährliche/halbjährliche Fortführung, Tagesaktuell, etc.)

Seit wann werden die Geodaten erfasst/ Zeitraum der Erfassung? (z. B. seit 01.01.1998, 2010 bis 2020, etc.)

Anzahl an bisherigen Fortführungszyklen
(z. B. 4 Fortführungen, etc.):

Auflistung der Stände/Versionen
(z. B. Jahrgang 2010, 2014, 2016, 2018, etc.):

Durch welche Fachattribute wird der Datensatz üblicherweise beschrieben?

Fachattribut (z. B. Jahrgang, Kachelnummer, etc.)

kurze Beschreibung (z. B. Jahr der Befliegung)

Fachattribut (z. B. Jahrgang, Kachelnummer, etc.)	kurze Beschreibung (z. B. Jahr der Befliegung)

Unterliegen die Daten dem Datenschutz?

Ja Nein Teilweise

Falls ja, welche? (z. B. Eigentümerangaben, etc.):

Gibt es Zugangs- oder Nutzungsbedingungen?

Ja Nein Teilweise

Falls ja, welche? (z. B. Lizenzen, Gebühren, etc.)

Gibt es festgelegte Aufbewahrungsfristen?

Ja Nein Teilweise

Falls ja, wie lange? (z. B. 10 Jahre, max. 30 Jahre, etc.)

Falls nein, wie lange werden die nicht-mehr-aktuellen Geodaten nachgefragt? (z. B. mind. 5 Jahre)

Gibt es eine Entscheidung zur Archivwürdigkeit?

Ja Nein Teilweise

Falls ja, welche? (z. B. Archivieren, Vernichten, etc.)

Falls teilweise, welche Aspekte? (z. B. Form und Farbe der Signatur sind in graphischer Form nicht archivwürdig.)

Sonstige Anmerkungen (z. B. Angaben zur Archivierung, Vollständigkeit, etc.):