

LZA UIS

Langzeitarchivierung von Umweltinformationen

*C. Keitel; R. Lang
Landesarchiv Baden-Württemberg, Abt. 2 und 1
Eugenstr. 7
70182 Stuttgart*

*T. Fritz
Landesarchiv Baden-Württemberg, Abt. Hauptstaatsarchiv Stuttgart
Konrad-Adenauer-Str. 4
70173 Stuttgart*

*A. Keitel; M. Müller
LUBW Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg
Griesbachstr. 1
76185 Karlsruhe*

*K. Naumann
Landesarchiv Baden-Württemberg, Abt. Staatsarchiv Ludwigsburg
Arsenalplatz 3
71638 Ludwigsburg*

*E. Schelkle; K.-P. Schulz; W. Uhrig
Ministerium für Umwelt, Naturschutz und Verkehr Baden-Württemberg
Kernerplatz 9
70182 Stuttgart*

*J. Treffeisen
Landesarchiv Baden-Württemberg, Abt. Generallandesarchiv Karlsruhe
Nördliche Hildapromenade 2
76133 Karlsruhe*

*F.-J. Ziwes
Landesarchiv Baden-Württemberg, Abt. Staatsarchiv Sigmaringen
Karlstr. 1+3
72488 Sigmaringen*

1. VORGESCHICHTE	87
2. PROJEKTPHASEN.....	87
3. ARCHIVIERUNGSSTRATEGIE.....	87
4. DATEIFORMATE	88
5. GRUNDSÄTZLICHE FRAGEN DER ÜBERNAHME.....	89
6. ÜBERNAHMEVERFAHREN	89
7. ÜBERNAHMEZEITPUNKTE	90
8. VALIDIERUNG.....	91
9. ZUSAMMENFASSUNG.....	92
10. LITERATUR.....	92

1. Vorgeschichte

Nach dem 1987 erstmals verabschiedeten Landesarchivgesetz Baden-Württemberg hat die Landesverwaltung dem Landesarchiv (vor 2005: der Archivverwaltung) sowohl konventionelle Unterlagen auf Papier als auch maschinenlesbare Unterlagen anzubieten /1/. Mitte der 1990er-Jahre fanden erste sporadische Gespräche zwischen Archiv- und Umweltverwaltung statt. Seit 1997 wurde in den baden-württembergischen Staatsarchiven über die archivische Bewertung von Umweltinformationssystemen nachgedacht /2/. 2002 konnte ein Modell zur Bewertung analoger Daten der Umweltverwaltung vorgelegt werden, das die Bewertung digitaler Daten der Umweltverwaltung in Aussicht stellte /3/. 2006 entwickelte das Landesarchiv das Digitale Magazin DIMAG und war nun auch organisatorisch und technisch in der Lage, digitale Umweltdaten in größerem Umfang zu übernehmen /4/ (vgl. Abbildung 1). Zugleich gab es keine veröffentlichten konkreten Angaben, wie solche Daten archiviert werden könnten. An diesem Zustand hat sich bis heute wenig geändert. Ende 2006 vereinbarte daher das damalige Umweltministerium mit dem Landesarchiv die Gründung einer gemeinsamen Arbeitsgruppe zur „Langzeitarchivierung von UIS-Daten“. Deren Ziel ist die Einrichtung dauerhafter Verfahren zur Übernahme archivwürdiger Umweltdaten aus dem Umweltinformationssystem Baden-Württemberg (UIS BW).

2. Projektphasen

Das Projekt ist in Entwicklungsphase, Aufbauphase und Regelbetrieb unterteilt. In der Entwicklungsphase wurden erste Pilotobjektarten übernommen, grundsätzliche Debatten geführt und eine Klassifikation der Objekte und Aufgaben vorgenommen. In der Aufbauphase werden die Erkenntnisse dieser Muster auf alle anderen Objektarten des UIS BW übertragen. Es erfolgt also eine archivische Bewertung und ggf. eine erste Übernahme dieser Objekte durch das Landesarchiv. Im Regelbetrieb werden die archivwürdig bewerteten Objektarten nach den entwickelten Vorgaben regelmäßig übernommen. Inhalte, Formate und Termine sind bekannt und stehen fest.

3. Archivierungsstrategie

Die Arbeitsgruppe diskutierte mit der Emulations- und der Migrationsstrategie die beiden derzeit relevanten Wege zur langfristigen Erhaltung digital gespeicherter Informationen /5/. Relativ rasch einigte sich die Arbeitsgruppe darauf, die Daten aus dem UIS BW wie alle anderen digitalen Archivalien des Landesarchivs mit der Migrationsstrategie zu archivieren. Die Informationen werden also übernommen, nicht erhalten werden die Hard- und die Software des Systems. Viel mehr Zeit benötigte die Diskussion der mit diesem Datenexport einhergehenden Informationsverluste:

- Die heutige Präsentationsebene geht verloren – einen gewissen Ersatz könnte hier die Archivierung einiger Screenshots oder Karten darstellen, die zusätzlich z.B. Symbole, Einfärbungen und erklärende Legenden enthalten.

- Implizites Wissen der Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg (LUBW) muss beschrieben werden, z.B. für die Interpretation notwendige zusätzliche Angaben. Schließlich können die heutigen Mitarbeiterinnen und Mitarbeiter der LUBW in fernerer Zukunft nicht mehr nach den Details befragt werden, die bereits heute für eine Interpretation der Daten nötig sind.
- Explizite Verknüpfungen können kaum erhalten werden, z.B. Links in HTML.
- Bei Migrationen in neue Dateiformate oder Koordinatensysteme sind die erhaltenswerten Eigenschaften sowie die hinnehmbaren Verluste zu bestimmen.

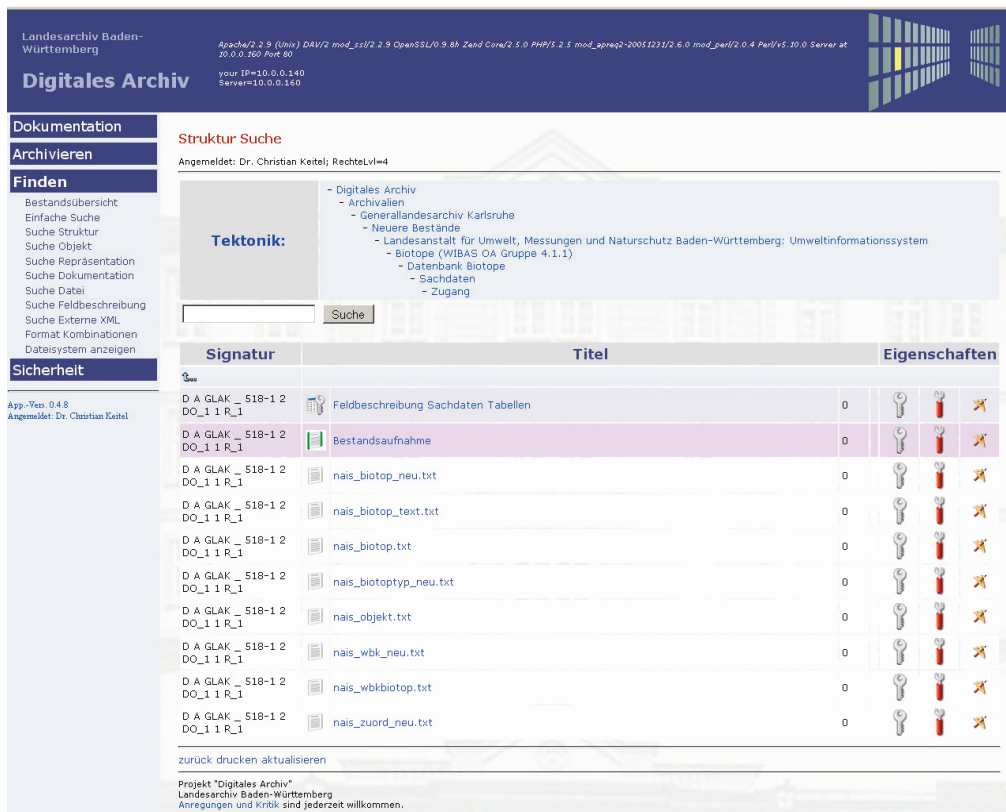


Abbildung 1: Archivobjekt „Biotopobjekt“ im Digitalen Magazin DIMAG

4. Dateiformate

Metadaten werden in den Dateiformaten XML oder CSW archiviert, Sachdaten in CSV oder XML. Länger diskutiert wurde über ein für die kartographische Darstellung geeignetes Format. Dabei konnte keines der diskutierten Formate vollständig überzeugen:

- TIFF ist zwar einfach; allerdings löst es die klare Zuordnung der Sachdaten zu definierten kartographischen Daten auf.
- ESRI/Shape ist zwar weit verbreitet, aber proprietär und damit für die dauerhafte Erhaltung nur wenig geeignet.
- Dagegen ist GML normiert, aber kaum verbreitet, was ebenfalls gegen eine Verwendung zur langfristigen Archivierung spricht.

Die Arbeitsgruppe entwickelte schließlich eine mehrstufige Lösung. Zunächst werden die Objekte im Format ESRI/Shape übernommen, manche von ihnen teilweise auch zusätzlich in

GML. 2011/12 soll die Alternative ESRI/Shape und GML erneut evaluiert werden. Die LUBW hat angeboten, im Falle einer Migration von ESRI/Shape zu GML das Landesarchiv zu unterstützen.

5. Grundsätzliche Fragen der Übernahme

Die Übernahme der Pilotobjektarten machte grundsätzliche Probleme bei der Übernahme von Daten aus georeferenzierten Fachverfahren deutlich. Manche Objektarten sind inhaltlich abgeschlossen (z.B. Biotopkartierung 2004), bei ihnen genügt eine einmalige Übernahme aller Daten. Andere Objektarten werden regelmäßig mit neuen Informationen angereichert. Bei ihnen liegt eine Übernahme in regelmäßigen Abständen (Zeitschnitten) nahe. Dabei könnten nach einer ersten vollständigen Übernahme bei den weiteren Zeitschnitten nur noch die veränderten Daten übernommen werden (Inkrementelle Übernahme). Die Daten der zweiten Übernahme (und aller weiteren Übernahmen) könnten dann nur sinnvoll zusammen mit den Daten der ersten Übernahme genutzt werden. Obwohl bei dieser Methode das Datenvolumen verringert und eine gemeinsame Ansicht mehrerer Zeitschnitte erleichtert wird, wäre diese Methode mit erheblichen Unsicherheiten verbunden: Wurden wirklich alle veränderten Daten übernommen? Können diese später mit den Daten der ersten Übernahme stimmig zusammengesetzt werden /6/? Könnten in diesem Fall alle Daten konkreten Zeitpunkten in der Vergangenheit (d.h. den Übernahmetagen) zugeordnet werden? Die mit diesem Vorgehen verbundenen Unsicherheiten erschienen der Arbeitsgruppe zu groß. Sie präferiert zum jetzigen Zeitpunkt, auch bei Zeitschnitten jeweils die ganze Objektart zu übernehmen.

Zeitschnitte können zyklisch oder anlassbezogen erfolgen. Die nähere Festlegung darüber sowie über die Zeitdauer zwischen zwei Schnitten muss auf die jeweilige Objektart bezogen werden.

6. Übernahmeverfahren

Vom Landesarchiv wurden zunächst verschiedene Objektarten in Pilotprojekten übernommen (u.a. Altlastenkataster und Biotopkartierung, Wasserschutzgebiete, Brunnen und Quellen; vgl. Abbildung 2).

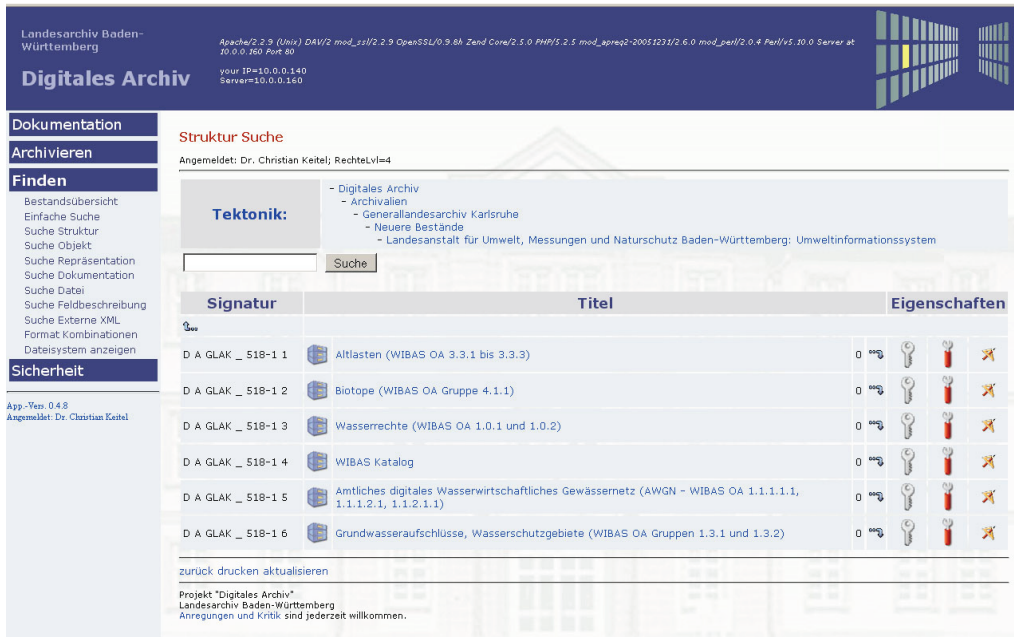


Abbildung 2: Übernommene Objektarten der Pilotphase

Zwar wurden dabei die wesentlichen konzeptionellen Probleme klar. Allerdings kann dieser Aufwand auf Dauer nicht betrieben werden, denn die Pilotobjektarten wurden jeweils über speziell angefertigte Abfragen exportiert. Die Arbeitsgruppe diskutierte daher intensiv über Möglichkeiten, ein dauerhaftes und möglichst ressourcenschonendes Export- und Übernahmeverfahren zu entwickeln. Dabei sollte die Übernahme auf jeden Fall in den Produktionsablauf der LUBW eingebunden werden. Wichtig ist die Beteiligung sowohl der fachlichen als auch der technischen Verantwortlichen der Umweltverwaltung, die häufig in unterschiedlichen Einrichtungen arbeiten. Dass das Landesarchiv bei einer Übernahme beteiligt ist, versteht sich von selbst.

7. Übernahmezeitpunkte

Der optimale Übernahmezeitpunkt ist von verschiedenen Determinanten abhängig:

- Anstehende Löschungen: Die Daten müssen vor einer aus Datenschutzgründen vorzunehmenden Löschung angeboten werden //.
- Bezug zu anderen Objektarten: Manche Objektarten können nur dann von künftigen Archivbenutzern gemeinsam ausgewertet werden, wenn sie auch zum selben Zeitpunkt übernommen wurden.
- Kapazitäten der fachlich und technisch Verantwortlichen: Eine Übernahme ist davon abhängig, dass entsprechende Ressourcen zur Verfügung stehen.
- Vollständigkeit der Daten: Manche Objektarten befinden sich erst im Aufbau. Hier kann es sinnvoll sein, zunächst von einer Übernahme abzusehen.
- Anstehende Veränderungen der Datenqualität: Die Informationen in manchen Objektarten werden regelmäßig auf ihre Stimmigkeit hin überprüft. Es empfiehlt sich, die Übernahme erst im Anschluss an eine derartige Prüfung vorzunehmen.
- Anstehende Veränderungen der Dateninhalte: Bei der Beschreibung des Datenumfangs oder der Dateninhalte ergeben sich über die „Lebenszeit“ einer Objektart teil-

weise Änderungen. Z.B. können Schlüssellisten im Laufe einer langjährigen Kartierung ergänzt werden. Es empfiehlt sich, die Auswirkungen inhaltlicher Veränderungen beim Festlegen eines Übernahmezeitpunktes zu berücksichtigen. Veränderungen sollten in den Metadaten vermerkt werden, um bei späteren statistischen Auswertungen verfahrensbedingte Sprungeffekte vermeiden zu können.

- Anstehende Veränderungen der Metadaten: Nach Möglichkeit sollten Metadaten normiert übernommen werden. Beispielsweise führen die von der EU-Richtlinie INSPIRE (Infrastructure for Spatial Information in the European Community) ausgehenden Vorhaben zu einer weiteren Normierung der Metadaten in Geographischen Informationssystemen. Zugleich wird die Zahl übernehmbarer Metadaten teilweise erweitert, was das Verständnis der beschriebenen Objektart durch künftige Archivbenutzer erleichtern dürfte.
- Anstehende Veränderungen des Koordinatennetzes: Um die verschiedenen archivierte Objektarten gemeinsam verwenden oder auswerten zu können, sollten sie auf dasselbe Koordinatennetz verweisen. Unmittelbar vor einer Änderung des Bezugssystems ist daher eine Übernahme von Daten nicht empfehlenswert.

8. Validierung

Nach der Bewertung sollten die Daten der archivwürdigen Objektarten ins Archiv übernommen werden, d.h., sie werden gemäß der Migrationsstrategie aus dem Umweltinformationssystem exportiert. Im Archiv muss geprüft werden, ob die Daten die erwarteten Inhalte besitzen. Ein erster Schritt ist die automatisierte Validierung. Wie können die exportierten Daten mit den Ausgangsdaten verglichen werden? Für diesen Zweck hat das Landesarchiv das Tool IngestList entwickelt (vgl. Abbildung 3).

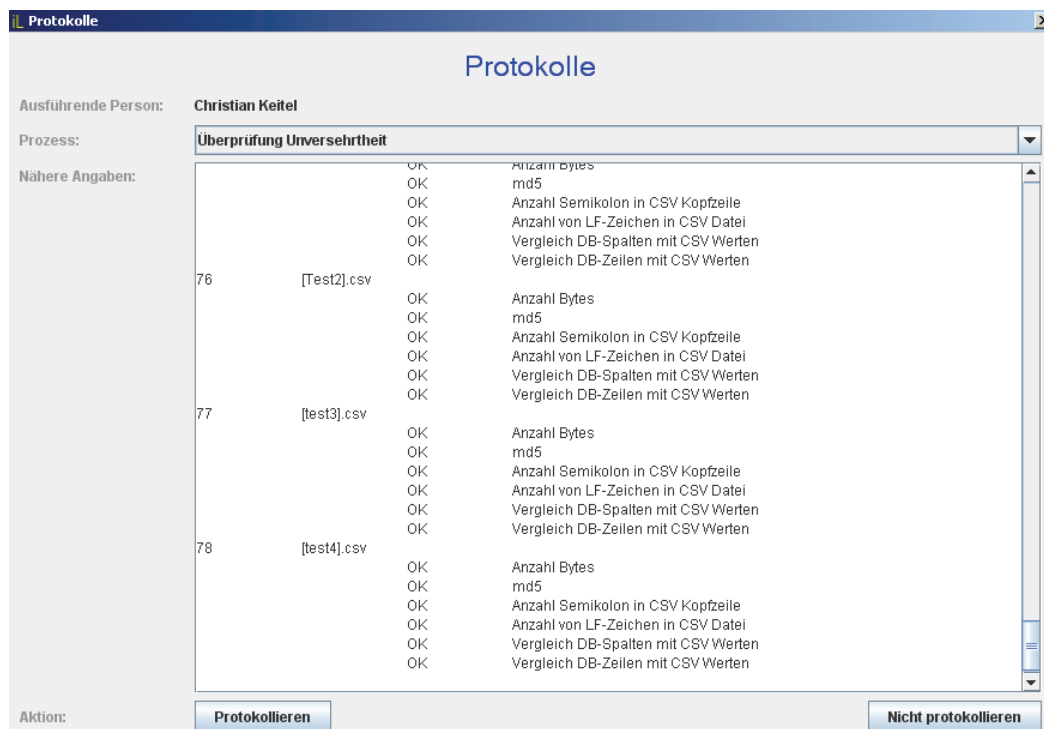


Abbildung 3: IngestList

Weitere Validierungsverfahren werden von der Arbeitsgruppe diskutiert. Inhaltlich können nur die Fachverantwortlichen die Qualität der exportierten Daten beurteilen. Trotz den noch auszubauenden Möglichkeiten einer voll- oder halbautomatischen Validierung wird die inhaltliche Validierung durch die Fachverantwortlichen auch in Zukunft notwendig sein.

9. Zusammenfassung

Die Arbeitsgruppe hat eine gemeinsame Sprache und ein gemeinsames Problemverständnis entwickelt, den Handlungsrahmen ausgeleuchtet, praktische Erfahrungen aufgebaut und fachliche und technische Lösungen erarbeitet.

Einige archivwürdige Geodaten konnten bereits gesichert werden. Übernommen wurden unter anderem das Altlastenkataster, das Biotopkataster, das Wasserrechtsbuch, das Amtliche Digitale Gewässernetz, die Wasserschutzgebiete, die Quellenschutzgebiete und die Grundwasseraufschlüsse. Nach Ablauf der archivgesetzlichen Sperrfristen (allgemeine Sperrfrist von 30 Jahren, personenbezogene Unterlagen sind bis 10 Jahre nach dem Tod der Person gesperrt) stehen die Daten zur Benutzung bereit.

10. Literatur

- /1/ Landesarchivgesetz § 2 Abs. 2.
- /2/ Erste Ergebnisse in Ziwes, F.-J. (1999): Überlegungen zur Bewertung von digitalen Unterlagen aus dem Umweltinformationssystem Baden-Württemberg. In: Schäfer, U., Bickhoff, N.; Hrsg.: Archivierung elektronischer Unterlagen. Verlag W. Kohlhammer, Stuttgart.
- /3/ Vertikale und horizontale Bewertung der Unterlagen der Umweltverwaltung in Baden-Württemberg, http://www.landesarchiv-bw.de/sixcms/media.php/120/44703/Bewertungsmodell_umweltverwaltung.pdf, Abruf 2.6.2010.
- /4/ Keitel, C. et al. (2007): Konzeption und Aufbau eines Digitalen Archivs: von der Skizze zum Prototypen. In: Ernst, K.; Hrsg.: Erfahrungen mit der Übernahme digitaler Daten. Elfte Tagung des AK „Archivierung von Unterlagen aus digitalen Systemen“ vom 20./21. März 2007, Hohenheim Verlag, Stuttgart/Leipzig.
- /5/ Zu den unterschiedlichen Strategieansätzen vgl. Neuroth, H. et al. (2009): nestor Handbuch. Eine kleine Enzyklopädie der Langzeitarchivierung. Version 2 vom Juni 2009, Verlag Werner Hülsbusch, Boizenburg.
- /6/ Dies erklären Buneman, P. et al. (2004): Archiving Scientific Data. ACM Transactions on Database Systems (TODS), 29, 1, p. 2-42, ACM, New York, USA, <http://portal.acm.org/citation.cfm?doid=974750.974752>, Abruf 15.6.2010
- /7/ Landesdatenschutzgesetz § 23 Abs. 3.