

Fachbeitrag

André Hartmann, Hendrik Herold, Anna Lisa Schwartz und Markus Schmalzl

FAIR Geodaten aus dem Archiv. Ein Tool zur automatisierten Metadatenextraktion aus historischen Karten

FAIR Geodata out of the Archive. A Tool for Automated Meta Data Extraction from Historical Maps

<https://doi.org/10.1515/abitech-2023-0045>

Zusammenfassung: In einem von der NFDI4Biodiversity geförderten Kooperationsprojekt der Generaldirektion der Staatlichen Archive Bayerns (GDA) und des Leibniz-Instituts für ökologische Raumentwicklung (IÖR) wurden Möglichkeiten der Verbesserung der Zugänglichkeit und Interoperabilität historischer Geoinformationen erprobt. Ziel des Projektes war die Erarbeitung von Digitalisierungsempfehlungen für historische Karten aus Archivbeständen und die Entwicklung sowie der Pilotbetrieb eines Tools zur Mobilisierung und FAIRification von analoggebundenen Informationen. Die über die Extraktion von Karteninhalten gewonnenen Geo- und Metadaten bedienen verschiedene Standards und stehen somit neben der Biodiversitätsforschung weiteren Communities zur Verfügung. Über eine generische Archivierungsschnittstelle werden die Daten in einem Digitalen Langzeitarchiv ingestiert und können in ausgewählten Fachportalen publiziert werden.

Schlüsselwörter: historische Karten, Map Content Extraction, Geodaten

Abstract: In a collaborative project between the State Archives of Bavaria and the Leibniz Institute of Ecological Urban and Regional Development (IOER), funded by NFDI4Biodiversity, possibilities for enhancing the accessibility and interoperability of historical geoinformation were developed and tested. The aim of the project was to develop digitization recommendations for historical maps from archival collections and to engineer and pilot a tool for the mobilization and FAIRification of analog-bound information. The geodata and metadata obtained through the automated extraction of map content serve various standards and are thus available to other science communities in addition to biodiversity research. Via a generic archiving in-

terface, the data are ingested in a digital long-term archive and can be published in selected research portals.

Keywords: historical maps, map content extraction, geodata

1 Neuentdeckung historischer Karten für unterschiedliche Wissenschaftsdisziplinen

Historische Kartensammlungen gehören zu den klassischen Sammlungsschwerpunkten in Museen und Bibliotheken und bilden häufig anzutreffende Pertinenz- und Sammlungsbestände in Archiven. Ihnen wird in den letzten Jahrzehnten vermehrt Aufmerksamkeit geschenkt, seitdem die Geisteswissenschaften und insbesondere die Geschichtswissenschaft den „Raum“ als Untersuchungsgegenstand und Bezugspunkt wieder stärker in den Blick genommen haben.¹ Wurden herausragende Karten und Pläne schon seit langem als Exponate gezeigt, widmen sich nun ganze Ausstellungen den Sammlungen und Beständen und thematisieren die Herstellungsmethoden und Vermessungstechniken vergangener Jahrhunderte.²

Zudem entschieden sich in den letzten zwanzig Jahren viele Institutionen, zunächst herausragende ältere Bestandteile und zunehmend auch ihre kompletten Kartensammlungen und Planbestände zu digitalisieren und online zu-

1 Schlögel, Karl. *Im Raume lesen wir die Zeit – Bücher. Über Zivilisationsgeschichte und Geopolitik*. München: Carl Hanser Verlag, 2003.

2 Leidel, Gerhard. *Von der gemalten Landschaft zum vermessenen Land: eine Ausstellung des Bayerischen Hauptstaatsarchivs zur Geschichte der handgezeichneten Karte in Bayern*. München, 6. Oktober bis 22. Dezember 2006. Ausstellungskataloge der Staatlichen Archive Bayerns, Bd. 48. München: Generaldirektion der Staatlichen Archive Bayerns, 2016.

gänglich zu machen. Die Digitalisierung bot damit auch die Chance, institutionell verstreute historisch bedeutsame Kartensammlungen virtuell wieder zusammenzuführen und über zentrale und spartenübergreifende Portallösungen, wie etwa bavarikon (Portal zur Kunst, Kultur und Landeskunde)³ oder LEO-BW (Landeskundliches Informationssystem Baden-Württemberg),⁴ zugänglich zu machen. Dies betrifft auch Geoinformationen großer ehemals zusammenhängender Kultur- und Wirtschaftsräume, wie dies etwa in einem langjährigen EU-geförderten Kooperationsprojekt der Staatlichen Archive Bayerns mit tschechischen Archiven gelungen ist, bei dem über das eigens aufgebaute Rechercheportal Porta Fontium historische Karten und Pläne zum bayerisch-tschechischen Grenzraum aus insgesamt 13 Archiven digitalisiert und erschlossen wurden.⁵ Dabei wurden Methoden der händischen Georeferenzierung sowie des automatisierten Auslesens verschriftlicher Informationen auf historischen Karten erprobt.⁶ Neben

KI-gestützten Verfahren setzen Bibliotheken und zunehmend auch Archive Crowdsourcingprojekte zur Georeferenzierung um.⁷ Bei vielen dieser Projekte wurden schon Methoden und Tools zur Map Content Extraction an bereits digitalisierten Kartensammlungen erprobt.⁸ Die konservatorischen Anforderungen bei der Digitalisierung geeigneter Kartenbestände, uneinheitliche Scanparameter bereits digitalisierter Sammlungen und der Langzeiterhalt der extrahierten Informationen stellen solche Projekte immer wieder vor Herausforderungen.⁹ Das Projekt Hist4Biodiv¹⁰ setzt bei diesen Punkten an: Neben einem Tool zur automatisierten Datenextraktion wurden Empfehlungen¹¹ und ein Workflow entwickelt, die bereits bei der Digitalisierung analogen Archivguts ansetzen. Die aus den Karten extrahierten Geoinformationen und Metadaten werden im Anschluss nicht nur in die Langzeitarchivierung überführt, sondern können durch die unterschiedlichen Ausgabefor-

³ Vgl. u. a. die Karten der Pfalz-Neuburgischen Landesaufnahme, die über das Kulturportal Bavariikon zugänglich gemacht wurden in: Hadry, Sarah. „Die Pfalz-Neuburgische Landesaufnahme, 1579/84-1604.“ *Bavariikon*. <https://www.bavariikon.de/object/bav:BSB-CMS-0000000000004084?lang=de>. Zuletzt geprüft am 09.09.2023; sowie die vorgestellten Projekte aus verschiedenen europäischen Ländern in: Benedetti, Benedetto u. a., Hrsg. *Cartography and Cadastral Maps Visions from the Past for a Vision of our Future*. Pisa: Edizioni della Normale, 2015. <https://edizioni.sns.it/prodotto/pdf-scariabile-3/>. Zuletzt geprüft am 09.08.2023; Zimmermann, Georg. „Historische Massensammlungen im virtuellen Verbund. Digitalisierung, Erschließung und Georeferenzierung historischer Karten im Kartenforum und im Virtuellen Kartenforum 2.0 der Sächsischen Landesbibliothek – Staats- und Universitätsbibliothek Dresden.“ In: *Die Masse macht's? Erschließungsmethoden und Erkenntnismöglichkeiten bei der Arbeit mit Massenbeständen*. Hrsg. von Wolfgang Hesse, Holger Starke. 41–48. Dresden: Stadtmuseum Dresden, 2019.

⁴ <https://www.leo-bw.de/web/guest/home>. Zuletzt geprüft am 16.08.2023.
⁵ Halla, Karel. „Porta fontium. Bayerisch-tschechisches Netzwerk digitaler Geschichtsquellen.“ In: *Archive im Web: Erfahrungen, Herausforderungen, Visionen/Archives on the Web: experiences, challenges, visions*. Hrsg. von Thomas Aigner, Stefanie Hohenbruck, Thomas Just u. a. 66–70. St. Pölten: Diözesanarchiv St. Pölten, 2010. Im Rahmen des NFDI Hackathons „Creating New Dimensions“ wurde ein Teilbestand aus Porta Fontium als Datenset aufbereitet: Generaldirektion der Staatlichen Archive Bayerns (Hg.). *Dataset: Porta fontium – Karten und Pläne zur bayerisch-tschechischen Geschichte. (Image) Metadaten Karten und Pläne zur bayerisch-tschechischen Geschichte*. 2022. DOI:10.22000/726. <https://radar4culture.radar-service.eu/radar/en/dataset/ktfsrSJfFlzjwje#>. Zuletzt geprüft am 09.08.2023.

⁶ Herold, Hendrik. *Geoinformation from the Past. Computational Retrieval and Retrospective Monitoring of Historical Land Use*. Wiesbaden: Springer Spektrum, 2018, 43–65; Chiang, Yao-Yi, Stefan Leyk, Craig A. Knoblock. „A Survey of Digital Map Processing Techniques.“ *ACM Computing Surveys* 47, 1 (2014): 1–44. DOI:10.1145/2557423.; Chiang, Yao-Yi u. a. „Creating Structured, Linked Geographic Data from Historical Maps: Challenges and Trends.“ In: *Using Historical Maps in Scientific*

Studies: Applications, Challenges, and Best Practices. Hrsg. von Yao-Yi Chiang, Weiwei Duan, Stefan Leyk u. a. 37–63. Cham: Springer, 2020. DOI:10.1007/978-3-319-66908-3_3.

⁷ Christoph, Andreas u. a. „Digitalisierung – Verortung – Recherche. Neue Perspektiven für historische Karten.“ *KN – Journal of Cartography and Geographic Information* 66, 3 (2016): 115–122. DOI:10.1007/BF03545231.; Bill, Ralf, Nils Koldrack, Kai Walter. „Georeferenzierung alter topographischer Karten – Crowdsourcing versus Bildverarbeitung.“ *AGIT Journal für Angewandte Geoinformatik* 1 (2015): 540–549. DOI:10.14627/537557074.; Sobotkova, Adela u. a. „Creating large, high-quality geospatial datasets from historical maps using novice volunteers.“ *Applied Geography* 155 (2023): 1–13. DOI: 10.1016/j.apgeog.2023.102967.; Bill, Ralf, Kai Walter. „Crowdsourcing zur Georeferenzierung alter topographischer Karten – Ansatz, Erfahrungen und Qualitätsanalyse.“ *Zfv – Zeitschrift für Geodäsie, Geoinformation und Landmanagement* 3 (2015): 172–179. DOI:10.12902/zfv-0060-2015.

⁸ Crom, Wolfgang. „Kartendigitalisierung – buntes Bild oder Mehrwert.“ *KN – Journal of Cartography and Geographic Information* 66 (2016): 243–248.; Frischknecht, Steffen, Entela Kanani. „Automatic interpretation of scanned topographic maps: A raster-based approach.“ In: *Graphics Recognition Algorithms and Systems*. Hrsg. von Karl Tombre, Atul K. Chhabra. Lecture Notes in Computer Science, Bd. 1389. Berlin, Heidelberg: Springer, 1998. DOI:10.1007/3-540-64381-8_50.

⁹ Herrault, Pierre-Alexis u. a. „Automatic Extraction of Forests from Historical Maps Based on Unsupervised Classification in the CIELab Color Space.“ In: *Geographic Information Science at the Heart of Europe*. Hrsg. von Danny Vandenbroucke, Bénédicte Bucher, und Joep Crompvoets. 95–112. Cham: Springer International Publishing, 2013. DOI:10.1007/978-3-319-00615-4_6.; Liu, Tiange, Pengfei Xu und Shihui Zhang. „A review of recent advances in scanned topographic map processing.“ *Neurocomputing* 328 (2018): 75–87. DOI: 10.1016/j.neucom.2018.02.102.

¹⁰ Das Kooperationsprojekt zwischen IÖR und GDA wurde 2022 durch Flexfundmittel aus NFDI4Biodiversity gefördert.

¹¹ Hellstern, Michael. *Empfehlungen für die Digitalisierung historischer Karten zur Vorbereitung auf die Extraktion von Karteninhalten (Projekt Hist4Biodiv)*. Hrsg. von der Generaldirektion der Staatlichen Archive Bayerns. München: 2023. DOI:10.5281/zenodo.10021336.

mate und Austauschstandards¹² im Sinne eines kompletten Datenlebenszyklus verschiedenen Forschungscommunities zur Nachnutzung zur Verfügung gestellt werden.

2 Historische Biodiversitätsdaten und Geoinformation aus mehreren Jahrhunderten

Die öffentlichen Archive in Deutschland archivieren in steigendem Maße digitale Informationen der öffentlichen Verwaltung von bleibendem Wert, um diese dauerhaft zu verwahren, zu erschließen und wieder zugänglich zu machen.¹³ Über dieses Aussonderungsverfahren auf Grundlage der Archivgesetze gelangen große Datenmengen in die Archive, die auch für die Beantwortung aktueller wissenschaftlicher Fragen beispielsweise der Biodiversitätsforschung (NFDI4Biodiversity), der Erdsystemwissenschaften (NFDI4Earth) und Agrosystemforschung (FAIRagro) sowie weitere Disziplinen und damit auch für die Nationale Forschungsdateninfrastruktur (NFDI) und ihre Konsortien relevant sind.¹⁴ Dies betrifft sowohl digitale Informationen der Verwaltung (born digital), die in den letzten Jahren entstanden sind, als auch historische Archivbestände, die bis in das frühe Mittelalter zurückreichen. So konnten in Vorbereitung auf die Beteiligung der Staatlichen Archive Bayerns in der NFDI einschlägige Archivalienbestände für die jeweiligen Forschungscommunities, u. a. historische Biodiversitätsdaten, identifiziert werden.¹⁵ Wie anderen Gedächtnisinstitutionen bieten sich den Archiven Chancen, mit Hilfe innovativer Methoden benachbarter Disziplinen, wie den Digital Humanities, diese analoggebundenen Daten zu mobilisieren und für die Wissenschaft zugänglich zu machen.¹⁶

¹² GML (<https://www.ogc.org/standard/gml/>), EAD (<https://wiki.deutsche-digitale-bibliothek.de/pages/viewpage.action?pageId=19010180>) sowie ABCD (<https://abcd.tdwg.org/xml/>). Alle zuletzt geprüft am 15.08.2023.

¹³ Puchta, Michael. *Fachkonzept für das Digitale Archiv der Staatlichen Archive Bayerns*. München: 2023. <https://zenodo.org/record/7743888#.ZImCyHZBxPY>. Zuletzt geprüft am 14.06.2023.

¹⁴ Grau, Bernhard, Markus Schmalzl und Michael Unger. „Die Beteiligung der Staatlichen Archive Bayerns in der NFDI. Ein Werkstattbericht.“ *Archiv. Theorie & Praxis*. 76,1 (2023): 40–42. DOI:10.5281/zenodo.8131770.

¹⁵ Mehr zu den historischen Daten im Kontext von NFDI4Biodiversity: <https://www.nfdi4biodiversity.org/de/gda-call-for-project-partner/>. Zuletzt geprüft am 15.08.2023.

¹⁶ Jaillant, Lise, Katie Aske, Eirini Goudarouli und Natasha Kitcher. „Introduction: Challenges and Prospects of Born-Digital and Digitized Archives in the Digital Humanities.“ *Archival Science* 22,3 (2022): 285–

In Hist4Biodiv wurden der Umgang mit solchen historischen Daten und ihre Verfügbarmachung für die Biodiversitätsforschung in einem Pilotprojekt erprobt. Denn viele Archivbestände, die bislang v.a. für geisteswissenschaftliche Fragestellungen ausgewertet wurden, bieten auch relevante Informationen für andere Disziplinen. So zieht die Wirtschafts- und Sozialgeschichte als frühes Beispiel mittelalterlicher Überlieferung zu Rechts- und Besitzverhältnissen geistlicher und weltlicher Herrschaftsträger die Urbare als Quellen heran, die teilweise und in Vorformen auch in Bayern seit dem 8. Jahrhundert überliefert sind.¹⁷ Sie enthalten, ähnlich wie die Ernteregister der Klöster, die teils ab dem Spätmittelalter bis ins 19. Jahrhundert in geschlossenen Serien in den Archiven verwahrt werden und Auskunft über Ernteerträge der Landwirtschaft geben, Informationen zur Landnutzung, die auch Rückschlüsse auf klimatische Veränderungen erlauben.¹⁸ Mit der dichteren Überlieferung einer sich professionalisierenden und ausdifferenzierenden Verwaltung seit der Frühen Neuzeit verbreitert sich auch das Spektrum einschlägiger Datenbestände für naturwissenschaftliche Fragestellungen.¹⁹ Neben der Urkunden- und Amtsbuchüberlieferung stehen ab dem 15. Jahrhundert umfangreiche Aktenbestände und Berichtsserien zur Verfügung, die mitunter auch Rückschlüsse zur Intensität der Landnutzung ganzer Territorien für einzelne Zeitschnitte erlauben, wie etwa die Hauptsteuerbeschreibungen der niederbayrischen Pflegerichte für 1612 und 1721.²⁰ Die statistischen Erhebungen des 19. Jahrhunderts erlauben teils ganz direkte Auswertungen zur Biodiversität. Beispielsweise erging 1845 der Auftrag an die bayerischen Forstverwaltungsbezirke, anhand eines Formularblatts die Fauna in den zugehörigen

291. DOI:10.1007/s10502-022-09396-1.; Ries, Thorsten und Gábor Palkó. „Born-digital archives.“ *International Journal of Digital Humanities* 1,1 (2019): 1–11. DOI:10.1007/s42803-019-00011-x.

17 Bader, Matthias. „Urbare.“ In: *Historisches Lexikon Bayerns*. 2014. <https://www.historisches-lexikon-bayerns.de/Lexikon/Urbare>. Zuletzt geprüft 14.08.23. Wild, Joachim u. a., Hrsg. *Die Fürstenkanzlei des Mittelalters: Anfänge weltlicher und geistlicher Zentralverwaltung in Bayern: Ausstellung des Bayerischen Hauptstaatsarchivs anlässlich des VI. Internationalen Kongresses für Diplomatik, München, 25. Oktober-18. Dezember 1983*. Ausstellungskataloge der Staatlichen Archive Bayerns, Bd 16. Neustadt a.d. Aisch: Kommissionsverlag Degener, 1983, 135–146; Bünz, Enno. „Urbare und verwandte Quellen zur Wirtschafts- und Sozialgeschichte.“ In: *Aufriß der historischen Wissenschaften*. Bd. 4: Quellen. Hrsg. von Michael Maurer. 7 Bde, 168–189. Stuttgart: Reclam, 2002.

18 So etwa die Zehnt- und Gültregister des Amts Alburg im Bestand Kloster St. Nikola aus den Jahren 1403–1802. BayHStA, Kloster St. Nikola 253–354.

19 Volkert, Wilhelm, Konrad Ackermann, Alois Schmid, Hrsg. *Staat und Verwaltung in Bayern: Festschrift für Wilhelm Volkert zum 75. Geburtstag*. Schriftenreihe zur bayerischen Landesgeschichte Bd. 139. München: C.H. Beck, 2003.

20 StALa, Landschafts Unterland Hauptsteuerbeschreibungen 1–49.

| | | |
|----|----------|---|
| 2. | Vorname | Wolfrat und sonstige Bezeichnungen. |
| | Wolfrat | |
| | Philipp | |
| | Valentin | Der Walpurgis ist vor vorzüllicher Größe und dauerndem Leben, wenn man den Namen stattet. Sie ist die einzige und alleinige Person, welche die Walpurgis in Mauren. |
| | Wolfrat | |
| | Phil. | Der vorzülliche Frey, der beginnt vor mehr als 1000 Jahren in Europa ist der Wolf, welche ist ein sehr alter, der selbige ist jetzt bei den vielen Forstgebäuden noch zu leicht zu erkennen, Sie ist aber sehr selten. |
| | Gambr | |

Abb. 1: Auszug aus dem Formularblatt zur Erfassung der Fauna in den zugehörigen Forstämtern

Forstämtern zu erfassen (s. Abb. 1). Das Ergebnis ist eine flächendeckende Erhebung des Vorkommens ausgewählter Tierarten in ganz Bayern für den Zeitschnitt 1845.²¹

Seit dem 15. Jahrhundert sind kartographische Darstellungen in den Staatlichen Archiven Bayerns überliefert, die zwar noch nicht auf exakter Vermessung der Erdoberfläche beruhen, aber dennoch wertvolle Aussagen zur Landnutzung und Landschaft erlauben. Mit der Herausbildung der amtlichen Vermessung, der Gründung eines topographischen Bureaus 1801 und der Nutzung der Trigonometrie mit dem Nordturm der Münchner Frauenkirche als Nullpunkt änderte sich das Verfahren grundlegend.²² Mit Hilfe moderner Messtechnik und großem Personaleinsatz wurde das Königreich Bayern komplett vermessen. Dies betraf nicht nur die 21 Millionen Grundstücke, sondern auch die Fläche des Landes. Zwischen 1808 und 1864 entstanden über 23 000 Messtisch- bzw. Urpositionsblätter im Maßstab

1:25 000, und damit die Basis für die Flurkarte, die heute mit Fernerkundung in digitaler Form erzeugt wird.²³

Die neuen Kartenwerke dienten nun nicht mehr nur militärischen Zwecken, sondern der Besteuerung und verschiedensten Aufgabenbereichen der zivilen Verwaltung. Entsprechend sind in den Staatlichen Archiven Bayerns nicht nur die älteren Flurkarten der Vermessungsverwaltung überliefert, die nach der Erstaufnahme auch für weitere Zeitabschnitte vorliegen. Vielmehr diente die Flurkarte ab der ersten Hälfte des 19. Jahrhunderts auch als Grundlage für Kartenwerke der verschiedenen Verwaltungszweige, wie der Vermessungs-, Steuer-, Umwelt und Landwirtschaftsverwaltung,

21 BayHStA, Zoologische Staatssammlung 112–119.

22 Habermeyer, Alfons. *Die topographische Landesaufnahme von Bayern im Wandel der Zeit*. Stuttgart: Wittwer, 1993, 25–38.

23 <https://www.ldbv.bayern.de/produkte/historisch/position.html>. Zuletzt geprüft am 16.08.2023;

Grau, Bernhard. *Uraufnahme: die Anfänge der Katastervermessung in Bayern*. Ausstellung des Bayerischen Hauptstaatsarchivs, 3. April–12. Mai 2006. München: Generaldirektion der Staatlichen Archive Bayerns, 2006.; Leidel, Von der gemalten Landschaft, 286–296; Habermeyer, Landesaufnahme, 79–85; 216. (Anhang 14: Abbildungsart und Blattschnitt der Katasterkarte [heute Flurkarte genannt]). Ziegler, Theodor. *Vom Grenzstein zur Landkarte*. 2. Aufl. Stuttgart: Wittwer, 1989.

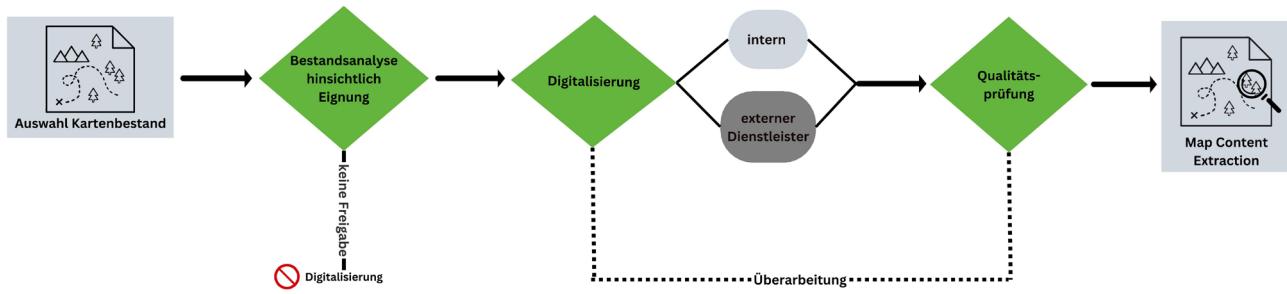


Abb. 2: Empfehlungen für die Digitalisierung historischer Karten

der Moorbewirtschaftung, der Bauverwaltung und Landesplanung oder für Planungszwecke zum Bau und Ausbau der Schienennetze im Königreich Bayern. Diese analog gebundenen Geofachdaten sind daher in den unterschiedlichsten Verwaltungszusammenhängen verschiedener Provenienzbildner überliefert, bislang nur in geringem Umfang digitalisiert, häufig nicht ausreichend mit Metadaten beschrieben und abgesehen von einer Zuweisung zu dem jeweiligen Flurkartenblatt nicht georeferenziert. Sie sind damit bislang in vielen Fällen nur schwer auffindbar, eingeschränkt zugänglich und nicht interoperabel auswertbar. Gerade eine gegenüberstellende Auswertung älterer und neuer Geoinformationen mit Hilfe von historischen Flurkarten eröffnet neue Forschungsperspektiven, wie dies in Einzelfallstudien bisher auf nicht-automatisiertem Wege erfolgt ist.²⁴ Der Bedarf an historischen Landnutzungsdaten²⁵ nimmt aktuell zu, wenn es beispielsweise um die Wiederverwässerung von Mooren zu Zwecken der CO₂-Bindung geht. Hierfür werden gerade auch ältere Kartenwerke herangezogen, die auf diesem Wege und neben aktuellen Geofach- und Geobasisdaten teils erst in die Archive gelangen und für künftige Auswertungen langzeitarchiviert werden.²⁶

24 Grau, Bernhard. „Der bayerische Wald im Kartenbild: von der Aufnahme zur digitalen Flurkarte.“ *Archive in Bayern* 7 (2012): 369–399. Die digitale Flurkarte Bayerns stellt das Landesamt für Digitalisierung, Breitband und Vermessung zur Verfügung: <https://www.ldbv.bayern.de/produkte/kataster/flurkarte.html>. Zuletzt geprüft am 17.08.2023. Ludwig, Robert. „Das Liegenschaftskataster auf dem Weg zum Geoinformationssystem.“ In: Es ist ein Maß in allen Dingen. 200 Jahre Bayerische Vermessungsverwaltung 1801–2001. Hrsg. von dem Bayerischen Staatsministerium der Finanzen, 200–211. München: 2001.; Hampp, Dieter, Franz Lindenthal. „Die Flurkarte erobert den PC. Der Einsatz der interaktiven Grafik bei der Bayerischen Vermessungsverwaltung.“ In: Es ist ein Maß 2001: 238–245.

25 Herold, Hendrik. „Big Historical Geodata for Urban and Environmental Research.“ In: *Handbook of Big Geospatial Data*. Hrsg. von Martin Werner, Yao-Yi Chiang. 475–486. Cham: Springer International Publishing, 2021. DOI: 10.1007/978-3-030-55462-0.

²⁶ So etwa im Projekt KliMoBay, in dem u. a. auch historische Karten des sogenannten Moorarchivs der Bayerischen Landesanstalt für Landwirtschaft ausgewertet wurden, die nun beim Bayerischen Haupt-

3 Digitalisierungsempfehlungen als einheitliche Ausgangsbasis für die Map Content Extraction

Das vollautomatisierte Auslesen maschinenlesbarer Inhaltsinformationen aus Karten stellt Projektvorhaben immer noch vor Herausforderungen.²⁷ In dem Projekt Hist4Biodiv wurden deshalb „Empfehlungen für die Digitalisierung historischer Karten“²⁸ entwickelt, um eine einheitliche Ausgangsbasis für die Anwendung eines Tools zur Extraktion von Karteninhalten zu schaffen und Hilfestellung für einen die Originale schonenden Umgang zu geben (s. Abb. 2).²⁹

In das Papier sind verschiedene Handreichungen zu Digitalisierungsempfehlungen in deutschen Archiven, Bibliotheken und Sammlungen eingeflossen.³⁰ Zwischen 2013 und 2015 beteiligte sich die Generaldirektion der Staatlichen Archive Bayerns an dem von der Archivschule Marburg koordinierten Projekt „Digitalisierung von archivalischen Quellen“.³¹ Auf dem Weg zu einer Vereinheitlichung von

staatsarchiv in digitalisierter und analoger Form dauerhaft archiviert werden <https://www.lfl.bayern.de/iab/kulturlandschaft/238568/index.php>. Zuletzt geprüft am 17.08.2023.

27 Schlegel, Inga. „A Holistic Workflow for Semi-automated Object Extraction from Large-Scale Historical Maps.“ *KN – Journal of Cartography and Geographic Information* 73, 1 (2023): 3–18. 4. DOI:10.1007/s42489-023-00131-z

28 Hellstern 2023

29 Beispiel für einen Digitalisierungsworkflow bei Glauert, Mario. „Dimensionen der Digitalisierung: Kosten, Kapazitäten und Konsequenzen.“ In: *Digital und analog. Die beiden Archivwelten. Rheinischer Archivtag, 21.–22. Juni 2012 in Ratingen*, Bd. 46, 48–59. Bonn: Habelt, 2012.

30 Holzapfl, Julian. „Digitalisierung und Bestandserhaltung – fachliche Synergie und planerisches Potential.“ *Archive in Bayern*, 9 (2016): 153–160.; Stumpf, Marcus. „Digitalisierungsstrategien in Deutschland – Versuch einer Bestandsaufnahme.“ *Archivpflege in Westfalen-Lippe*, 80 (2014): 27–35.

31 Eine Zusammenstellung von Handreichungen und Beiträgen sind hier zu finden: <https://www.archivschule.de/DE/forschung/digitalisierung-archivalischer-quellen/>. Zuletzt geprüft am 13.08.2023.

Digitalisierungsvorhaben im deutschen Archivwesen sind zudem die „Handreichung zur Digitalisierung von Archivgut“³² sowie das Grundlagenpapier „Archiv- und Bibliotheksgut schonend digitalisieren“ als wichtige Meilensteine zu nennen.³³ Die Digitalisierung von kulturellem Erbe wird weiterhin intensiv diskutiert³⁴ und auch die DFG hat ihre 2016 erstmals veröffentlichten Praxisregeln, insbesondere vor dem Hintergrund einer Nationalen Forschungsdateninfrastruktur, überarbeitet.³⁵

Zur genauen Kalkulation des zeitlichen Umfangs sowie eines Ressourcen- und Kostenplans für ein Digitalisierungsprojekt dient in einem ersten Schritt die gründliche Analyse eines in Frage kommenden Bestands. Im Projekt wurden Empfehlungen erarbeitet, welche Aspekte hierbei und zu einer Reihe weiterer wichtiger Fragen entlang des Digitalisierungsprozesses von Karten und Plänen zu berücksichtigen sind:³⁶

- Eignen sich Trägermaterial und Formate der Karten und Pläne und haben diese evtl. Auswirkung auf die Auswahl der Digitalisierungstechniken?³⁷
- Sind Schäden nach einer einheitlichen Matrix erfasst und ggf. Zustandsprotokolle für herausragende Einzelstücke erstellt?³⁸

³² Bundeskonferenz der Kommunalarchive beim Deutschen Städtetag (BKK) (Hg.). *Handreichung zur Digitalisierung von Archivgut*. Wolfsburg, 2017. https://www.bundeskongress-kommunalarchive.de/empfehlungen/Handreichung_Digitalisierung_von_Archivgut_endfassung.pdf. Zuletzt geprüft am 18.09.2023.

³³ Bundesarchiv (Hg.). *Archiv- und Bibliotheksgut schonend digitalisieren*. 2019. https://www.bundesarchiv.de/DE/Content/Downloads/KLA/digitalisierung-grundlagenpapier.pdf?__blob=publicationFile. Zur Verzeichnung von Schäden: Bachmann, Christoph u. a. *Richtlinien zur Verzeichnung von Archivgut der Staatlichen Archive Bayerns (Version 1.0)*. Hrsg. von der Generaldirektion der Staatlichen Archive Bayerns. 2020. 14. https://www.gda.bayern.de/fileadmin/user_upload/Medien_fuer_Unterseiten/Verzeichnungsrichtlinien_11_Version_1.0_-14.pdf. Alle zuletzt geprüft am 08.08.2023.

³⁴ Wilkens, Jan Gerd. „Tagungsbericht: Herausforderungen, Chancen und Grenzen digitaler Sammlungen.“ *O-Bib. Das Offene Bibliotheksjournal* 10,2 (2023): 1–7. DOI:10.5282/o-bib/5925.

³⁵ Deutsche Forschungsgemeinschaft (Hg.). *DFG-Praxisregeln „Digitalisierung“*. Aktualisierte Fassung. 2022. DOI:10.5281/zenodo.7435724.

³⁶ Hellstern 2023: 4–6.

³⁷ Stockert, Harald: „Digitalisierung von Großformaten – Methoden und Anforderungen an verwaltungsinterne Kooperationspartner oder Dienstleister.“ *Archivpflege in Westfalen-Lippe*, 85 (2016): 44–50.

³⁸ Holzapfl, Julian, Karin Marth. *Schadenserfassung als Wegweiser. Bestandserhaltungsmaßnahmen im Archiv systematisch planen*. 2016. https://www.gda.bayern.de/fileadmin/user_upload/Medien_fuer_Unterseiten/Handreichung_Schadenserfassung.pdf. Zuletzt geprüft am 09.08.2023.; Bundesarchiv 2019: 3–9.

- Gibt es urheber- oder archivrechtliche Einschränkungen für die unterschiedlichen Nutzungsszenarien und die Publikation der Digitalisate?³⁹
- Ist die korrekte Grunderschließung aller Archivalieneinheiten des Bestands gegeben,⁴⁰ die den jeweiligen Erschließungsstandards der bestandshaltenden Institution genügt?⁴¹
- Wie und unter welchen Bedingungen können die einzelnen Karten und Pläne digitalisiert werden oder müssen von der Maßnahme aus konservatorischen Gründen ausgeschlossen werden?
- Welche Verbesserungen müssen künftig bei der Lagerung der Archivalien geschaffen werden, um künftige Schäden zu verhindern?
- Welche Aspekte müssen bei der Beauftragung eines Dienstleisters, beim Transport der Karten und Pläne und bei der Qualitätskontrolle beachtet werden.⁴²

Zur Optimierung der Digitalisate für die automatisierte Extraktion von Karteninhalten im Projekt Hist4Biodiv wurden zudem technische Vorgaben zusammengestellt:

- Entsprächen die erzeugten Dateien einschließlich der Probedigitalisate den festgelegten technischen Parametern, insbesondere hinsichtlich des TIFF-Formats zur Vermeidung von Einspielfehlern?⁴³
- Enthalten die Dateien die notwendigen technischen Metadaten zum Scanprozess und folgen diese der vorab festzulegenden Dateinamenskonvention?

³⁹ Puchta: 2023: 49; DFG 2022: 11.

⁴⁰ Friedrich, Tanja, Jonas Recker. „Auffindbarkeit und Nutzbarkeit von Daten“. In: *Praxishandbuch Forschungsdatenmanagement*, Hrsg. von Markus Putnings, Heike Neuroth, Janna Neumann. 405–426. Berlin/Boston: De Gruyter Saur, 2021. DOI: 10.1515/9783110657807; Iglezakis, Dorothea u. a. „Interoperabilität von Metadaten innerhalb der NFDI.“ *Bausteine Forschungsdatenmanagement* 2 (2021): 124–135.

⁴¹ Bachmann u. a. 2020: 43–46; Hellstern 2023: 5–6. Zum EAD-Format siehe: <https://wiki.deutsche-digitale-bibliothek.de/pages/viewpage.action?pageId=19010180>. Zuletzt geprüft am 13.08.2023.

⁴² Hellstern 2023: 6–8. Zu externen Vertragsaufgaben: DFG 2022: 9–11; 36–37. Zu Transportfragen: Bundesarchiv, Hrsg. *Empfehlungen der KLA zum bestandsschonenden Transport von Archivgut*. 2020. https://www.bundesarchiv.de/DE/Content/Downloads/KLA/bestandsschonender-transport-archivgut.pdf?__blob=publicationFile#:~:text=Archivgut%20sollte%20grunds%C3%A4tzlich%20liegend%20transportiert,ten%2D%20%2FB%C3%BCcherwagen%20transportiert%20werden. Zuletzt geprüft am 09.08.2023.

⁴³ Hellstern 2023: 10–12. Zu TIFF als Format für die Langzeitarchivierung: Enders, Markus. „Bildddokumente“. In: *nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Hrsg. von Heike Neuroth u. a. Aufl. 2.3., 8–18. Göttingen: Nestor, 2010. http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_422.pdf 2010, 17:11. Zuletzt geprüft am 18.09.2023.



Abb. 3: Digitalisat eines Flurkartenblattes. Anhand dieser und ähnlicher Flurkarten wurde die Metadaten-Extraktion getestet

- Gibt es Richtlinien für die mobilen Datenträger bei der Zwischenspeicherung der Digitalisate?
- Entsprächen Aufnahmeformat und Auflösung den Zielvorgaben für die automatisierte Datenextraktion (in Hist4Biodiv 508 dpi)?⁴⁴
- Werden die Qualitätsstandards für eine Langzeitspeicherung bzw. Langzeitarchivierung der Digitalisate in dem jeweiligen Repositorium bzw. Digitalen Langzeitarchiv erreicht?⁴⁵

Anhand dieser Empfehlungen wurden Digitalisate analoger Flurkarten aus dem Bayerischen Hauptstaatsarchiv erstellt, um diese bestmöglich für die Anwendung eines vom Leibniz-Institut für ökologische Raumentwicklung entwickelten Tools zur Datenextraktion bereitzustellen (s. Abb. 3).

4 Extraktion von Karteninhalten und Metadaten

Durch die Extraktion von Karteninhalten wird angestrebt, nicht nur den Kartenscan selbst, sondern auch die extrahierten Informationen so gut zugänglich wie möglich zu machen. Dies hängt mit dem eher technischen Teil der Interoperabilität zusammen, d. h. mit der Verwendung offener Standards, Formate und Software. Indem wir die Exportfunktionalität für eine Vielzahl von Dateiformaten vorbereiten, ermöglichen wir die freie Wahl der Software auf der Benutzeroberseite. Außerdem ermöglicht die Implementierung in Open-Source-Code und -Software, die Definition einer geeigneten Exportschnittstelle für die Archivierung sowie die Bereitstellung ausreichenden Dokumentationsmaterials die Reproduzierbarkeit und Wiederverwendbarkeit des Ansatzes zur Datenextraktion und den Langzeiterhalt der extrahierten Informationen.

⁴⁴ Hellstern 2023: 10–13.

⁴⁵ Die „Empfehlungen“ enthalten beispielhaft den Workflow zur Langzeitspeicherung der Staatlichen Archive Bayerns sowie technische Vorgaben für die Masterdigitalisate (Hellstern 2023: 13).

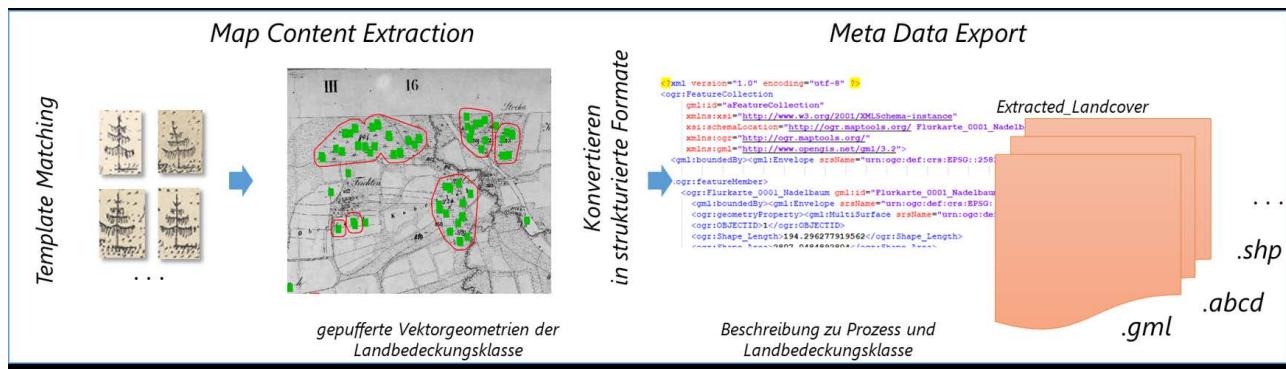


Abb. 4: Workflow der Extraktion von Karteninhalt und Metadaten

4.1 Map Content Extraction

Der im Folgenden beschriebene Workflow ermöglicht es, die Kartenscans um strukturierte und möglichst interoperable Metadaten anzureichern. Hierfür wurde ein zweistufiges Verfahren implementiert. Zunächst wird darauf abgezielt, mit Methoden der Computer Vision (d. h. Farbseparation und Template Matching) Karteninhalt zu extrahieren (Map Content Extraction). Aus dem Ergebnis dieser Extraktion wird dann ein Metadatensatz aufgebaut, der den Karteninhalt beschreibt und in möglichst viele Ausgangsformate exportiert werden kann (s. Abb. 4).

Eine erste Analyse betrifft also den Farinhalt der gescannten Karten. Damit ist nicht die rein technische Frage gemeint, ob Farbkanäle im Bildformat gespeichert sind, sondern vielmehr die durch das menschliche Auge wahrnehmbare Farbkodierung der semantischen Karteninhalte. Ersteres ist ohne großen Aufwand zu beantworten, der zweite Sachverhalt lässt sich nur klären, wenn man die Mischung der Pixelfarben zu Grauwerten, oder die Schwellen für den Weißwert berücksichtigt. Dazu stelle man sich vor, dass eine historische Karte ohne farbliche Inhalte auf vergilbtem Papier eingescannt wurde. Das gescannte Bild wird rein technisch verschiedene Farbkanäle (d. h. Pixel mit Rot- oder Blauwerten) enthalten, obwohl der Inhalt der Karte für das menschliche Auge nur aus einfarbigen Grauwerten der graphischen Zeichen besteht. Ob eine solche Karte tatsächlich farbliche Inhalte enthält kann, über eine statistische Auswertung der einzelnen Pixelwerte beantwortet werden. Mit dieser Information kann dann entschieden werden, ob eine Farbseparation für ein solches Kartenwerk in Frage kommt. Mittels der Farbseparation können die farblich hinterlegten Inhalte direkt über ihre Farbbereiche selektiert und extrahiert werden. In den meisten Fällen sind Liniensignaturen der Gewässer, Höhenlinien oder Flächen der Vegetation auf diesem Wege extrahierbar.

Template Matching wird verwendet, um in einem gescannten Kartenwerk wiederkehrende Signaturen zu identifizieren. Das Verfahren basiert darauf, aus einem Bestand an kleinen Bildausschnitten (Templates) ähnliche Bereiche in größeren Bildern zu finden. Diese Bildausschnitte müssen zunächst aus dem Kartenwerk zusammengestellt werden, d. h. es müssen einige Signaturen in einem Kartenscan ausgewählt und als Beispiel gespeichert werden. In den meisten historischen Karten sind die Signaturen handgezeichnet und weisen dadurch eine gewisse Variabilität auf. Die gewählten Templates sollten diese Variabilität widerspiegeln, und dennoch möglichst repräsentativ sein. Wenn die semantische Tiefe des Kartenwerkes groß genug ist, sind mit diesem Verfahren sehr performante inhaltliche Auswertungen möglich. So lassen sich Signaturen für Laub- und Nadelbäume identifizieren und die entsprechenden Waldflächen geometrisch abgrenzen. Ziel der hier erwähnten Verfahren ist es letztlich, die Landbedeckung (LB) aus einer historischen Karte abzuleiten und in strukturierte Metadatenformate überführen zu können. Für den folgenden Schritt des Metadaten-Exports ist es nachrangig, mit welchem Extraktionsverfahren die Vektorgeometrien der Landbedeckung erstellt werden.

4.2 Generierung und Export von Metadaten

Am Beispiel des Template Matchings wird an dieser Stelle die Extraktion der Metadaten dargestellt. Als Eingangsdaten dienen Scans der oben erwähnten bayerischen Urpositionsblätter, die mit 23 000 Blatt ganz Bayern abdecken, in ihren Maßstäben je nach Region aber unterschiedlich sind (1 : 1250, 1 : 2500, 1 : 5 000).⁴⁶

46 https://www.ldbv.bayern.de/file/pdf/14548/Faltblatt_HistorischeKarten.pdf. Zuletzt geprüft am 18.08.2023.

```
<?xml version="1.0" encoding="utf-8" ?>
<DataSets xmlns:gml="http://www.opengis.net/gml/3.2"
  xmlns:ogr="http://ogr.maptools.org/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.tdwg.org/schemas/abcd/3.0
  http://rs.tdwg.org/abcd/3.0/ABCD_3.0.xsd
  http://ogr.maptools.org/
  gml:id="aFeatureCollection">
  <DataSet>
    <TechnicalContacts>
    <ContentContacts>
    <Metadata>
      <Description>
      <RevisionData>
    </Metadata>
    <Units>
      <Unit>
        <SourceInstitutionID>GDA_Institution_ID</SourceInstitutionID>
        <SourceID>ScannedTIF_ID</SourceID>
        <UnitID>ExtractedGeom_ID</UnitID>
        <DateLastEdited>2023-01-26T13:31:12</DateLastEdited>
        <Identifications>
          <RecordBasis>DrawingOrPhotograph</RecordBasis>
          <Gathering>
        </Unit>
      </Units>
    </DataSet>
  </DataSets>
```

Abb. 5: Die Zielformate für die Metadaten sind am ABCD-Schema orientiert

Der kartografische Zeichenschlüssel der Flurkarten gestattet die Unterscheidung verschiedener Bodenbedeckungen (u. a. Feuchtwiesen/Moore) und Waldarten. Anhand dieser Signaturen wurden Templates erstellt, je Bodenbedeckung 20 Stück. Die Anzahl der Templates ist dabei Abwägungssache, aber eine niedrige zweistellige Zahl an Templates fängt in der Regel die grafische Variabilität einer Signatur ganz gut ein. Für jedes dieser Templates werden in den Eingangskarten ähnlich aussehende Bildteile gesucht, markiert und in einer Ausgabemaske gespeichert. Diese Ausgabemaske wird anschließend in Vektordarstellung konvertiert, somit wird jede maskierte Position durch ein Polygon dargestellt. Diese Polygonflächen werden nun zusammengefasst und gepuffert, d. h. in gewisser Weise die Umwandlung von „Fläche mit Waldsignaturen“ in „Waldfläche“ durchgeführt. Diese Waldfläche verfügt, dank der Vektordarstellung, über Koordinaten und kann mit Attributen versehen werden. Die Koordinaten werden letztlich benötigt, um den Fundort der Waldfläche als Metadatum zu extrahieren.

Da die Zielformate für die Metadaten am ABCD-Schema orientiert sein sollten (s. Abb. 5), werden die Koordinaten im gml-Format ausgespielt. Dieser gml Abschnitt kann im ABCD-Schema eingefügt werden. Daneben können die Koordinaten leicht in andere Dateiformate (ESRI-Shapefile, Geopackage, NAS etc.) überführt werden. Aus Gründen der Interoperabilität und der langfristigen Interpretier-

barkeit werden offene XML-basierte Formate als Standardeinstellung genutzt, was auch weitestgehend den Anforderungen an die Langzeitarchivierung Rechnung trägt. Die Implementierung des Verfahrens erfolgte in offener Software, mit Hilfe der Programmiersprache Python und der Schnittstellen des QGIS-Geoinformationssystems. Auch die verwendeten Zusatz-Bibliotheken (openCV, numpy) sind frei zugänglich.

5 Sicherung der langfristigen Verfügbarkeit – Schnittstellen für die digitale Langzeitarchivierung

Neben der Festlegung von allgemeingültigen Parametern für die Kartendigitalisierung und der Extraktion von Daten aus historischen Karten war das Ziel des Projektes, diese auch langfristig nachnutzbar und interpretierbar zu halten. Die Überführung in die Langzeitarchivierung ermöglichen generische logische Archivierungs- und Exportschnittstellen, die in der GDA konzipiert wurden. Hierfür wurden wichtige Elemente wie die Nutzung möglichst offener, dokumentierter und nicht proprietärer Formate sowie dokumentierte Metadatenstandards berücksichtigt. Mit Hilfe eines vorgegebenen Schemas für die Bildung des Submission Information Package (SIP), liegen die Daten in der

ZIP-Datei PROLOG

Dateiname:
Archivkürzel_PROLOG_UUID.zip

- 1 XSD-Datei zur ABCD-XML
- 1 XSD-Datei zur INSPIRE-XML
- 1 PDF/A Empfehlungspapier Digitalisierung (Hellstern 2023)

ZIP-Datei DATEN

Dateiname:
Archivkürzel_Daten_UUID.zip

- 1 XML-Datei mit grundsätzlichen Erschließungsmetadaten (ABCD-konform) zum Gesamtpaket (ZIP-Datei PROLOG)
 - *Dateiname:* Archivkürzel_Erschließung_UUID.xml
 - *Inhalt:* datenhaltende Stelle, Entstehungsjahr, Titel, allgemeine Beschreibung, UUID, Lizenz, Herkunft Metadaten
- 1 CSV-Datei mit grundsätzlichen Erschließungsmetadaten zum Gesamtpaket (ZIP-Datei PROLOG)
 - *Dateiname:* Archivkürzel_Erschließung_UUID.csv
 - *Inhalt:* datenhaltende Stelle, Entstehungsjahr, Titel, allgemeine Beschreibung, UUID, Lizenz, Herkunft Metadaten
- 1 XML-Datei mit Metadaten zur Karte (INSPIRE-konform)
- 1 NAS-File
- 1 Shape-File
- 1 TIFF-Datei Digitalisat

Abb. 6: Das vorgegebene Schema für die Bildung des Submission Information Package (SIP)

korrekten Struktur für den Ingest über den Generalisierten XML-Client (GXAC) der Staatlichen Archive Bayerns vor (s. Abb. 6). Die nachvollziehbare Struktur, die Verwendung möglichst offener, weitgehend nicht proprietärer und gut dokumentierter Formate, die Auswahl geeigneten Dokumentationsmaterials und die Extraktion der Metadaten in eine zusätzliche CSV-Datei ermöglicht zudem als generische Archivierungsschnittstelle die Langzeitarchivierung in anderen Repositoryn.

Das SIP besteht aus zwei ZIP-Containern, „Prolog“ und „Daten“. Beide werden nach einer Dateinamenskonvention benannt, die das Archivkürzel der datenhaltenden Stelle sowie eine UUID enthält. Die ZIP „Prolog“ enthält eine CSV-Datei mit den Erschließungsmetadaten (Mindestanforderungen, s. o.) für ein Archivfachinformationssystem (AFIS) sowie Dokumentationsmaterial zu den Daten, um sie interpretierbar zu halten. Dazu zählen die Empfehlungen für die

Digitalisierung und die Schemata der bei der Metadatenextraktion bedienten Exportformate. Der Container „Daten“ enthält neben dem Digitalisat auch die extrahierten Metadaten gemäß unterschiedlicher XML-Austauschstandards. Mit dem XML-Standard ABCD⁴⁷ lässt sich die Publikation der Daten in Forschungsportalen der Biodiversitätscommunity realisieren, die innerhalb von NFDI4Biodiversity weiter ausgebaut werden. Durch die INSPIRE-Konformität⁴⁸ der Geoinformationen stehen die Daten zusätzlich den Erdsystemwissenschaften und der NFDI4Earth sowie weiteren wissenschaftlichen Communities zur Nutzung zur Verfügung.

⁴⁷ Holetschek, Jörg u. a. „The ABCD of Primary Biodiversity Data Access.“ *Plant Biosystems – An International Journal Dealing with all Aspects of Plant Biology* 146,4 (2012): 771–779. DOI:10.1080/11263504.2012.740085.

⁴⁸ <https://www.gdi-de.org/INSPIRE>. Zuletzt geprüft am 21.08.2023.

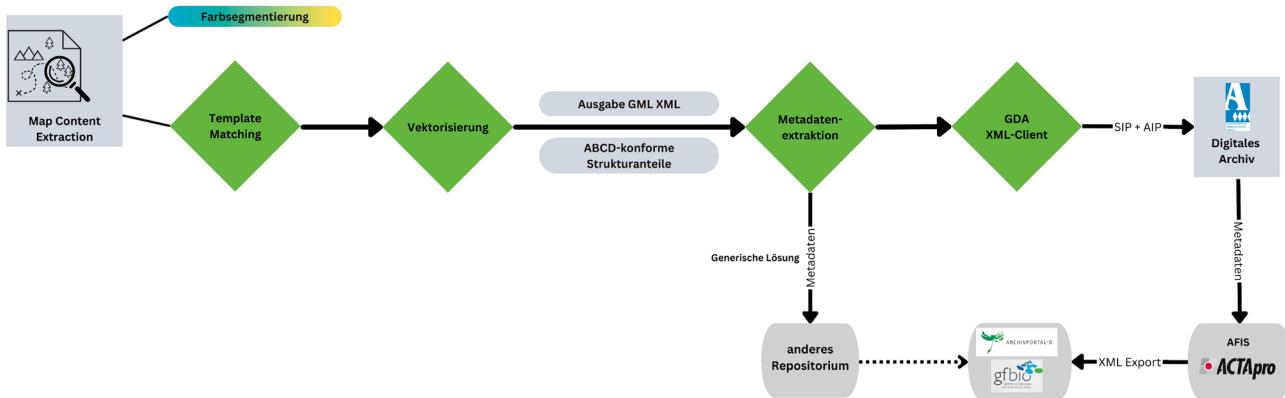


Abb. 7: Workflow: wie Daten über die vorhandenen Schnittstellen in den Standards EAD- und ABCD-XML für verschiedene Wissenschaftscommunities publiziert werden können

Der GXAC wurde für die Archivierung personen-, objekt- oder fallbezogener Daten aus komplexen Fachverfahren entwickelt.⁴⁹ Dieses Tool erlaubt es, unterschiedliche Daten einer Vielzahl von Einzelobjekten automatisiert in ein angeschlossenes Digitales Magazin zu ingestieren. Dabei können flexible Daten, also die hier extrahierten Metadaten, zusammen mit den logisch zugehörigen fixierten Daten, in diesem Falle den Kartendigitalisaten, strukturiert verarbeitet und jeweils in ein Archival Information Package (AIP) portioniert werden.⁵⁰ Bei der Verarbeitung der Daten aus der Metadatenextraktion werden also die extrahierten Metadaten und die zugehörige digitalisierte Karte gemeinsam zu einem AIP strukturiert (s. o.). Durch die automatisierte Dokumentation aller Verarbeitungsschritte wird deren Nachvollziehbarkeit und damit die data provenance gesichert. Außerdem ermöglicht es der GXAC, mit Blick auf eine künftige automatisierte Verarbeitung für einzelne Unterlagengruppen oder Ursprungssysteme jeweils zu konfigurieren, welche Abschnitte einer XML-Datei mit welchen definierten Dokumenten zu einem AIP strukturiert werden sollen und welche Daten aus der XML-Datei zur Grunderschließung im angeschlossenen Archivfachinformationsystem (AFIS) nachgenutzt werden sollen.

Von dort aus können die Daten über die vorhandenen Schnittstellen in den Standards EAD- und ABCD-XML für verschiedene Wissenschaftscommunities u. a. über das Archiv-

portal-D⁵¹ und GFBio⁵² publiziert werden (s. Abb. 7). Mit der umgesetzten Exportstruktur, der Dokumentation und den beim Export erzeugten XML-Erschließungsinformationen wird außerdem zusätzlich garantiert, die Informationen auf Grundlage anderer technischer Infrastrukturen in ein Repository überführen und andernorts publizieren zu können.

6 Zusammenfassung und Ausblick

Im Projekt Hist4Biodiv werden Aspekte des gesamten Datenlebenszykluses bei der Mobilisierung analoggebundener Daten aus historischen Kartenwerken adressiert: von Empfehlungen für die Digitalisierung über die tatsächliche Metadatenextraktion bis hin zur Langzeitarchivierung, Erschlie-

⁵¹ <https://www.archivportal-d.de/>. Zuletzt geprüft am 18.08.2023. Wolf, Christina. „Eines für alle: das Archivportal-D. Neue Zugangswege zu Archivgut.“ In: *Neue Wege ins Archiv – Nutzer, Nutzung, Nutzen. 84. Deutscher Archivtag 2014 in Magdeburg. Deutscher Archivtag*. Fulda: De Gruyter, 2016, 48–63; Wolf, Christina, Gerald Maier. „Das Archivportal-D und die Deutsche Digitale Bibliothek: neue übergreifende Recherchemöglichkeiten nach Quellen für die historische Forschung.“ In: *Recherche und Weiterverarbeitung: digitale Angebote der Archive für die historische Forschung im Netz. Beiträge einer Sektion auf dem 51. Deutschen Historikertag 2016 in Hamburg*. Hrsg. von Rainer Hering, Robert Kretzschmar. 10–35. Stuttgart: Kohlhammer, 2017.

⁵² Helbing, Franziska u. a. „GFBio: Schnittstelle zwischen institutionellem und disziplinärem Forschungsdatenmanagement im Bereich Biodiversitätsforschung, Ökologie und Umweltwissenschaften.“ In: *Forschung + Datenmanagement= Forschungs-datenmanagement – Wann geht die Gleichung auf? Bericht zur Abschlussstagung des PODMAN-Projektes*. Hrsg. von Lea Gerhards, Marina Lemaire, Stefan Kellendonk. 6–7. Trier: Online-Publikations-Server der Universität Trier, 2020. DOI: 10.25353/ubtr-xxxx-4728-6420; Grobe, Peter u. a. „Long-Term Reusability of Biodiversity and Collection Data using a National Federated Data Infrastructure.“ *Biodiversity Information Science and Standards* 3: e37414 (2019). DOI:10.3897/biss.3.37414.

⁴⁹ Puchta 2023: 13–17; Puchta, Michael, Markus Schmalzl. *Archival Preservation of Digital Databases – Teilautomatisierung oder Standardformat?* Vortrag gehalten auf: „Database History – Histories of Datebasing and Databasing of History“, 23. Juni, Basel, Switzerland. DOI: 10.5281/zenodo.8133541, <https://zenodo.org/record/8133541>. Zuletzt geprüft am 18.08.2023, 11–16.

⁵⁰ Keitel, Christian. *Zwölf Wege ins Archiv. Umrisse einer offenen und praktischen Archivwissenschaft*. Stuttgart: Franz Steiner Verlag, 2018, 140.

ßung und Publikation der extrahierten Informationen. Aus den im Test verfügbaren hochauflösten Digitalisaten konnten über Verfahren der Computer Vision Abgrenzungen für definierte Landbedeckungen abgeleitet werden. Diese können als Vektorgeometrien in strukturierte XML-basierte Metadatenformate eingebettet werden und gemeinsam mit wesentlichen Prozessparametern den ursprünglichen Archivdatensatz näher beschreiben. Günstig für die Analyse ist das Vorliegen eines umfangreicheren Kartenwerks mit festgelegtem Zeichen- bzw. Signaturenschlüssel.

In der bisherigen Implementierung müssen einige Parameter noch interaktiv gesetzt werden. Dies betrifft z. B. die zulässige Differenz zwischen Template und Bildausschnitt. Auch die Templates als solche sind zunächst eine Eingabegröße, bei der das MCE-Tool auf Interaktion angewiesen ist. An diesen Stellen kann zukünftig darauf hingearbeitet werden, dass Parameterwerte in bestimmten Bereichen voreingestellt sind und falls nötig nur feingestellt werden. Des Weiteren ist es denkbar, für verschiedene Landbedeckungsklassen eine Bibliothek an Templates anzulegen, die auf beliebige Kartenwerke angewendet werden kann.

Ein hier nicht adressierter Aspekt ist die Extraktion von Textinformation aus dem Kartenwerk. Die gestalterische Variabilität, gerade historischer Karten, macht die maschinelle Texterkennung zu einer besonderen Herausforderung. Dennoch können historische Topo- bzw. Choronyme wertvolle Hinweise der historischen Landbedeckung liefern, weshalb hier Potential für Ausbau und Weiterentwicklung des Tools gesehen wird. Weiterentwicklungsmöglichkeiten bestehen darüber hinaus auch bei der Extraktion von Fachinformationen, die durch weitere Symbole in die Karten eingetragen wurden oder beim georeferenzierten Extrahieren kartographischer Aufnahmen, die noch nicht auf exakter Vermessung beruhen.

Autorin und Autoren

**André Hartmann**

Raumbezogene Information und Modellierung
Leibniz-Institut für ökologische Raumentwicklung e.V.
Dresden
a.hartmann@ioer.de
orcid.org/0000-0003-0384-8501

**Hendrik Herold**

Raumbezogene Information und Modellierung
Leibniz-Institut für ökologische Raumentwicklung e.V.
Dresden
h.herold@ioer.de
orcid.org/0000-0002-2806-3121

**Anna Lisa Schwartz**

Stabstelle Forschung
Generaldirektion der Staatlichen Archive
Bayerns
München
annalisa.schwartz@gda.bayern.de
orcid.org/0000-0003-0391-3879

**Markus Schmalz**

Stabstelle Forschung
Generaldirektion der Staatlichen Archive
Bayerns
München
markus.schmalz@gda.bayern.de
orcid.org/0009-0005-7547-2699