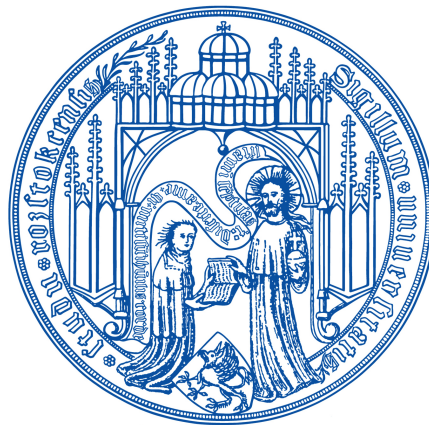

Bildarchive nutzbar machen
Eine Studie am Beispiel des Bildarchives des Freilichtmuseums
Mueß

Bachelorarbeit

Universität Rostock
Fakultät für Informatik und Elektrotechnik
Institut für Informatik



vorgelegt von: Andreas Görres
Matrikelnummer: 215204685
geboren am: 05. April 1986 in Frankfurt am Main
Erstgutachter: Dr.-Ing. Holger Meyer
Zweitgutachter: Professor Dr. rer. nat. habil. Andreas Heuer
Betreuer: Dr.-Ing. Holger Meyer
Volker Janke
Abgabedatum: 7. November 2018

Inhaltsverzeichnis

1	Einleitung und Problemstellung	5
2	Bedeutung der Citizen Science in den Digital Humanities	9
2.0.1	Gründe für Citizen Science	10
2.0.2	Einteilung von Citizen Science Projekten	10
2.0.3	Abgrenzung gegenüber CrowdSourcing	12
3	Stand der Technik und deren Grundlagen	13
3.1	Online Bildarchive	13
3.2	In der Anwendung eingesetzte Systeme	14
3.2.1	Media Pro SE	14
3.2.2	MediaWiki	19
3.3	Bilder und Metadaten-Formate	24
3.3.1	Exif	24
3.3.2	IPTC IIM	25
3.3.3	XMP	26
3.3.4	Zusammenfassende Bewertung der Formate Exif, IPTC und XMP	26
3.4	Bewertung der Metadatenverwaltung in Media Pro	27
3.5	Empfehlungen der Metadata Working Group	28
3.6	Hierarchische Schlüsselwörter	31
4	Ein System zur Kopplung von Bildarchiv und Online-Portal	35
4.1	Komponentenarchitektur der Software	35
4.2	Workflow	37
4.3	Datenbankarchitektur	40
5	Prototypische Implementierung und Bewertung	43
5.1	Prototypische Implementierung	43
5.2	Vergleich mit existierenden Systemen	44
5.3	Alternativansatz: Importieren und Exportieren in Echtzeit	45
6	Zusammenfassung und Ausblick	47
6.1	Zusammenfassung	47
6.2	Ausblick	47

Literaturverzeichnis

49

Kapitel 1

Einleitung und Problemstellung

Das Freilichtmuseum für Volkskunde Schwerin-Mueß verfügt über eine umfangreiche Sammlung historischer Fotografien. Der Altbestand von 70.000 analogen Aufnahmen wurde 2016 um 300.000 Fotografien aus den Archiv der Schweriner Volkszeitung erweitert (Abbildung 1.1). 2018 kamen weitere 150.000 Aufnahmen von Ernst Höhne, einem Fotografen der Schweriner Volkszeitung, hinzu. Um die Bilder zu sichern, zu erschließen und für weitere Recherchen nutzbar zu machen, werden sie digitalisiert und gleichzeitig mit Metadaten versehen. Zu diesen Zweck wurde ein Workflow definiert, der für Fotoalben aus den folgenden Schritten besteht[JH16]:

1. Nummerierung der Albumsseiten und Fotografien
2. Identifizierung schadhafter oder verunreinigter Objekte
3. Digitalisierung des Albums, der Seiten und der Fotografien
4. fachgerechte Aufbewahrung der Alben
5. Festlegen von Metadaten nach dem IPTC Core Standard
6. Sicherung der Digitalisate
7. Veröffentlichung in einer Online-Bilddatenbank

Die vorliegende Arbeit widmet sich dem letzten Schritt dieses Workflows, indem sie ein Konzept vorstellt, wie das „Ortschronikenportal Mecklenburg-Vorpommern“ [Ort18] als Veröffentlichungsplattform der digitalisierten Fotografien und zugehörige Metadaten des Freilichtmuseums genutzt werden kann. Die Nutzer des Portals sollen die Metadaten nicht nur perzipieren, sondern in begrenztem Ausmaß auch vervollständigen können.

Der bisherige Zustand der Metadaten-Erschließung und die Form der Bereitstellung erlauben nur einen sehr eingeschränkten Zugriff des Publikums auf die Ergebnisse des Digitalisierungsprojektes in Mueß. Eine Einsicht der Digitalisate ist nur durch persönliche Absprache mit dem stellvertretenden Museumsleiter des Museums möglich. Metadaten können zwar teilweise mit Bildbearbeitungs- bzw. Katalogisierungssoftware aus den Bildern selbst ausgelesen werden, es kann jedoch weder abgeschätzt werden, ob dem Publikum alle Metadaten-Felder auf diese Weise zur Verfügung stehen, noch, ob die Felder die beabsichtigten Namen tragen.

Ein weiteres Problem ist also die Vielfältigkeit der existierenden Metadaten-Formate und die Variabilität, mit der Anwendungsprogramme mit diesen umgehen. Nicht zuletzt aufgrund seiner guten Kompatibilität mit der bei der Digitalisierung verwendeten Aufnahmesoftware wird vom Freilichtmuseum Mueß die Katalogisierungssoftware Phase One Media Pro SE eingesetzt. Der Media Pro-Katalog selbst enthält zwar alle benötigten Metadaten der verwendeten Bilder, um jedoch auf diesen zuzugreifen, ist die Installation der Media Pro-Software nötig, was dem allgemeinen Publikum nicht zumutbar ist.

Darüber hinaus ist eine Speicherung und Verwaltung der Daten ausschließlich mit Media Pro nicht zukunftsträchtig, da die Unterstützung von Media Pro durch Phase One während der Arbeit am vorliegenden Projekt eingestellt wurde. Zwar konzentriert sich diese Arbeit auf Media Pro, welches nach wie vor in Mueß eingesetzt wird, zukünftige Erweiterungen könnten aber auch eine Anbindung an alternative Softwaresysteme einschließen. Von wesentlicher Bedeutung ist, den Bild- und Metadaten-Bestand in einem unabhängigen, potentiell erweiterbarem System zwischen zu speichern.

Diese Sicherung ist auch aus einem anderen Grund essentiell: Zur Zeit werden alle analogen Aufnahmen, digitalisierten Bilder und Metadaten (d.h. der Server des Digitalisierungsprojektes) in einem einzelnen Gebäude des Freilichtmuseums ohne Plan einer Datensicherung aufbewahrt. Ein Nebeneffekt der vorliegenden Arbeit ist die Übertragung der Daten auf einen Server der Universität Rostock, um das Sicherungskonzept dieser Server auszunutzen.

Das Ziel der Arbeit ist also die Erstellung eines Konzeptes zur Verfügbarmachung digitalisierter Fotografien aus dem Bilderarchiven des Freilichtmuseums Mueß zusammen mit ihren Metadaten. Ein Teil der Metadaten soll durch den Nutzer anpassbar sein, wobei eine Integration dieser Änderungen in das Bildarchiv ermöglicht werden soll. Darüber hinaus dient die Speicherung der digitalen Fotografien und zugeordnete Metadaten auf Servern der Universität Rostock dem Zweck der Datensicherung. Anhand einer prototypischen Implementierung soll die Tragfähigkeit des Konzeptes gezeigt und eine Rückkopplung zur Nutzbarkeit durch die Anwender ermöglicht werden.

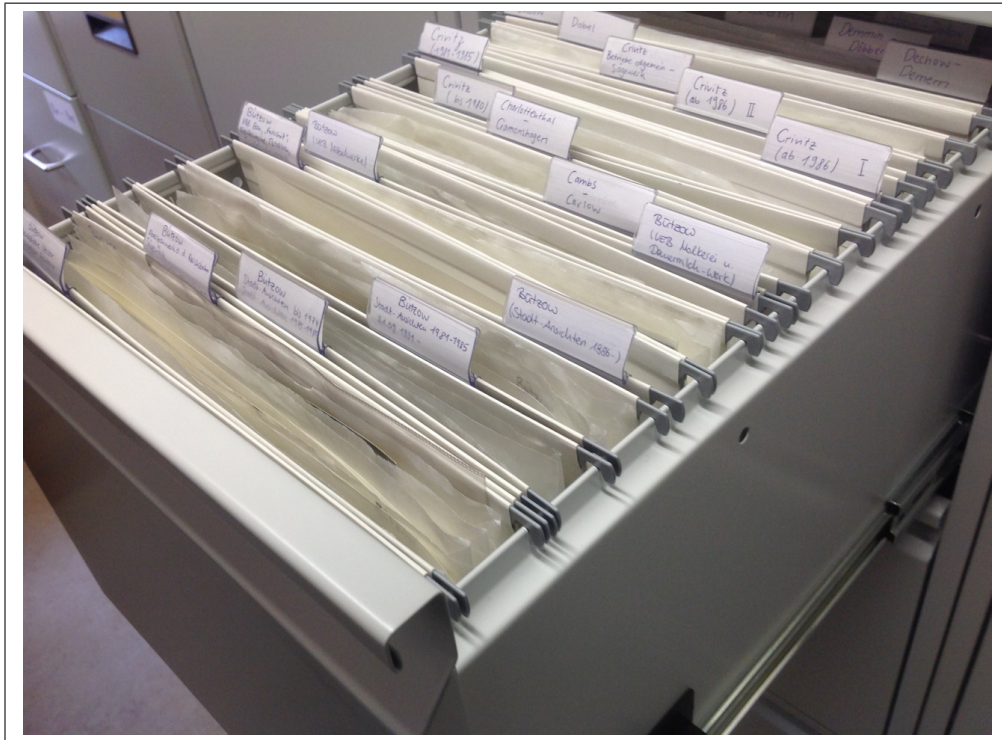
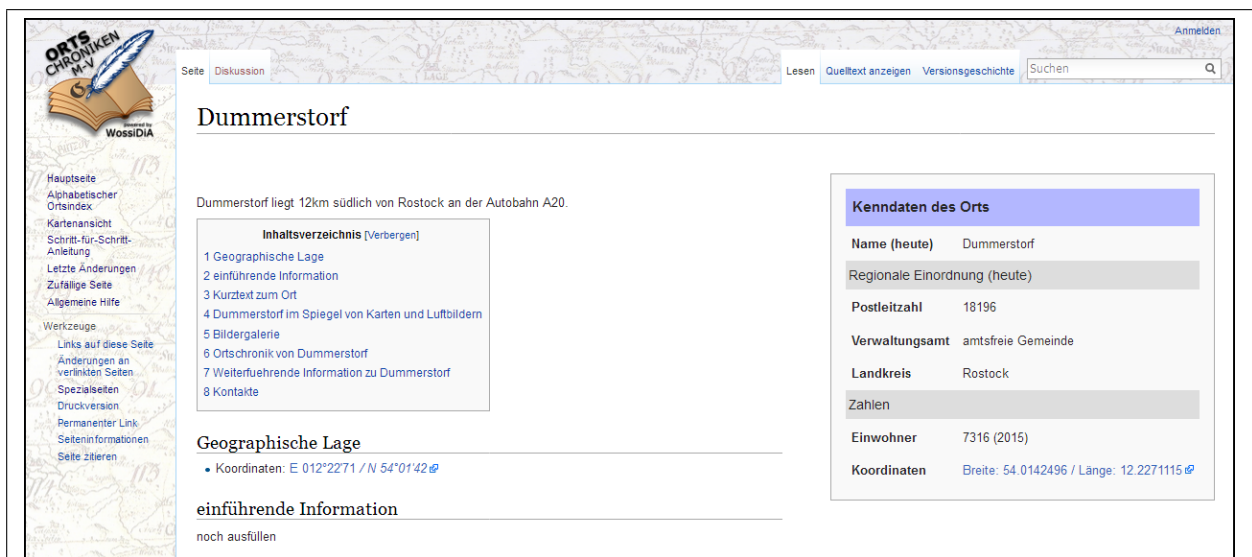


Abbildung 1.1: Analoge Fotografien aus dem Archiv der Schweriner Volkszeitung. Während ein Teil des Bildbestandes wie hier gezeigt nach Städtenamen kategorisiert vorliegt, sind andere Fotografien nach Personennamen, Ländernamen oder einem Thesaurus hierarchischer Schlüsselwörter sortiert.

Kapitel 2

Bedeutung der Citizen Science in den Digital Humanities

In der vorliegenden Arbeit soll die Expertise von Ortschronisten genutzt werden, um fotografische Metadaten des Freilichtmuseum Mueß zu erweitern. Die aktive Beteiligung der Bevölkerung an wissenschaftlichen Projekten ist ein altes Konzept, so wurde etwa die weihnachtliche Vogelzählung der „National Audubon Society“ um 1900 zum ersten Mal durchgeführt und weist heute zehntausende von Teilnehmern auf. Der Begriff „Citizen Science“ – etwa „Bürgerwissenschaft“ – für derartige Projekte wurde jedoch erst in den neunziger Jahren populär. Die Aufgaben der nichtwissenschaftlichen Teilnehmer wurden anspruchsvoller und schlossen zunehmend die Nutzung komplexer Ausrüstung mit ein[Coh08].



The screenshot shows the website 'Ortschroniken M-V' (Ortschroniken Mecklenburg-Vorpommern) with a focus on the entry for 'Dummerstorf'. The page layout includes a navigation menu on the left, a main content area, and a sidebar on the right.

Navigation Menu (Left):

- Hauptseite
- Alphabetischer Ortsindex
- Kartenansicht
- Schritt-für-Schritt-Anleitung
- Letzte Änderungen
- Zufällige Seite
- Allgemeine Hilfe
- Werkzeuge
- Links auf diese Seite
- Änderungen an verlinkten Seiten
- Spezialseiten
- Druckversion
- Permanenter Link
- Seiteninformationen
- Seite zitieren

Main Content Area:

Dummerstorf

Dummerstorf liegt 12km südlich von Rostock an der Autobahn A20.

Inhaltsverzeichnis [Verbergen]

- 1 Geographische Lage
- 2 einführende Information
- 3 Kurztext zum Ort
- 4 Dummerstorf im Spiegel von Karten und Luftbildern
- 5 Bildergalerie
- 6 Ortschronik von Dummerstorf
- 7 Weiterführende Information zu Dummerstorf
- 8 Kontakte

Geographische Lage

- Koordinaten: [E 012°22'71" / N 54°01'42"](#)

einführende Information

noch ausfüllen

Kenndaten des Orts

Name (heute)	Dummerstorf
Regionale Einordnung (heute)	
Postleitzahl	18196
Verwaltungsamt	amtsfreie Gemeinde
Landkreis	Rostock
Zahlen	
Einwohner	7316 (2015)
Koordinaten	Breite: 54.0142496 / Länge: 12.2271115

Abbildung 2.1: Das Ortschronikenportal Mecklenburg-Vorpommern ist ein Beispiel für Citizen-Science in den Digital Humanities.

2.0.1 Gründe für Citizen Science

Zwei Hauptgründe für den Einsatz von Citizen Science in Digital Humanities Projekten sind Geld- und Zeitersparnis. Citizen Science ermöglicht die Beteiligung von wesentlich mehr Personen an einem Projekt, als dies bei ausschließlichem Einsatz wissenschaftlicher Hilfskräften möglich wäre. Dies bedeutet allerdings nicht, dass jedes Citizen Science Projekt alleine mit Haushaltsmitteln bezahlt werden kann, zum Teil sind zusätzliche Sponsoren nötig[WC11].

Für die vorliegende Arbeit ist von besonderem Interesse, dass durch Citizen Science große geographische Gebiete abgedeckt werden können. Ortschronisten wohnen oder arbeiten in der Regel in der Nähe des Ortes, den sie untersuchen, was ihnen erleichtert, lokale Fotografien zu interpretieren oder Informationen von Zeitzeugen einzuholen.

Diese Möglichkeit interessierter Laien, „vor Ort“ und in der Fläche aktiv zu werden, stellt ein Grundbaustein des Beiträger- und Gewährsleute-Prinzip dar, welches für die volkskundlicher Feldforschung von zentraler Bedeutung ist.

Zwar ist davon auszugehen, dass ungeübten Teilnehmern (inklusive ungeübten Wissenschaftlern) mehr Fehler unterlaufen als geübten Wissenschaftlern, diese Situation sollte sich jedoch im Laufe des Projektes verbessern. Der Lernvorgang der Citizen Science Teilnehmer bedarf selbstverständlich der Überwachung durch die wissenschaftlichen Organisatoren, um nachvollziehbare Ergebnisse entstehen zu lassen. An dieser Stelle soll darauf hingewiesen werden, dass manche Citizen Science Projekte die Bildung der Bürger als Haupt- oder zumindest Nebenziel aufweisen (siehe nächster Abschnitt). Ein wesentlicher Aspekt für die vorliegende Arbeit ist die Doppelrolle, die Ortschronisten als Nutzer des Ortschronistenportals und als Teilnehmer am Citizen Science Projekt spielen. Indem das Hinzufügen von Schlüsselwörtern an die Personengruppe übertragen wird, die diese auch zum Suchen der Bilder nutzt, vereinfacht sich das anschließende Finden der Fotografien[Mat06].

Im vorliegenden Text wird häufig von „nicht-Wissenschaftlern“ gesprochen, wenn die Citizen Science Teilnehmer gemeint sind, die nicht zu den wissenschaftlichen Initiatoren des Projektes gehören (wobei die Initiative nicht immer von Wissenschaftlern ausgehen muss, siehe nächster Abschnitt). Dies bedeutet jedoch nicht, dass Wissenschaftler oder Akademiker an der Teilnahme in diesen Rollen ausgeschlossen wären. Auch bedeutet es nicht, dass keine Expertise hinsichtlich der Thematik vorhanden ist. Tatsächlich ist sogar davon auszugehen, dass durch das Einbringen „interessierter Laien“ zusätzliche Fähigkeiten und Kenntnisse für das Projekt nutzbar gemacht werden – seien es fotografische Kenntnisse durch Amateurfotografen oder eine andere Herangehensweise an Rechercheaufgaben durch Journalisten.

2.0.2 Einteilung von Citizen Science Projekten

Citizen Science ist ein weit gefasster Begriff, der je nach Zielstellung unterschiedlich implementiert werden kann. Eine Möglichkeit der Kategorisierung unterscheidet „Contributory“, „Collaborative“ und „Co-Created“ Projekte, fokussiert sich also auf die Rolle der nicht-wissenschaftlichen Teilnehmern im Projekt[BCD⁺09]. Eine andere Form der Klassifizierung bezieht sich auf die jeweiligen Zielstellungen und unterscheidet fünf Projekttypen[WC11]:

1. „Action“
2. „Conservation“

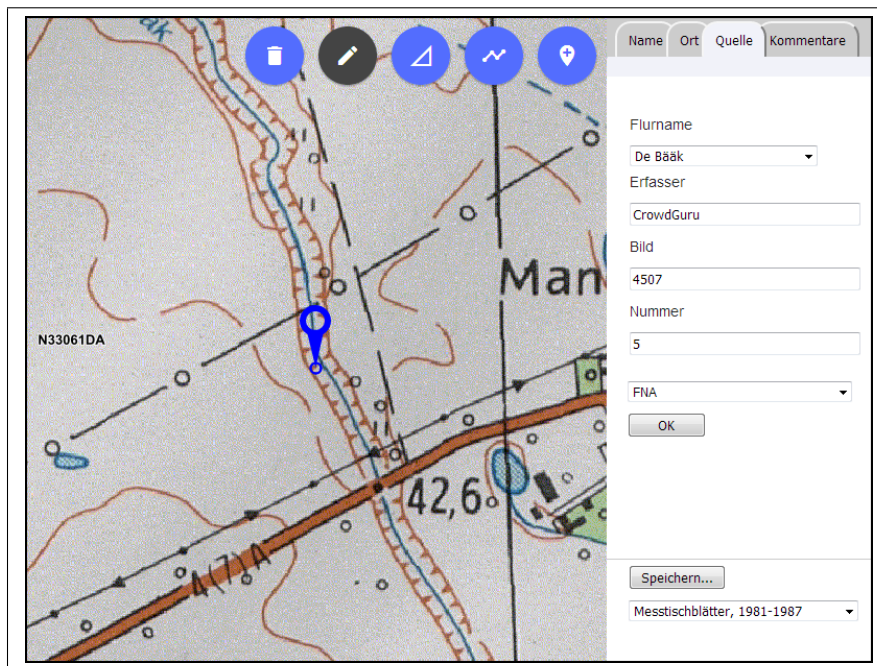


Abbildung 2.2: Historische Flurnamen des Flurnamenarchivs können über eine interaktive Kartenanwendung eingesehen und ergänzt werden. Für die ursprüngliche Transkription der Flurnamen wurde CrowdSourcing eingesetzt.

3. „Investigation“
4. „Virtual“
5. „Education“

„Action“ und „Conservation“-Projekte sind in der Regel naturschutzorientiert. „Action“ Projekte sind „bottom-up“ organisiert, werden also nicht von Wissenschaftlern initiiert sondern beziehen Wissenschaftler als Berater ein. Ziel des „Action“-Projektes ist es nicht, die fachliche Wissensbasis zu erweitern, sondern Hinweise für Umweltgefährdungen zu untersuchen und den Eingriff von Naturschützern zu rechtfertigen.

„Conservation“-Projekte hingegen sind entweder „top-down“ oder „middle-out“ organisiert, d.h. die Initiative geht entweder von Wissenschaftlern oder vom Management aus. Ziel von „Conservation“-Projekten ist es, die Entscheidungsfindung im Ressourcen-Management zu unterstützen (bezogen auf ökologische Fragestellungen). Ein Nebenziel von „Conservation“-Projekten ist es, die Bevölkerung zu bilden und ökologische Probleme ins allgemeine Bewusstsein zu rücken.

„Investigation“-Projekte sind von Wissenschaftlern initiiert und dienen hauptsächlich der wissenschaftlichen Forschung. Sie beruhen auf dem Sammeln von Daten in der „physischen“ (d.h. nicht rein virtuellen) Umgebung, zum Teil durch zehntausende von Teilnehmern. Gelegentlich bieten sie Weiterbildungsmöglichkeiten für die Teilnehmer, dies steht jedoch nicht im Fokus des Projektes. Ein Problem, dem sich „Investigation“-Projekte stellen müssen, ist die möglicherweise ungleichmäßige räumliche Verteilung der Teilnehmer.

„Virtual“-Projekte gleichen „Investigation“-Projekten, allerdings erfolgen alle Projektaktivitäten vermittelt durch Informations- und Kommunikationstechnologie. Die räumliche Verteilung der Projektteilnehmer, die für „Investigation“-Projekte bedeutsam ist, um repräsentative Daten zu generieren, spielt für „Virtual“-Projekte keine Rolle. Ein zentrales Element von „Virtual“-Projekten ist die Replikation der Datengenerierung: Zahlreiche Teilnehmer führen die selbe Aufgabe aus (z.B. Interpretation eines astronomischen Bildes), das vermutlich korrekte Ergebnis wird anschließend mit statistischen Mitteln interpoliert. Zweifelhafte Fälle werden durch Algorithmen erkannt und einem wissenschaftlichen Gutachter vorgelegt.

„Education“-Projekte konzentrieren sich darauf, die Teilnehmer zu unterrichten und ihr kritisches Denkvermögen zu schulen. Wissenschaftliche Erkenntnis steht nicht im Mittelpunkt, zum Teil handelt es sich um Unterrichtsprojekte mit wissenschaftlichem Organisator.

Die vorliegende Arbeit weist einige Eigenschaften eines „Virtual“-Projektes auf, so steht die Gestaltung einer Online-Plattform, welche die virtuelle Infrastruktur des Ortschronikenportals nutzt, im Zentrum des Projektes. Da das Ortschronikenportal eine Wiki ist, können zahlreiche Nutzer die generierten Daten überprüfen. Prinzipiell ist auch die Überwachung dieses Editier- und Revidier-Prozesses mit Hilfe von Algorithmen möglich, die in der vorliegenden Arbeit jedoch nicht spezifiziert worden sind. Obwohl dieser Korrekturprozess ein wünschenswertes Ergebnis wäre, ist es realistischer anzunehmen, dass Ortschronisten sich nicht mit Metadaten befassen, die „ihren“ Ort nicht betreffen. Offenbar ist die räumliche Verteilung der Teilnehmer also wichtig für das Projekt. Ein wesentlicher Vorteil, den Ortschronisten gegenüber statisch in Mueß oder Rostock agierenden Forschern haben, ist ihre räumliche Präsenz am zu untersuchenden Ort. Tatsächlich entspricht die vorliegende Arbeit (wie auch das Ortschronikenportal als Ganzes) also eher der „Investigation“-Kategorie von Citizen Science.

2.0.3 Abgrenzung gegenüber CrowdSourcing

CrowdSourcing beschreibt das Auslagern von Aufgaben an freiwillige Nutzer, die häufig in hierfür konzipierten Portalen angeworben und für ihre Unterstützung bezahlt werden. Die Gründe für CrowdSourcing sind reduzierte Kosten und Zeitersparnisse, was auch eine Basismotivation für CitizenScience bildet. CrowdSourcing grenzt sich gegenüber CitizenScience durch eine weit größere Teilnehmerzahl und eine geringere fachliche Expertise der einzelnen Teilnehmer ab. Allerdings werden Citizen Science Projekte, in denen die Zusammenarbeit ausschließlich über Informations- und Kommunikationstechnologie vermittelt wird, auch als eine Form des CrowdSourcing angesehen[WC11].

Kapitel 3

Stand der Technik und deren Grundlagen

3.1 Online Bildarchive

Es können zwei Typen von Online-Bildarchiven unterschieden werden: Von einer zentralen Stelle – etwa einem Museum – aufgebaute Sammlungen und Webseiten, die zum Teilen der eigenen Fotografien anregen[Mat06]. Bildarchive der ersten Kategorie spezifizieren Metadaten auf Basis eines kontrollierten Vokabulars und mit einer standardisierten Menge von Metadaten-Feldern. Häufig kommen hier Hierarchien von Metadaten zum Einsatz, zum Beispiel hierarchische Schlüsselwörter oder geographische Hierarchien. Obwohl die gewählten Metadatenfelder verschiedener Archive nicht notwendigerweise identisch oder auch nur kompatibel sind, ist häufig eine Zuordnung zu existierenden Metadaten-Standards (zum Beispiel Dublin Core Feldern bei Sammlungen, die das CONTENTdm System verwenden) zu beobachten. Die Katalogisierung erfolgt durch Fachleute und ist zeit- und kostenaufwendig. Webseiten zum Teilen von Bildern hingegen basieren auf dem dezentralen Einfügen von Bildern, die nachträglich durch seitenspezifische Metadaten beschrieben werden können. Diese seitenspezifischen Beschreibungen sind oft nicht mit den zuvor beschriebenen Metadaten-Formaten kompatibel und beruhen auf dem Hinzufügen von „Tags“ zum entsprechenden Bild. Die Aufgabe von Tags besteht nicht nur in der Hinterlegung von Metainformationen, sondern auch in der Navigation zwischen Bildern mit identischen Tags. Im Gegensatz zu formalen Metadaten-Feldern sind Tags häufig nicht auf ein spezifisches Thema beschränkt, so stehen Tags zu Ort, Zeit und dargestellten Personen gleichrangig nebeneinander. Tags entstammen meist keinem beschränkten Vokabular, sodass Beugungsformen des selben Schlüsselwörters in verschiedenen Tags resultieren, gleiches gilt für Singular- und Pluralformen sowie Synonyme. Dies bietet jedoch die Möglichkeit, Tags in verschiedene Sprachen übersetzt einzufügen, was in zentral organisierten online Bildarchiven weit weniger verbreitet ist, obwohl der XMP-Metadatenstandard sprachspezifische Metadatenfelder erlauben würde. Die Nutzung der Bevölkerung beim Klassifizieren von Bildern wird als Social Classification oder auch als Folksonomy bezeichnet, wobei letzterer Begriff (gebildet aus „Folk“ und „Taxonomy“) die Generierung einer Taxonomie, also einer hierarchischen Struktur, impliziert – derartige Hierarchien sind jedoch selten zu beobachten.

Es gibt zahlreiche Online-Archive zum Teilen von Bildern, oftmals mit dem Charakter von So-

zialen Netzwerken, z.B. Flickr, Fotki, Photobucket, Imgur, Instagram und Google Fotos. Im Folgenden soll kurz darauf eingegangen werden, welche Möglichkeiten Flickr und Fotki hinsichtlich des Einsehens und Erstellens von Metadaten bieten.

Flickr ist in der Literatur gut beschrieben und stellt eine häufige Quelle von Bildern und nutzergenerierten Metadaten dar [Mat06, SJ09, LCH09]. Bei Auswahl eines Bildes werden einige grundsätzliche technische Daten (Kameramodel, Blende, Brennweite, Belichtungszeit und Isowert) automatisch angezeigt, wenn diese in den Exif-Daten enthalten sind. Die Gesamtübersicht der Exif-Felder (inklusive der ohnehin angezeigten grundsätzlichen Daten) lässt sich bei Bedarf ebenfalls einblenden. Darüber hinaus wird der Ort der Aufnahme auf einer Karte dargestellt, wenn diesbezügliche Ortsinformationen in den Exif-Daten des Bildes vorhanden sind. Flickr gibt neben dem Namen und der Beschreibung des Bildes das Hochladedatum an. Weitere Datumsangaben (wie das Erstellungsdatum) sind nur über das entsprechende Exif-Feld zugänglich. Flickr vergibt selbstständig Tags für Fotografien, diese weisen einen weißen Hintergrund auf, um sie von den vom Nutzer vergebenen Tags zu unterscheiden. Weder Tags, noch die Gruppen, Galerien und Alben, denen man Bilder hinzufügen kann, sind hierarchisch organisiert.

Wie Flickr erlaubt auch Fotki das Hinzufügen von Tags – hier „Hashtags“ genannt – zu eigenen Bildern. Bezogen auf ein einzelnes Bild sind die Hashtags nicht hierarchisch organisiert, nur manche Bilder mit dem Hashtag „New York“ weisen also auch den Hashtag „City“ auf. Nichtsdestotrotz kann Fotki auch eine hierarchische Gruppierung von Tags durchführen, hierbei werden verbreitete Tags unter bestimmten Oberbegriffen – wie City, Car, People oder Travel – angezeigt. Diese zweistufige Hierarchie ist jedoch nicht zweiseitig navigierbar, d.h. nach Auswählen eines untergeordneten Hashtags kann der Oberbegriff nicht mehr eingesehen werden. Zusätzlich zu den Hashtags kann ein Bildtitel, eine Beschreibung und eine Urheberrechtshinweis angegeben werden. Bei Bedarf werden Exif-Daten in einem Popup-Fenster angezeigt. Hierarchische Informationen können in Form der Albenorganisation hinterlegt werden. Jedes Album kann mehrere untergeordnete Alben oder Fotografien enthalten. Die Anzeige der Albenhierarchie eines Bildes erfolgt in Form des Breadcrumb-Patterns, während die Darstellung der gesamten Albenorganisation eines Nutzers in Form eines Dateibaumes erfolgt. Die Hierarchie ist nutzerspezifisch und nicht standardisiert, ein Album mag den Titel „2018“ tragen und eine mittlere Hierarchieebene einnehmen, während ein zufällig gleichnamiges Album eines anderen Nutzers die höchste Hierarchieebene bildet. Darüber hinaus ist auf diese Weise – im Gegensatz zu hierarchischen Schlüsselwörtern von Metadaten-Formaten – nur eine einzelne Abfolge hierarchischer Beschreibungen pro Bild möglich.

3.2 In der Anwendung eingesetzte Systeme

Die vorliegende Arbeit beschreibt eine Schnittstelle für Metadaten zwischen Phase One Media Pro und dem auf MediaWiki basierenden Ortschronikenportal, zwei Systeme, die im Folgenden vorgestellt werden sollen.

3.2.1 Media Pro SE

Media Pro (Abbildung 3.1) ist eine Katalogisierungssoftware für Bilder und andere Medientypen. Ursprünglich wurde die Software durch iView Multimedia entwickelt und unter dem Namen iView Media Pro vertrieben. Microsoft erwarb die Software und vertrieb sie unter dem Namen

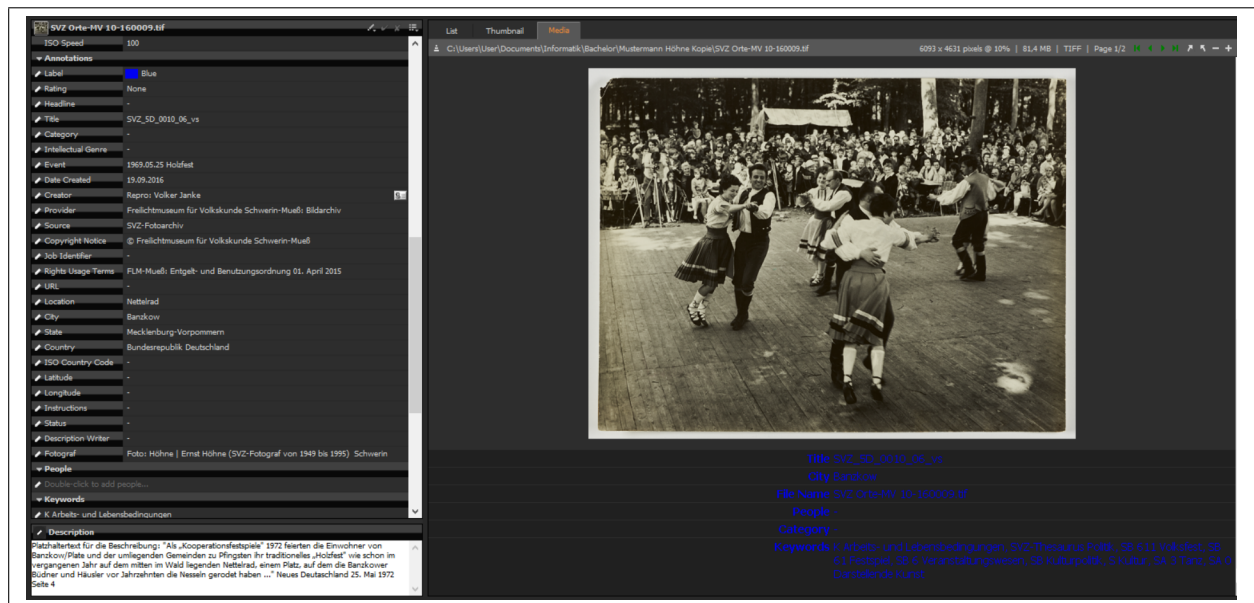


Abbildung 3.1: Benutzeroberfläche von Phase One Media Pro SE. Das „Info Panel“ zeigt Metadaten eines ausgewählten Bildes an, rechts daneben sind die Thumbnails des zugehörigen Kataloges dargestellt.

ExpressionMedia. Nachdem Phase One die Software im Jahr 2010 von Microsoft erworben hatte, wurde sie unter dem neuen Namen Phase One Media Pro SE veröffentlicht. Im Verlauf dieser Arbeit, 2018, gab Phase One das Ende von Media Pro bekannt, die Software wird vorerst jedoch weiter im Freilichtmuseum Mueß genutzt. Ein wesentlicher Vorteil von Phase One Media Pro SE ist die gute Zusammenarbeit mit der Aufnahmesoftware Capture One, die ebenfalls von Phase One veröffentlicht wurde und beim Digitalisierungsprojekt in Mueß eingesetzt wird.

Hat man für eine Auswahl von Bildern einen Media Pro-Katalog erstellt, so speichert Media Pro die Metadaten der Bilder nicht nur in den Bildern selbst ab, sondern auch in einer eigenen Katalogdatei. Katalog und Bilder können in bestimmten Fällen unabhängig voneinander manipuliert werden. Bilder eines Katalogs müssen nicht zwangsläufig in einem Dateiordner abgespeichert sein, da der Dateipfad des Bildes in seinen Metadaten codiert ist. Wenn Bilder von ihrer Position im Dateisystem entfernt werden, wird der Katalog hierdurch nicht geschädigt, obwohl Media Pro eine Warnung ausgibt. Wenn in Abwesenheit des Bildes seine im Katalog abgespeicherten Metadaten manipuliert werden, so können diese mit den im Bild abgespeicherten Metadaten synchronisiert werden.

Die im Katalog abgespeicherten Metadaten können als Textdateien, als XMP-Datei, als XML-Datei oder als html-Datei exportiert werden. Die so erzeugten XML-Dateien können anschließend wieder importiert werden. Eine wesentliche Grundidee der vorliegenden Arbeit beruht darauf, dass Metadaten nicht in der Fotografie oder im Katalog selbst manipuliert werden, sondern dass dies über die Zwischenstufe der XML-Exportdatei geschieht. Anschließend wird die XML-Datei von Media Pro importiert, was eine Synchronisation mit den Metadaten des Media Pro-Katalog und den in der Fotografie abgespeicherten Metadaten zur Folge hat.

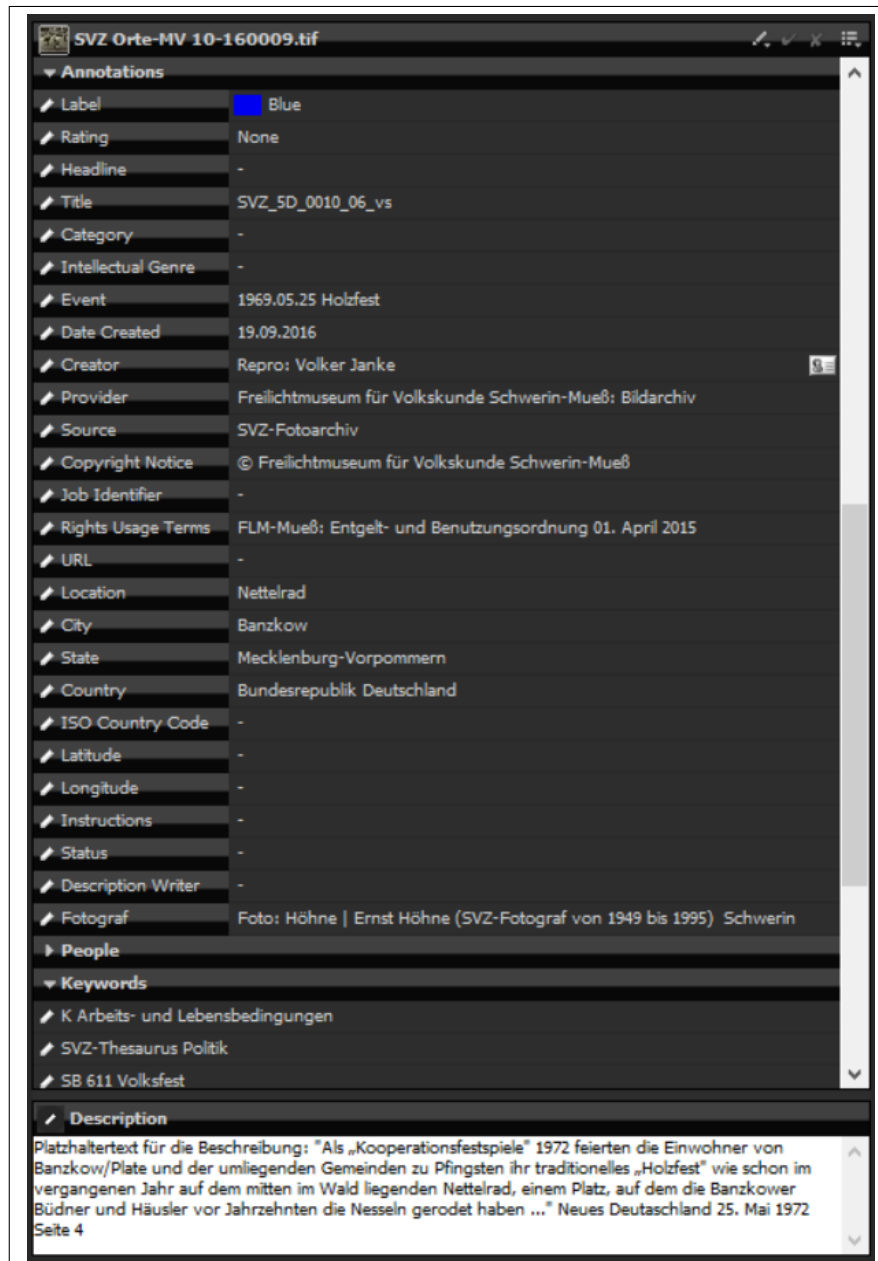


Abbildung 3.2: Detailansicht des „Info Panels“

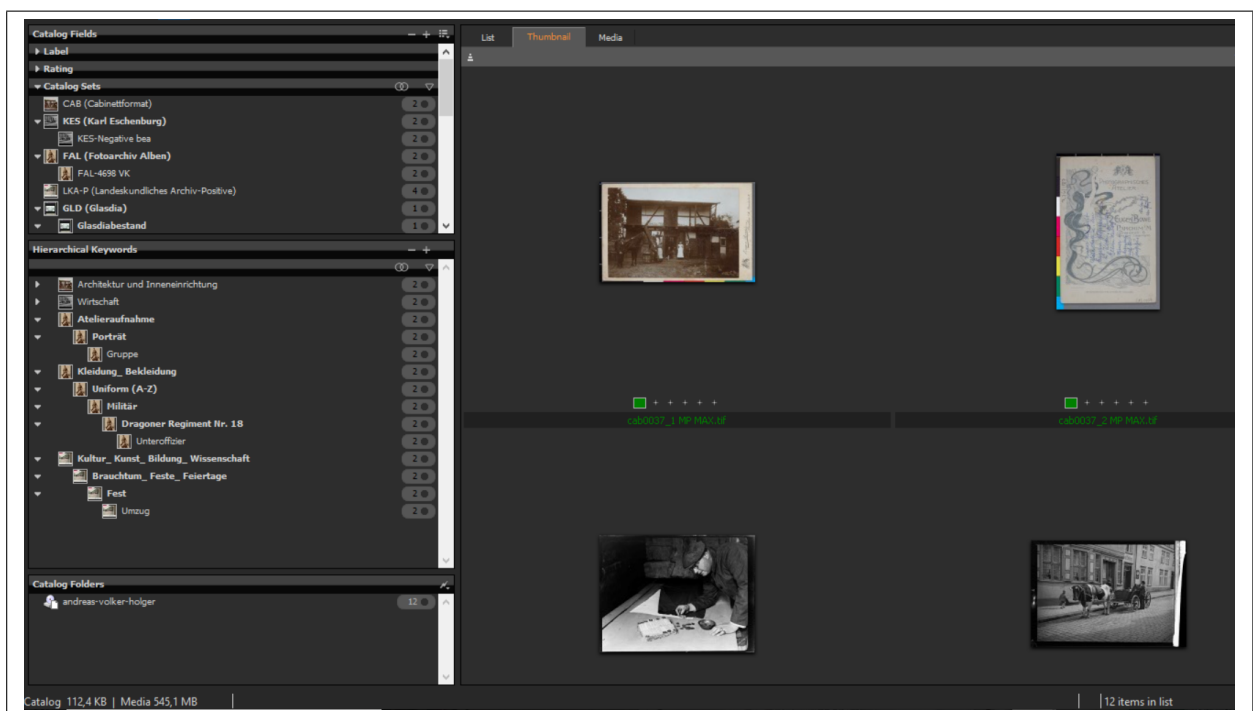


Abbildung 3.3: Metadaten können über das „Organize Panel“ auf Ebene des Katalogs dargestellt werden. Die Darstellung mehrerer Thumbnails anstelle eines hochauflösenden Einzelbildes erfolgt unabhängig hiervon.



Abbildung 3.4: Detailansicht des „Organize Panels“ mit hierarchischen Schlüsselwörtern des Thesaurus Politik, der von der Schweriner Volkszeitung adaptiert worden ist.

3.2.2 MediaWiki

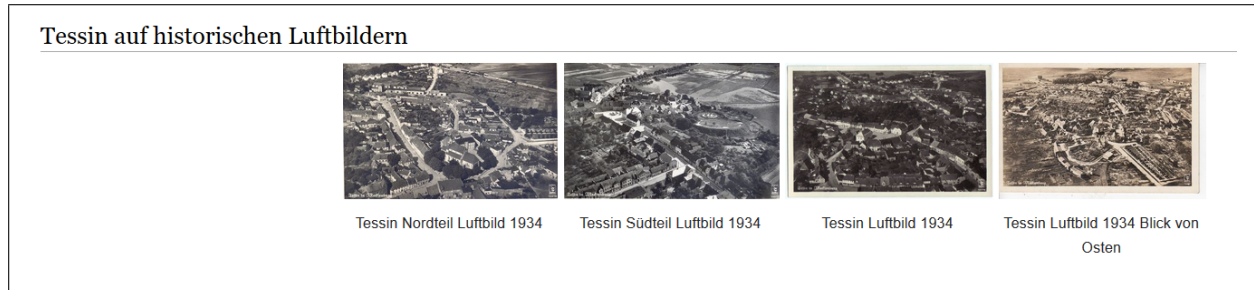


Abbildung 3.5: MediaWiki kann Bilder in Form einer Galerie darstellen. Das Beispiel ist dem Ortschronikenportal entnommen.

Die MediaWiki ist eine in PHP geschriebene Open Source Software für die Verwaltung von Inhalten, die die Form einer Wiki besitzen. Eine Wiki ist aus „Seiten“ zusammen gesetzt, die einen textuellen Inhalt haben, in die aber auch Dateien, z.B. in Form von Bildergalerien, eingebettet werden können (Abbildung 3.5, 3.6).

Zum Speichern der Inhalte nutzt MediaWiki ein Datenbank-Mangagement-System. Kompatibilität ist unter anderen mit MySQL, MariaDB, PostgreSQL, Oracle und SQLite gegeben. Da das Ortschronikenportal Mecklenburg Vorpommern PostgreSQL nutzt, wurde die vorliegende Arbeit auf dieses Datenbank-Mangagement-System hin ausgerichtet.


Seiten der Wiki können in verschiedenen Namensräumen stehen, welche durch Doppelpunkte vom nachfolgenden Seitennamen abgetrennt werden. Ein wichtiger Namensraum ist „Vorlage“ (bzw. „Template“, Namensräume werden von der deutsche MediaWiki automatisch übersetzt). Ist eine Vorlage definiert – in der vorliegenden Arbeit etwa die Vorlage für Metadaten-Infoboxen (Abbildung 3.7) – kann diese durch Angabe der Vorlage in jede Seite eingefügt werden. Der Vorlage können Parameter übergeben werden, deren Wert an der spezifizierten Stelle der Vorlage angezeigt wird.

Seiten jedes Namensraumes können durch Hinzufügen des Kürzels „[[Kategorie:Kategorienname]]“ (oder „[[Category:Kategorienname]]“) der Kategorie „Kategorienname“ zugeordnet werden (Abbildung 3.8). Ähnlich der im Kapitel „Online Bildarchiv“ beschriebenen Tags dienen MediaWiki-Kategorien der Klassifizierung, der Seitenbeschreibung und der Navigation zwischen Seiten der gleichen Kategorie. Da sowohl Dateien als auch Kategorien stets eine Seite des entsprechenden Namensraumes zugeordnet ist, ist eine hierarchische Kategorisierung von Bildern möglich. Intern sind die komplexen Kategoriebeziehungen – die durchaus zyklisch sein können – in einer einzigen Tabelle, „Categorylinks“, abgespeichert.

MediaWiki lässt sich beliebig durch sogenannte Extension erweitern und anpassen. In der vorliegenden Arbeit geschieht dies zum Beispiel durch die „ParserFunction“-Extension, welche eine Fallunterscheidung bei der Darstellung des Infobox-Templates realisiert – ein Metadaten-Feld der Infobox soll nämlich nur dargestellt werden, wenn der entsprechende Metadaten-Wert bekannt ist.

Bilder in MediaWiki gehören dem Namensraum „Datei“ an. Einige ihrer Metadaten – inklusive einiger Exif-Daten und z.B. Schlüsselwörtern, falls diese vorhanden sind – können bei Bedarf eingblendet werden (Abbildung 3.8). Zunächst ist nur eine Auswahl der Metadaten sichtbar.

Datei 1



SVZ Orte-MV 10-160009.jpg
640 × 486, 74 KB

Quelldatei: SVZ Orte-MV 10-160009.jpg

Maximale Dateigröße: 2 MB

Erlaubte Dateitypen: png, gif, jpg, jpeg, webp.

Zielname:

Beschreibung/Quelle:

Lizenz: ▾

Diese Datei beobachten

Warnungen ignorieren

Datei 2

Quelldatei: Keine Datei ausgewählt.

Abbildung 3.6: Nach Auswahl der gewünschten Dateien können diese hochgeladen und somit in das MediaWiki-System integriert werden. Hochgeladene Dateien erhalten jeweils ihre eigene MediaWiki-Seite.

Metadaten des Bildes	
Name	SVZ Orte-MV 10-160009.JPG
Annotationen	
Farbcode	3
Bewertung	0
Produkt	SVZ_5D_0010_06_vs
Ereignis	1969.05.25 Holzfest
Datum des Ereignisses	2016:09:19
Beschreibung	"Als „Kooperationsfestspiele" 1972 feierten die Einwohner von Banzkow/Plate und der umliegenden Gemeinden zu Pfingsten ihr traditionelles „Holzfest" wie schon im vergangenen Jahr auf dem mitten im Wald liegenden Nettelrad, einem Platz, auf dem die Banzkower Büdner und Häusler vor Jahrzehnten die Nesseln gerodet haben ..." Neues Deutaschland 25. Mai 1972 Seite 4
Schlüsselwörter	K Arbeits- und Lebensbedingungen; SVZ-Thesaurus Politik; SB 611 Volksfest; SB 61 Festspiel; SB 6 Veranstaltungswesen; SB Kulturpolitik; S Kultur; SA 3 Tanz; SA 0 Darstellende Kunst
Fotograf	Foto: Höhne
Autor	Repro: Volker Janke

Abbildung 3.7: Die Metadaten-Infobox wird mit Hilfe eines MediaWiki-Templates generiert.

Hersteller	HP
Modell	HP Scanjet 3800
Erfassungszeitpunkt	15:46, 23. Sep. 2016
Horizontale Auflösung	200 dpi
Vertikale Auflösung	200 dpi
Y und C Positionierung	Benachbart
Schwarz/Weiß-Referenzpunkte	<ul style="list-style-type: none"> • 0 • 255 • 128 • 255 • 128 • 255
Exif-Version	2.2
Bedeutung einzelner Komponenten	<ol style="list-style-type: none"> 1. Y 2. Cb 3. Cr 4. Existiert nicht
unterstützte Flashpix-Version	0.100
Farbraum	sRGB
Sättigung	Gering
Schärfe	Gering
Erweiterte Details ausblenden	

Abbildung 3.9: MediaWiki liest einige, hauptsächlich technische, Metadaten aus der Bilddatei aus und stellt sie tabellarisch dar.

3.3 Bilder und Metadaten-Formate

Es gibt eine unüberschaubare Anzahl von Metadatenstandards. Die Standards lassen sich auf verschiedene Weise charakterisieren, unter anderem anhand der Zielgruppen (z.B. Museen oder Archive), den Arten der beschriebenen Daten (z.B. visuelle Ressourcen oder kulturelle Objekte), der Funktion des Metadatenformats (z.B. kontrollierte Vokabulare oder Markup-Sprachen) und der Kategorien der Metadaten, die mit dem Format beschrieben werden sollen[Ril18]. Zu Metadatenkategorien zählen:

- Daten: Standard beschreibt Ressource, möglicherweise zusammen mit Markupinformationen und den Metadaten
- Metadata Wrapper: Format umschließt mehrere Unterformate, möglicherweise inklusive der beschriebenen Ressource
- Deskriptive Metadaten: Metadaten dienen dem Entdecken, Verstehen und Interpretieren der Ressource (in der folgenden Arbeit meist „Annotationen“ genannt)
- Bewahrende Metadaten: Informationen, die nötig sind, die Ressource auch in Zukunft auslesen und nutzen zu können
- Rechtslage: Metadaten, welche Zugriffsberechtigungen und Beschränkungen beschreiben
- Strukturelle Metadaten: Angabe anderer Versionen oder anderer Elemente der Ressource
- Technische Metadaten: Digitale und physische Eigenschaften einer Ressource

Im Folgenden gehen wir auf drei Metadatenstandards näher ein, die bei der Beschreibung von Fotografien Verwendung finden. Zusammen decken sie (mit Einschränkungen) alle genannten Metadaten-Kategorien ab, wobei in den folgenden Kapiteln der Fokus auf den beschreibenden Metadaten liegen wird.

3.3.1 Exif

Das Exif-Format wird durch die Japan Electronics and Information Technology industries Association (JEITA) und die Camera and Imaging Products Association (CIPA) reguliert. Es ermöglicht, technische Informationen in Form von TIFF 6.0-Tags in der Datei des Bildes zu speichern. Abgesehen von TIFF sind Exif-Header auch mit anderen Bildformaten, wie JPEG oder PSD kompatibel. Eine wesentliche Verwendung ist die Speicherung von Kamera- und Blitzlichteinstellungen durch die Kamera selbst bei der Erstellung des Bildes, aber auch andere Geräte (z.B. Scanner) verwenden Exif für die Speicherung technischer Daten. Das Exif-Format ist dazu ausgelegt, im Verlauf des Bildbearbeitungsprozesses von unterschiedlicher Stelle ausgelesen und angepasst zu werden. Als primär binäres Format ist Exif besser als XMP dafür geeignet, Thumbnails von Bildern als Teil der Bildmetadaten selbst abspeichern (Media Pro speichert Thumbnails intern als Teil des Media Pro-Katalogs ab, kann sie jedoch auch exportieren und enthält dann im XML-ExportPfadangaben zu den Thumbnails). Der Exif-Standard erlaubt sowohl komprimierte als auch unkomprimierte Thumbnail und sieht keine Größenbeschränkungen für diese vor – allerdings kann einem komprimiertes Bild kein komprimiertes Thumbnail zugewiesen werden[CJ16].

In der Bilddatei sind die Exif-Header in einem eigenen Image File Directory (IFD), dem EXIF IFD, gespeichert. Zusätzlich zum Exif IFD spezifiziert der Exif-Standard noch den GPS IFD für GPS-Daten und den Interoperabilitäts IFD.

Insgesamt sind in der Exif IFD 77 und in der GPS IFD 32 verschiedene Exif Header definiert, hinzu kommen 30 TIFF Tags, die außerhalb der genannten IFDs abgespeichert sind. In diesen zusätzlichen Tags ist beispielsweise Datum und Zeit der letzten Dateiänderung abgespeichert, während im Exif IFD sowohl der Zeitpunkt der ursprünglichen Bildentstehung als auch der Zeitpunkt der Digitalisierung enthalten sind. Im GPS IFD ist eine vierte Datumsangabe abgespeichert, allerdings folgt diese einem anderen Format und verwendet für Datum und Uhrzeit zwei verschiedene Header. Eine Exif Headerangabe besteht, dem TIFF Standard folgend, aus mindestens vier Elementen.

Der Header beginnt mit der zwei Byte lange Codennummer des jeweiligen Tags. In weiteren zwei Bytes ist der Datentyp des Tags codiert. Vier Bytes legen die Anzahl der Werte fest (z.B. die Anzahl Zeichen, wenn der Datentyp ASCII ist). Wenn der Wert des Headers kürzer als vier Bytes ist, steht er direkt hinter der Längenangabe, wenn nicht, ist in den verbliebenen vier Byte die Position des Wertes codiert. In allen drei zuvor genannten Headern mit kombinierten Datums- und Zeitangaben ist als Datentyp ASCII und als Anzahl zwanzig angegeben. Das zu verwendende Datumsformat ist jeweils YYYY:MM:DD_HH:MM:SS, wobei Doppelpunkte, Leerzeichen und der abschließende Null-Character mitgezählt werden, um auf eine Anzahl von zwanzig ASCII-Zeichen zu kommen.

3.3.2 IPTC IIM

Das International Press Telecommunication Council (IPTC) wurde 1965 gegründet und stellt ein internationales Konsortium der Hauptnachrichtenagenturen, Nachrichten-Verlagen und der Nachrichten-Händler dar. Der IPTC IIM (Information Interchange Model) Standard wurde 1990 entwickelt und definiert ein Hüllformat für die Übertragung von Fotografien und Textdokumenten. Er ist wohl wohl der einflussreichste, wenn auch nicht der einzige der vom IPTC geschaffenen Standards – andere Standards sind beispielsweise IPTC 7901 für den Austausch von Texten und NewsCodes, eine Taxonomie von Medien-Kategorien. Die tatsächlich für Bilder verwendeten IPTC-Header stellen nur eine Untermenge der im IPTC IIM definierten Metadaten-Felder dar [Int18].

Der IPTC IIM-Standard war für eine einzelne Sprache konzipiert (z.B. Englisch) und beschränkte die Zeichenanzahl pro Metadatenfeld. Heute hat IPTC IIM vor allem historische Bedeutung, die IPTC selbst rät zur Nutzung der Nachfolgestandards[Gro10]. Die Weiterentwicklungen von IIM, IPTC Core (seit 2004) und IPTC Extension (seit 2007) sind als Namensräume innerhalb von XMP konzipiert. Sie verfügen daher über die Möglichkeiten der Sprachspezifizierung über das XML-Language-Attribut. Auch die Beschränkung der Feldlänge ist nur noch Implementationsabhängig (siehe Beispiel Media Pro).

Das IIM besteht aus neun „Records“, die fünf Schichten zugeordnet sind: zur ersten Schicht gehört ein Object Envelope Record, der zweiten Schicht zählen fünf Application Records, in der dritten Schicht befindet sich ein Pre-ObjectData Descriptor Record, in der vierten Schicht ist der ObjectData Record zu finden (in unserem Fall die Fotografie selbst) und zur fünften Schicht zählt ein Post-ObjectData Descriptor Record[IN14]. Jeder Record besteht aus einem oder mehreren DataSets, welche wiederum aus einem Tag und den eigentlichem Wert des Metadaten-Feldes, dem

Data Field, bestehen. Ein Tag besteht aus dem Tag-Marker (8 bit), der Nummer des Records (8 bit), der Nummer des Datasets (8 bit) und entweder der Länge des Data Fields oder – im Fall eines zu langen Datafields – der Länge des nachfolgenden Feldes, welches die Länge des Data Fields angibt (in beiden Fällen 16 bit).

IPTC Core sieht, anders als gerade für IPTC IIM beschrieben, eine Aufteilung der Metadatenfelder in vier Abschnitte vor: IPTC Contact, IPTC Content, IPTC Image und IPTC Status. Sowohl die Angabe des Datums (DateCreated) als auch die Angabe der Ortsinformationen (Location, City, State, Country) erfolgt im IPTC Image Abschnitt, während zum Beispiel Schlüsselwörter Teil des IPTC Content sind. Im IPTC IIM waren alle diese Angaben im zweiten Record, einem der Application Records, enthalten.

Wie Metadaten im Bild selbst gespeichert werden, ist im Standard des jeweiligen Bildformats (also jpg, tiff, png ...) festgelegt und nicht Teil des IPTC-Standards – werden die Metadaten jedoch außerhalb des Bildes gespeichert, legt IPTC IIM ein binäres Format fest[Cou18].

3.3.3 XMP

XMP (Adobe's Extensible Metadata Platform) wurde 2002 von Adobe spezifiziert und heute durch die ISO kontrolliert. XMP verwendet das Datenmodell des W3C Resource Description Frameworks (RDF). Da XMP auf XML basiert (und als XML-Code serialisiert werden kann), bietet es hervorragende – wenn auch aufwendige – Möglichkeiten, mehrschichtige hierarchische Strukturen aufzubauen, wie im Kapitel „Hierarchische Schlüsselwörter“ demonstriert werden soll. Eine Instanz des XMP-Datenmodells wird Paket genannt[Inc12]. Ein Paket beschreibt eine Resource, zum Beispiel eine Bilddatei oder eine Fotografie, die in ein pdf-Dokument eingebettet ist. Ein Paket weist höchstens eine Eigenschaft (d.h. Metadatenfeld) eines bestimmten Namens auf. Metadaten mit mehreren Werten – z.B. Schlüsselwörter – werden als Array definiert. Diese Behandlung von Schlüsselwörtern unterscheidet sich sowohl vom IPTC IIM Standard, welcher für jedes Schlüsselwort ein unabhängiges Dataset mit identischer DataSet-Nummer zuweist (25), als auch vom Media Pro-XML-Export, welcher für jedes Schlüsselwort ein eigenes XML-Tag des Typs „Keyword“ vorsieht.

Tatsächlich stellt XMP, im Gegensatz zu Exif und IPTC IIM, weniger eine eigenständige Sammlung standardisierter Metadatenfelder dar, sondern erlaubt vielmehr die Koexistenz mehrerer derartiger Sammlungen in jeweils unterschiedlichen Ästen der XML Baumstruktur. Eine derartige Metadatensammlung bildet einen Namensraum. Neben dem bereits erwähnten IPTC Core-Namensraum (Iptc4XMPCore) nennt der XMP-Standard unter anderem die Namensräume Adobe PDF (pdf), Photoshop (photoshop), Camera Raw (crs) und Dublin Core (dc)[Inc16].

Wie zuvor angedeutet, ist es prinzipiell möglich, Thumbnails im XMP-Format einzubinden, hierbei wird das base64-Format verwendet. Informationen über das Thumbnail wie auch das Thumbnail selbst sind im Namensraum XMPGImg zu finden. Auch die Erweiterung eines Namensraumes um zusätzliche Felder sowie das Einführen neuer Namensräume wird vom XMP-Standard unterstützt.

3.3.4 Zusammenfassende Bewertung der Formate Exif, IPTC und XMP

Die zu Beginn des Kapitels genannten Kategorien der Metadaten werden durch die Formate Exif, IPTC IIM und XMP im Wesentlichen abgedeckt. Exif setzt hierbei einen Fokus auf technische

– und damit auch erhaltende – Metadaten, wohingegen bei der Konzeption von IPTC IIM bzw. der Nachfolgestandards IPTC Core und IPTC Extension besonderen Wert auf beschreibende Metadaten gelegt wurde. XMP ist ein Beispiel für einen Metadata Wrapper und schließt somit eine Vielzahl von Metadatenkategorien ein, die beliebig erweiterbar ist.

Während die Beschreibung technischer Metadaten durch Exif klar festgelegt ist, kann die Vielfaltigkeit der Namensräume von XMP zu einer gewissen Willkür der Zuordnung beschreibender Metadaten führen. In den XMP-Exportdatei von Media Pro SE wird dies besonders deutlich: Obwohl eine Zuordnung der Annotationen zu IPTC Core-Feldern definiert ist, werden die Metadaten in mehreren Namensräumen redundant abgespeichert.

3.4 Bewertung der Metadatenverwaltung in Media Pro

Abhängig von der Darstellungsform – XML-Export, XMP-Export, Skript-Interface oder GUI der Software – klassifiziert Media Pro die Metadaten in unterschiedlicher Weise.

Media Pro ermöglicht die Erstellung einer XML-Datei, welche alle Metadaten (inklusive Pfadangaben von Originalfotos und Thumbnail-Fotos, aber ohne Bitinformationen der tatsächlichen Fotos) eines Bildkatalogs enthält. Die Datei basiert auf der Katalogdatei (d.h. der Filialdatei-Speicherung der Metadaten) und nicht auf den im Foto selbst abgespeicherten Metadaten. Der XMP-Export, ebenfalls eine XML-Datei, bezieht sich im Gegensatz hierzu nur auf eine einzelne Fotografie. Wie im folgenden Kapitel gezeigt wird, hat dieser Unterschied insbesondere Konsequenzen auf die Speicherung von katalogübergreifend gültigen Daten, wie der Keyword Hierarchie. Während die Struktur von XMP tief verschachtelt ist, und die historische Entwicklung von XMP berücksichtigt indem beispielsweise die verschiedenen Namensräume von XMP (z.B. Dublin Core, Photoshop ...) in unterschiedlichen Ästen des XML-Dokuments abgespeichert sind, haben einzelne Media Items des XML-Exports nur eine Klassifizierungsebene oberhalb der tatsächlichen Metadaten. In dieser Ebene sind die Metadaten in Asset Properties, Media Properties, Annotation Fields, User Fields und Meta Data Fields unterteilt. Zusätzlich befindet sich auf dieser Ebene der Hierarchie der Pfad des Thumbnails (wenn beim Export die Möglichkeit genutzt wurde, Thumbnails zu exportieren).

Zu den Asset Properties gehören Eigenschaften, die für den Dateizugriff von Bedeutung sind, wie Dateiname, Dateipfad und mehrere Datumsangaben, wie das Erstellungsdatum und das letzte Änderungsdatum der Datei. Interessanterweise sind auch „Label“ und „Rating“ Asset Properties. „Label“ bietet dem Nutzer die Möglichkeit, dem Bild einen Zahlenwert (der in der Benutzeroberfläche als Farbwert visualisiert wird) zuzuweisen. In der vorliegenden Arbeit wird das Label verwendet, um den Verarbeitungsstatus des Bildes zu kennzeichnen, wobei Bilder, die vollständig mit Metadaten versehen wurden und zur Veröffentlichung bereit sind das Label „0“ erhalten. „Rating“ bezieht sich auf eine Bewertung des Bildes mit einer Zahl zwischen 0 und 5, welche als Anzahl von Sterne visualisiert wird. Diese Bewertung ist normiert und z.B. Teil des XMP-Schemas.

Media Properties sind technische Eigenschaften, die zur Darstellung des Bildes nötig sind, wie die Dimensionen, Auflösung, Farbraum und Komprimierung des Bildes. Zusätzliche technische Bildeigenschaften sind in der Kategorie „Meta Data Fields“ zu finden, allerdings handelt es sich hierbei vor allem um Geräteeinstellungen, wie die verwendeten Blitzlichteinstellungen. Die technischen Eigenschaften korrespondieren zum größten Teil mit den im Exif-Format codierten

Metadaten, und tatsächlich ist die verwendete Exif-Version eine Eigenschaften der Kategorie „Meta Data Fields“. Von größerer Bedeutung für die vorliegende Arbeit war die Kategorie „Annotation Fields“. (Nicht-hierarchische) Schlüsselwörter, die Spezifizierung des Ortes – mit den Hierarchieebenen „Location“, „City“, „State“ und „Country“ – sowie Angaben zu den abgebildeten Personen sind ebenso Elemente dieser Kategorie wie Angaben zu Copyright und Rechteinhabern. Obwohl XML ausgezeichnet für die Beschreibung von hierarchischen Strukturen geeignet ist – wie im nächsten Kapitel am Beispiel der hierarchischen Schlüsselwörter gezeigt werden soll – sind für einzelne Bilder keine derartigen hierarchischen XML-Strukturen realisiert, die Angaben zum Abbildungsort sind durch vier unabhängige Tags realisiert, obwohl die Visualisierung dieser Informationen in der Benutzeroberfläche von Media Pro als Baumstruktur erfolgt. Trotz der fehlenden strukturellen Representation in XML liegen (zumindest in der vorliegenden Bildersammlung) funktionale Abhängigkeiten zwischen „City“, „State“ und „Country“ vor. In früheren Versionen von Media Pro – z.B. in iView Media Pro 2.6 – beruhten die Metadaten der Kategorie „Annotation Fields“ auf den im IPTC IIM Schema definierten Eigenschaften. Das Handbuch von iView Media Pro nannte ausdrücklich die Größenbeschränkungen der einzelnen Felder, die auf IPTC IIM beruhte (wobei die umfangreichsten Felder, wie „Description“, bis zu 2000 Zeichen enthalten konnten)[iML]. In der aktuellen Version von Media Pro, die in der vorliegenden Arbeit Verwendung fand (Version 2.3.0.266), beruhen die „Annotation Fields“ hingegen auf den IPTC Core-Eigenschaften, die ein Namensraum von XMP darstellen (Abbildung 3.1). Obwohl XMP keine Größenbeschränkungen vorschreibt, lässt Media Pro in der getesteten Implementation für keines der Felder Einträge von mehr als 4096 ($= 2^{12}$) Zeichen zu (für Feldern mit mehreren Einträgen, wie Schlüsselwörter und Personen, kann jeder einzelne Eintrag 4096 Zeichen lang sein). Die erwähnte Abbildung von „Annotation Fields“ auf IPTC Core-Eigenschaften wird nicht für das Erstellen der XMP-Exportdatei berücksichtigt, tatsächlich findet der Namensraum Iptc4XMPCore nur für ausgewählte Metadaten (z.B. Location) Verwendung.

Obwohl der XMP-Standard (einschließlich des darin enthaltenen IPTC Core-Standards) eine komplexe Beschreibung einer Fotografie zulässt, die beispielsweise zwischen Aufnahmeort und dargestelltem Ort sowie Aufnahmedatum und Digitalisierungsdatum unterscheidet (siehe Empfehlungen im nächsten Kapitel), ist die Realisierung durch Media Pro wesentlich einschränkender. Um diese Beschränkungen zu umgehen, ist die Erstellung zusätzlicher „User Field“ möglich. Im XML-Export heißen diese nutzerdefinierten Felder `User_Field<X>` (wobei `<X>` für eine fortlaufende Nummer steht), die Visualisierung von Media Pro hingegen zeigt den vom Nutzer gewählten Namen an. Im Fall der Fotosammlung Mueß wurde das nutzerdefinierte Feld „Fotograf“ erstellt, welches Namen des Fotografen sowie Orts- und Datumsangaben enthält[JH16]. Für bessere Kompatibilität mit anderen Katalogisierungsprogrammen könnte diese Zeichenkette geparkt und aufgeteilt werden, für die vorliegende Arbeit wurde dies jedoch nicht weiter berücksichtigt und Strings mit komplexen Informationen als atomare Feldwerte betrachtet.

3.5 Empfehlungen der Metadata Working Group

Die Metadata Working Group wurden 2007 gegründet und bestanden ursprünglich aus den Unternehmen Apple, Adobe, Canon, Microsoft und Nokia; 2008 trat Sony der Working Group als sechstes Mitglied bei. Hauptziel der Working Group ist die Bewahrung und Interoperabilität von digitalen Bildmetadaten wodurch die Verfügbarkeit von Metadaten für alle Anwendungen, Geräte

Media Pro	IPTC Core
Label	Label
Rating	Rating
Title	Headline
Product	Title
Genre	Category
Intellectual Genre	Intellectual Genre
Event	Event
Event Date	Date Created
Author	Creator
Credit	Provider
Source	Source
Copyright	Copyright
Transmission	Job Identifier
Rights Usage Terms	Rights Usage Terms
URL	URL
Location	Location
City	City
State	State/Province
Country	Country
ISO Country Code	ISO Country Code
Instructions	Instructions
Status	Status
Writer	Description Writer
People *	People *
Categories *	Supplemental Categories *
Keywords *	Keyword *
Scenes *	Scenes *
Subject Codes *	Subject Codes *
Caption	Description

Tabelle 3.1: Abbildung der Media Pro Annotationen auf Felder des IPTC Core-Standards. Mit einem Stern (*) markierte Felder können mehrere Einträge aufnehmen.

und Dienste ermöglicht werden soll. Ein zentrales Thema der Working Group ist die Gestaltung eines auf Metadaten basierenden Workflows. Die an Metadaten interessierten Stakeholder werden hierfür in drei verschiedene Gruppen eingeteilt – Creator, Changer und Consumer, wobei nutzergruppenspezifische Empfehlungen für die Behandlung von Metadaten ausgesprochen werden. Tatsächlich ist es nicht im Interesse der Working Group, neue Metadatenformate zu erstellen, sondern durch Empfehlungen das Arbeiten mit bereits existierenden Formaten zu erleichtern – allerdings decken die Empfehlungen keinesfalls alle existierenden Formate ab, sondern beschränken sich auf die drei Formate, die auch in dieser Arbeit angesprochen wurden (Exif, IPTC IIM/Core und XMP). Die Richtlinien[Gro10] sind auf der Webseite der Metadata Working Group hinterlegt. Sie sind keinesfalls als allgemein akzeptierter Standard anzusehen. Wie im nächsten Abschnitt deutlich wird, weicht beispielsweise die Implementierung hierarchischer Schlüsselwörter durch Media Pro deutlich von ihnen ab. Ein Problem bezüglich der Empfehlungen könnte ihre fehlende Aktualität sein – die letzte Version der Guidelines For Handling Image Metadata erschien 2010, der Einfluss neuerer Entwicklungen des Gebiets wie die Festlegung von Teilen des XMP-Formats als ISO-Standard ISO 16684-1 im Jahr 2012 bleiben folglich unberücksichtigt. Beim Vergleich der drei Metadatenformate Exif, IPTC IIM und XMP fand die Metadata Working Group nur geringe Überlappungen zwischen Exif und den anderen beiden Formaten – einzig die Exif Felder Date/Time, Description, Copyright und Creator sind gesondert zu betrachten: Da ähnliche Felder auch in XMP vorhanden sind, schlägt die Working Group zum einen ein direktes Mapping vor (z.B. wird Copyright auf das XMP-Feld dc:rights abgebildet), zum anderen wäre es möglich, Exif als zusätzlichen Namensraum in XMP zu integrieren (Copyright würde in diesem Fall auf das neu einzuführende XMP-Feld exif:copyright gemappt). In beiden Fällen wären also die Exif-Informationen in das XMP-Format integriert. Im Fall widersprüchlicher Informationen empfiehlt die Metadata Working Group, die Exif Metadaten zu bevorzugen. Bei näherer Betrachtung der existierenden Exif-Header fällt auf, dass die Metadata Working Group anscheinend mehrere Exif-Header zusammengefasst betrachtet – wie bereits zuvor erwähnt, unterscheidet Exif zwischen drei verschiedenen Datumsangaben. Auch die Bildbeschreibung könnte in zwei verschiedenen Exif-Feldern gespeichert werden, dem ImageDescription Header und dem UserComment Header. Im Gegensatz zu Exif teilen sich IPTC IIM und XMP zahlreiche Felder. Durch die Einführung von IPTC Core als XMP Namensraum existiert jedoch bereits ein standardisiertes Mapping zwischen IPTC und einem Namensraum in XMP. Die Metadata Working Group empfiehlt, im Zweifelsfall die XMP Information (d.h. die Metadaten des IPTC Core Namensraumes) gegenüber den IPTC IIM Daten zu bevorzugen. Diese Empfehlung entspricht der oben genannten Einschätzung des IPTC selbst bezüglich der Weiterverwendung von IPTC IIM. Aus den obigen Beispielen wird deutlich, dass nicht nur Richtlinien über Mapping und Bevorzugung von Metadatenfeldern verschiedener Metadatenformate existieren müssen, sondern auch innerhalb von XMP selbst eine klare Aufgabentrennung der verschiedenen Namensräume nötig ist. Dies wurde zumindest teilweise von der Metadata Working Group in weiteren Empfehlungen vorgenommen.

Wie im Abschnitt über Exif bereits angedeutet, können für eine Fotografie eine ganze Reihe von Datumsangaben eine Rolle spielen – etwa die Zeit des abgebildeten Motivs, die Zeit der ursprünglichen Fotografie, die Zeit der Restauration, die Zeit der Reproduktion (bzw. Zeitangaben der Reproduktionen) sowie Datumsangaben bezüglich Bearbeitung und Metadatenerfassung des Bildes. Die Metadata Working Group empfiehlt, für das Erstellungsdatum des intellektuellen Inhalts (also der Fotografie, nicht das Datum des Motivs) das Feld photoshop:datecreated und für

das Datum der Digitalisierung das Feld XMP:createdate zu verwenden – zwei annähernd gleichnamige Tags aus verschiedenen Namensräumen. Für den Zeitpunkt der letzten Veränderung soll das Feld XMP:ModifyDate dienen. Das hier dargestellte Beispiel der Datumsangaben soll die Kombination von Metadaten verschiedener XMP:Namensräume illustrieren, tatsächlich ist eine Abspeicherung mehrerer Datumsangaben auch mit IPTC IIM möglich.

Obwohl also Empfehlungen bezüglich drei verschiedener Datumsangaben vorliegen, heißt dies nicht, dass diese auch tatsächlich in dieser Form in Fotoarchivierungssoftware implementiert sind – und schon gar nicht, dass drei Datumsangaben für alle Nutzer ausreichend sind. Wie zuvor erwähnt unterscheidet Media Pro Datumsangaben in mindestens drei verschiedenen Kategorien – auf Ebene des Dateisystems Erstellungs- und Änderungsdatum der Datei, auf Ebene der technischen Bilddaten die von der Kamera abgelegten Exif-Datumsangabe, vom Nutzer frei gestaltete Annotationen (im vorliegenden Fall „Event“ für das Datum des Ereignisses) und eventuell vom Nutzer definierte zusätzliche Annotationsfelder. Eine standardisierten Nutzung dieser Felder durch den Anwender, welcher versucht, Datumsangaben semantisch zu unterscheiden, ist auf diese Weise nicht vollständig unterstützt. Im vorliegenden Fall des Fotoarchivs Mueß wurde im zusätzlichen Metadatenfeld „Fotograf“ Datumsinformation bezüglich des Fotografen innerhalb einer Zeichenkette codiert.

3.6 Hierarchische Schlüsselwörter

Im Gegensatz zur „sozialen Klassifikation“ (der Erstellung einer Metadatataxonomie durch den Nutzer) sind wissenschaftliche Klassifizierungssysteme (verwendet z.B. von Museen und Journalisten) häufig tief hierarchisch gegliedert: Auch wenn zwei Fotografien durch unterschiedliche Schlüsselwörter beschrieben wurden, kann durch Analyse der (standardisierte) Struktur der Schlüsselwörterhierarchie eine semantische Ähnlichkeit der Bilder abgeleitet werden, was insbesondere die Suche nach Fotografien mit bestimmten Motiven erleichtert. Eine der Problemstellungen der vorliegenden Arbeit war es, eine solche Schlüsselwörterhierarchie in den Ansatz der „sozialen Klassifikation“ zu integrieren – Nutzer sollen die (standardisierte und daher unveränderliche) Hierarchie für Navigationszwecke verwenden, und gleichzeitig andere Felder der Metadaten überarbeiten können. Im Folgenden soll daher auf verschiedene Möglichkeiten eingegangen werden, hierarchische Schlüsselwörter in den Metadaten zu codieren, wobei die Empfehlungen der Metadata Working Group bezüglich XMP mit der Realisierung in Media Pro verglichen wird.

Die Metadata Working Group empfiehlt aus Kompatibilitätsgründen, flache und hierarchische Schlüsselwörter zusammen im rdf:Description Feld abzuspeichern (hierfür werden die hierarchischen Schlüsselwörter in „flache“ Schlüsselwörter umgewandelt, d.h. sie werden redundant abgespeichert, Abbildung 3.10). Während die „flachen“ Schlüsselwörter dem Dublin Core-Feld „Subject“ zugeordnet werden, wird für „hierarchische“ Schlüsselwörter das Tag mwg-kw:Keywords vorgeschlagen (d.h. die Metadata Working Group leitet aus ihren Initialen einen eigenen XMP-Namensraum ab, mwg-kw). Für die Definition eines Keywords sind bis zu drei XML-Hierarchieebenen nötig: rdf:li, mwg-kw:Children (auf gleicher Ebene mwg-kw:Keyword für den eigentlichen Namen des Schlüsselwörters) und rdf:bag. Darüber hinaus sind drei Hierarchieebenen nötig, um innerhalb von rdf:Description den Wurzelknoten der Hierarchie zu definieren: Das bereits genannte <mwg-kw:Keywords>, <mwg-kw:Hierarchy> und <rdf:Bag>.

Im Gegensatz hierzu erfolgt die Implementation hierarchischer Schlüsselwörter im XMP-Export

```

1 <rdf:RDF XMLns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
2   <rdf:Description XMLns:dc="http://purl.org/dc/elements/1.1/">
3     <dc:subject>
4       <!-- flat keywords -->
5     </dc:subject>
6   </rdf:Description>
7
8   <rdf:Description
9     XMLns:mwg-kw="http://www.metadataworkinggroup.com/schemas/keywords/">
10     <mwg-kw:Keywords rdf:parseType="Resource">
11       <mwg-kw:Hierarchy>
12         <rdf:Bag>
13           <rdf:li rdf:parseType="Resource">
14             <mwg-kw:Keyword>SVZ-Thesaurus Politik</mwg-kw:Keyword>
15             <mwg-kw:Applied>False</mwg-kw:Applied>
16             <mwg-kw:Children>
17               <rdf:Bag>
18                 <rdf:li rdf:parseType="Resource">
19                   <mwg-kw:Keyword>K Arbeits- und
20                     Lebensbedingungen</mwg-kw:Keyword>
21                   <mwg-kw:Applied>True</mwg-kw:Applied>
22                 </rdf:li>
23               </rdf:Bag>
24             </mwg-kw:Children>
25           </rdf:li>
26         </rdf:Bag>
27       </mwg-kw:Hierarchy>
28     </mwg-kw:Keywords>
29   </rdf:Description>
30 </rdf:RDF>

```

Abbildung 3.10: Empfehlung der Metadata Working Group für die Codierung hierarchischer Schlüsselwörter im serialisierten XMP-Format.


```

1 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
2   <rdf:Description rdf:about=""
3     xmlns:lightroom="http://ns.adobe.com/lightroom/1.0/"
4     <lightroom:hierarchicalSubject>
5       <rdf:Bag>
6         <rdf:li>SVZ-Thesaurus Politik</rdf:li>
7         <rdf:li>SVZ-Thesaurus Politik|K Arbeits- und
8           Lebensbedingungen</rdf:li>
9       </rdf:Bag>
10    </lightroom:hierarchicalSubject>
11  </rdf:Description>
12 </rdf:RDF>

```

Abbildung 3.11: Die Organisation hierarchischer Schlüsselwörter in XMP-Exportdateien von Media Pro weicht deutlich von den Empfehlungen der Metadata Working Group ab.

von Media Pro in nur wenigen XML-Ebenen: Innerhalb von `rdf:Description` stehen nur die Tags `<lightroom:hierarchicalSubject>` und `<rdf:Bag>` oberhalb der einzelnen Schlüsselwörter (Abbildung 3.11). Die Schlüsselwothierarchie wird hierbei nicht auf die XML-Hierarchie abgebildet, stattdessen befinden alle Elemente auf einer Ebene. Hierarchische Strukturen sind textuell definiert, indem dem Namen des Schlüsselwörters alle Vorgängerknoten vorangestellt sind (durch „|“ abgetrennt). In der vorliegenden Arbeit wurde – zusätzlich zu der hierarchischen Abbildung auf die MediaWiki-Kategorien – auch eine flache Abbildung auf einen Textstring (der in einer Infobox dargestellt werden kann) implementiert.

```

1 <SetList>
2   <Set>
3     <SetName>@KeywordsSet</SetName>
4     <Set>
5       <SetName>SVZ-Thesaurus Politik</SetName>
6       <UniqueID>1</UniqueID>
7       <UniqueID>2</UniqueID>
8     <Set>
9       <SetName>K Arbeits- und Lebensbedingungen</SetName>
10      <UniqueID>1</UniqueID>
11      <UniqueID>2</UniqueID>
12    </Set>
13  </Set>
14 </SetList>

```

Abbildung 3.12: Die XML-Exportdatei von Media Pro speichert hierarchische Metadaten getrennt von den individuellen Metadaten der Fotografien ab. Dem einzelnen hierarchischen Schlüsselwort werden die „UniqueIDs“ der zugehörigen Bilder zugeordnet.

Die Implementation hierarchischer Schlüsselwörter im XML-Export nutzt im Gegensatz hierzu

die XML-Hierarchie aus und verwendet eine Ebene pro Keyword weniger als die Metadata Working Group für XMP empfiehlt (Abbildung 3.12). Ein Keyword besteht aus den Ebenen `<set>` und `<setname>`, wobei die `<set>`-Tags der untergeordneten Schlüsselwörter auf gleicher Ebene wie `<setname>` stehen. Da für der XML-Export für einen Katalog von Fotos gilt, wird die Schlüsselwörterhierarchie nicht beim jeweiligen Foto eingetragen, stattdessen wird die Hierarchie am Ende des XML-Dokuments definiert. Die ids der einzelnen Fotos werden – invers zu den Empfehlungen für XMP – unter dem Tag `<UniqueID>` auf gleicher Ebene wie `<setname>` eingeordnet. Die Hierarchie der Keywords ist hierbei nur eine der möglichen Hierarchien, die auf diese Weise definiert werden können (andere betreffen z.B. Kategorien). Vom Freilichtmuseum Mueß werden zwei verschiedene Systeme hierarchischer Schlüsselwörter verwendet. In der vorliegenden Arbeit beschränkt sich jedoch auf eines der beiden Hierarchien – das Wurzelement trägt den Setname „@KeywordsSet“, das bisher außen vor gelassene System hingegen verwendet ein Wurzelement mit Setname „SVZ-Fotoarchiv“.

Kapitel 4

Ein System zur Kopplung von Bildarchiv und Online-Portal

Im Folgenden wird auf Basis der vorangehenden Analyse der eingesetzten Werkzeuge und Metadatenformate ein Vorschlag zur Kopplung des Bildarchives und des Chronistenportals vorgestellt. Ziele dabei sind die Verfügbarmachung digitalisierter Fotografien zusammen mit den zugeordneten Metadaten und die Möglichkeit, diese Metadaten durch den Anwender erweitern zu lassen. Gegebenenfalls sollen diese Änderungen in das Bildarchiv integriert werden. Des Weiteren soll die Speicherung der Metadaten und digitalisierten Fotografien auf Servern der Universität Rostock eine Sicherung der Daten erlauben.

4.1 Komponentenarchitektur der Software

Die Software besteht aus zwei Hauptkomponenten: Importer und Exporter (Abbildung 4.1). Die Aktivitäten dieser Komponenten sind im Kapitel „Workflow“ näher beschrieben. Eine weitere Komponente, die am Importprozess teilnimmt, dient der Erstellung einer Karte, in der alle Orte mit assoziierten Fotografien markiert sind. Die Kartendarstellung selbst basiert auf der JavaScript-Bibliothek OpenLayers. Die Karten-Komponente vervollständigt ein JavaScript-Template mit aktuellen Informationen der Datenbank und generiert Bildergalerie-Seiten für jeden Ort. Da jedes Bild einer Kategorie zugewiesen wird, die dem assoziierten Ort entspricht, kann eine Ortsgalerie in Form einer Kategoriengalerie realisiert werden. Derartige Galerien, die alle Bilder einer Kategorie umfassen, können mit geeigneten MediaWiki Extensions erzeugt werden. Alle drei genannten Hauptkomponenten benötigen Zugriff auf die Datenbank des Ortschronikenportals. Dies betrifft erstens Lese- und Schreibzugriff auf die neu zu erstellenden Metadaten-Tabellen, zweitens das Auslesen des Seiteninhalts von MediaWiki-Seiten und drittens die Ableiten der MediaWiki Kategorien-Hierarchie. Die hierfür benötigte Datenbank-Manager-Komponente besteht intern aus Treiberbibliotheken, und einer Adapterkomponente, welche konkreten SQL-Code an die API der Datenbank übermittelt und das Ergebnis der Anfrage in ein einfacher handhabbares Format überträgt.

Während das Auslesen des Seiteninhalts und das Ermitteln der Kategoriebeziehungen durch Datenbankabfragen erfolgen können, ist es deutlich komplexer, Änderungen an den Standardtabellen von MediaWiki auf dieser Abstraktionsebene vorzunehmen. Wir verwenden daher an dieser Stelle

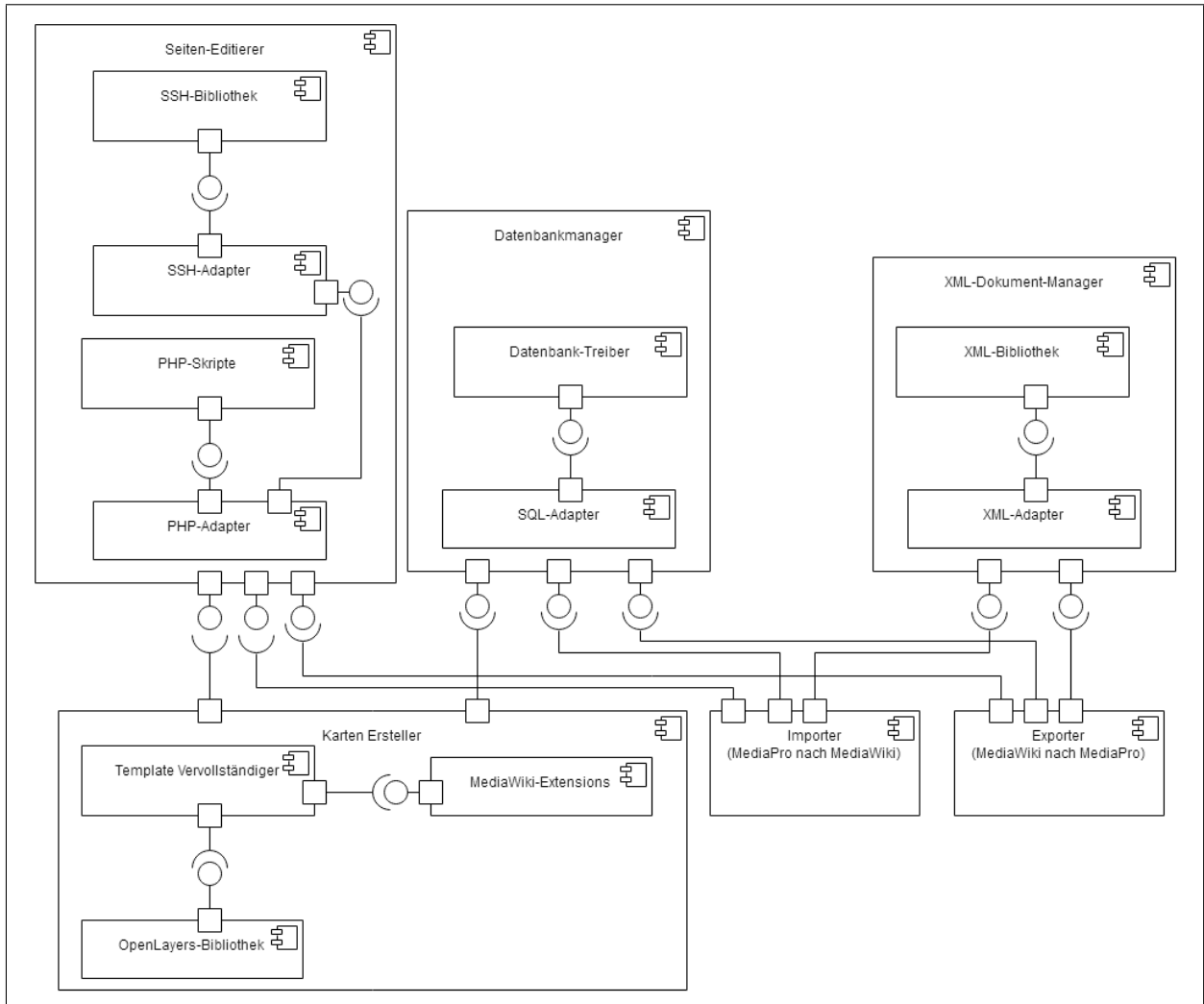


Abbildung 4.1: Architektur der Softwarekomponenten

Maintenance-Skripte, d.h. am Standardbetrieb von MediaWiki beteiligte PHP-Skripte. Auf diese Weise werden nicht nur alle beteiligten Tabellen angepasst, sondern auch Konflikte mit anderen im Hintergrund arbeitenden Maintenance-Skripten von MediaWiki vermieden. Falls die Software auf dem Server des Ortschronikenportals installiert wird, ist ein direkter Zugriff auf die dort vorhandenen PHP-Skripte möglich. Falls der Zugriff jedoch von einem anderen Computer heraus erfolgen soll, ist erst eine Verbindung zum Server herzustellen, zum Beispiel über SSH. Hierfür wird zum einen eine geeignete SSH-Bibliothek, und zum anderen eine Adapterkomponente für diese SSH-Bibliothek benötigt.

Sowohl Importer als auch Exporter benötigen Zugriff auf das von Media Pro generierte XML-Export-Dokument. Ersterer liest die Metadaten der einzelnen Bilder aus, um sie in Datenbank zu schreiben, während letzterer Metadaten bei Bedarf ändert und Daten unveränderter Bilder aus dem Dokument löscht. Die hierfür benötigte XML-Management-Komponente besteht zum einen aus einer XML-Bibliothek, zum anderen aus einem Adapter für diese Bibliothek, welche spezifisch auf die DTD des XML-Export-Dokumentes abgestimmt ist.

4.2 Workflow

Für die vorliegende Arbeit wurden drei verschiedene Nutzer bzw. Nutzergruppen definiert: Vertreter des Freilichtmuseums Mueß (der Archivar), Vertreter des Ortschronikenportals und die Nutzer des Ortschronikenportals, im folgenden „Ortschronisten“ genannt (Abbildung 4.2). Tatsächlich handelt es sich um eine diverse Gruppe von Personen mit Interesse am Fachgebiet, die zum Teil ihr Wissen für ein sekundäres Publikum aufbereitet darstellen (z.B. Journalisten, Historiker, Fotografen ...). Neben Nutzern, die sich selbst als Ortschronisten verstehen, sind die aufbereiteten Bildmetadaten auch für Genealogen von Interesse. Der Workflow der vorliegenden Arbeit ist in drei Grobphasen unterteilt: Vorbereitung, Import und Export. Die Vorbereitungsphase besteht aus einer Aktivität – dem Erstellen der nötigen Datenbanktabellen für Metadaten. Die Importphase beginnt damit, dass der Vertreter Mueß den XML-Exportbefehl der Media Pro Software nutzt, um die Metadaten von fertig bearbeiteten Bildern zu exportieren. Zusammen mit der XML-Datei werden auch Thumbnail-Bilder und Bilder in Originalauflösung exportiert. Bereite Bilder sind (gemäß des internen Workflows des Museums Mueß) mit einem speziellen Label gekennzeichnet. Thumbnailbilder tragen den selben Namen wie die Originalbilder, verwenden jedoch das jpeg Format statt des tiff-Formats. In den exportierten Metadaten ist der Name der Thumbnails nicht explizit genannt, kann jedoch aus den Informationen des „Filename“-Feldes ermittelt werden. Wie die Bilder in der Praxis an den Vertreter des Ortschronikenportals übermittelt werden – z.B über Dropbox – ist in dieser Arbeit nicht festgelegt, da der Fokus auf den Metadaten liegt. Für die XML-Datei bietet sich ein Kopieren in eine festgelegte Seite des Ortschronikenportals an, es kann aber auch der selbe Übertragungsmodus verwendet werden, der für die Fotos verwendet wird. Im folgenden gehen wir davon aus, dass die XML-Daten als Dateien vorliegen, nicht als MediaWiki-Seiten. Diese Dateien werden für den Re-Export der Bildmetadaten benötigt, weshalb der Dateiname der verantwortlichen XML-Datei in der Datenbanktabelle für Metadaten festgehalten ist.

Die XML-Daten werden geparkt und in die Tabellen für Metadaten eingefügt, gegebenenfalls werden die vorhandenen Daten aktualisiert. Der Vertreter des Ortschronikenportals lädt Thumbnails und Originalbilder über eine MediaWiki-Extension hoch. Eine zusätzliche (bereits installierte)

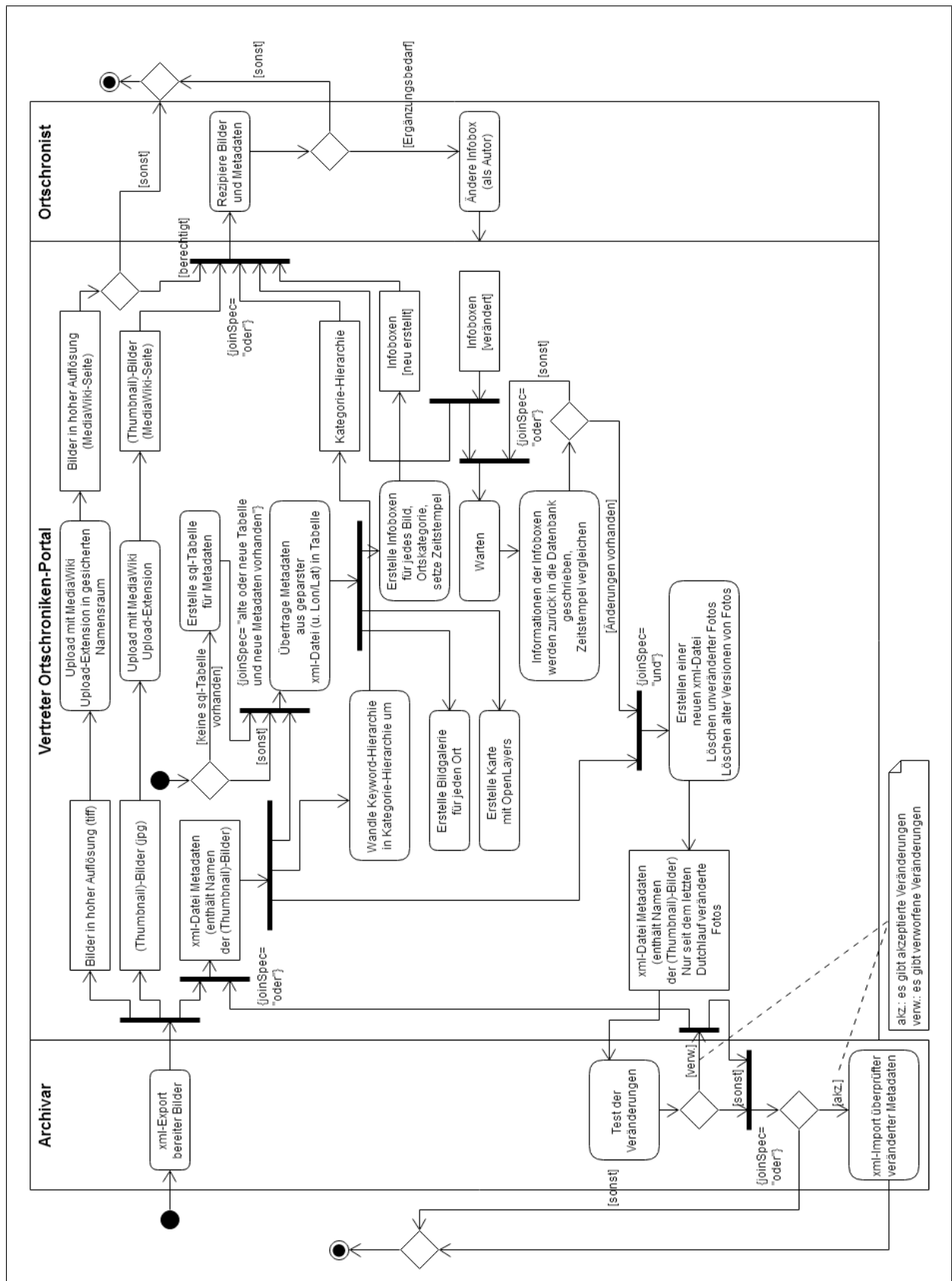


Abbildung 4.2: Darstellung des Arbeitsablaufes in Form eines Aktivitätsdiagrammes.

Sicherheitsextension schützt den Namensraum der Originalbildern vor unerlaubtem Zugriff. Im weiteren Workflow werden die Bilddaten selbst nicht weiter bearbeitet, allerdings erstellt MediaWiki automatisch eine Seite für jede hochgeladene Datei. Von vornherein zeigt diese Seite einige Metadaten des Bildes an, allerdings lassen sich diese Informationen schlecht manipulieren, weshalb wir zusätzliche Infoboxen für Metadaten des XML-Export-Dokuments erstellen. Das Erstellen dieser Infoboxen und die Einbindung der Datei-Seiten in MediaWiki-Kategorien sind die beiden abschließenden Schritte des Bildimports. Das hierarchische Kategoriensystem von MediaWiki soll das hierarchische Schlüsselwortsystem von Media Pro repräsentieren. Zusätzlich erhält jedes Bild die Kategorie des dem Bildes zugeordneten Ortes. Hierdurch wird es später möglich, Bildergalerien für jeden Ort zu generieren.

Zwischen der Import- und der Exportphase werden die Bilder und Metadaten von Ortschronisten eingesehen. Ein voraussichtlich großer Anteil der Nutzer (insbesondere nicht angemeldete Nutzer des Portals) werden es beim Perzipieren der Media Pro-Thumbnail und Metadaten belassen. Bilder in Originalauflösung können nur von Benutzern mit besonderen Rechten angesehen werden, doch das Rechtemanagement des Ortschronikenportals und die hierfür genutzten MediaWiki-Extensions sind nicht Teil dieser Arbeit. Im Gegensatz hierzu von besonderem Interesse sind Ortschronisten, die Metadaten ergänzen bzw. korrigieren und Ortschronisten, welche diese Ergänzungen überarbeiten. Aufgrund der Möglichkeit, gemachte Veränderungen wieder zu revidieren, ist es nicht trivial zu bestimmen, wann ein zeitweiser stabiler Zustand erreicht ist oder ob überhaupt Veränderungen gegenüber dem ursprünglichen XML-Export vorgenommen wurden.

Entweder in regelmäßigen Zeitabschnitten – möglicherweise gesteuert durch ein Skript – oder nur bei auffälligen Änderungen der Infoboxen wird die Exportphase eingeleitet. Zunächst werden die Daten aus den Infoboxen aller Bilder extrahiert und zurück in die Datenbanktabelle geschrieben. Indem der Datenbankeintrag auf Veränderungen überprüft wird, kann getestet werden, ob ein Export dieses spezifischen Bildes nötig ist. Ein veränderter Metadatensatz wird durch ein Ändern des Labels markiert. Darüber hinaus werden die zu exportierenden Bilder gezählt: Falls keine Bilder exportiert werden müssen, wird der Vorgang abgebrochen. Die hierarchischen Schlüsselwörter werden nicht berücksichtigt, da sie als wenig geeignet für Verbesserungen durch „soziale Klassifizierung“ eingeschätzt wurden.

Aus den so markierten Metadatensätzen der Datenbank werden XML-Dokumente gewonnen. Hierbei werden nicht etwas XML-Dokumente nur auf Basis der Datenbankeinträge aufgebaut (hierfür während zusätzliche Daten, z.B. Exif-Daten, in der Datenbanktabelle nötig), sondern vielmehr werden die ursprünglichen XML-Export-Dateien der jeweiligen Bilder angepasst. Alle derartigen Dateien (bzw. MediaWiki-Seiten mit zugehörigem XML-Code) müssen also bewahrt werden, die Dateinamen (bzw. Seitennamen) sind, wie zuvor angesprochen, in der Tabelle der Metadaten für jedes Bild abgespeichert. Nicht veränderte Bilder werden aus dem XML-Dokument entfernt.

Das Ergebnis dies Re-Exports sind entweder eine Reihe von XML-Dateien (eine für jeden beteiligten ursprünglichen Import) oder eine spezielle Seite des Ortschronikenportals, die den Code aller erzeugten XML-Dokumente enthält. In beiden Fällen ist es auch möglich, die erzeugten XML-Dokumente vor ihrer Übertragung zu einem einzigen Dokument zu vereinen. Die Frage, welche Methode der Übertragung letztlich gewählt wird, hängt von Usability-Überlegungen ab. Nachdem die XML-Dokumente nach Mueß übermittelt wurden (bzw. nachdem sie im Ortschronikenportal hochgeladen und in Mueß eine Reihe von XML-Dateien erstellt wurden), werden sie von

Media Pro als Quelle für den Import verwendet. Diese Arbeit ist zumindest teilweise manuell zu erledigen. Innerhalb von Media Pro können die veränderten Datensätze entweder akzeptiert und der ursprüngliche Media Pro-Katalog aktualisiert werden, oder die Änderungen werden verworfen bzw. angepasst. Aus letzteren Datensätzen wird ein neues XML-Export-Dokument erzeugt, welches durch das Ortschronikenportal erneut importiert werden kann (in diesem Fall aber ohne beiliegende Thumbnail-Bilder).

4.3 Datenbankarchitektur

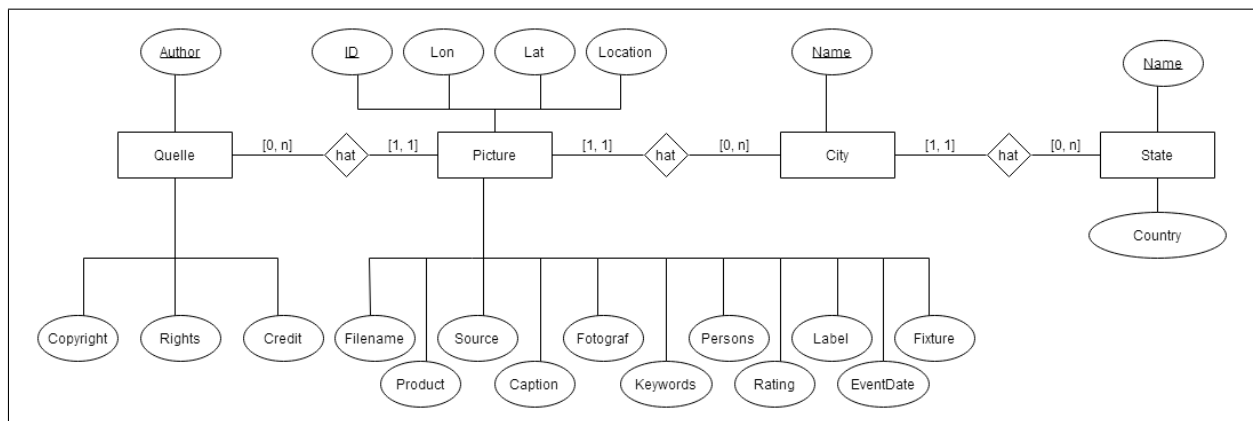


Abbildung 4.3: Entity-Relationship-Diagramm der verwendeten Metadaten.

MediaWiki speichert Metadaten in Form eines einzigen Datenbankeintrags des Typs Medium Blob ab, der einen serialisierten PHP-Array repräsentiert. Da dies für die vorliegende Arbeit nicht ausreichend ist, werden die aus dem geparsten XML-Export-Dokument gewonnenen Metadaten stattdessen in mehreren hierfür generierten Tabellen gesichert, welche beim Export mit den Metadateneinträgen der einzelnen Bild-Infofoxen synchronisiert werden und schließlich als Grundlage eines neuen XML-Export-Dokuments dienen.

Abhängigkeiten zwischen den einzelnen Metadaten-Feldern werden von Media Pro nicht erzwungen und lassen sich auch nicht aus der Tagstruktur des XML-Export-Dokumentes ableiten. Semantisch lassen sich zwei Gruppen von Abhängigkeiten identifizieren: Die hierarchische Angabe des Ortes und quellspezifische Metadaten.

Media Pro folgt dem IPTC Core-Standard indem es vier geographische Hierarchieebenen unterscheidet: „Location“, „City“ (im Kontext des Ortschronikenportals ein Dorf oder eine Ortschaft), „State“ (im vorliegenden Projekt als Bundesland interpretiert) und „Country“. Während eine „Location“, also eine Landmarke innerhalb einer Stadt, nicht notwendigerweise einen einzigartigen Namen tragen muss, können wir für die anderen Hierarchieebenen funktionale Abhängigkeiten voraussetzen: Eine (eindeutig benannte) Stadt befindet sich in einem Bundesland, und ein Bundesland befindet sich in einem Land. Tatsächlich tragen Städte nicht notwendigerweise eindeutige Namen, da im Metadatensatz jedoch zusätzliche identifizierende Daten (wie die Angabe einer Städte-ID oder einer Region der Stadt) fehlen, muss der Wert des Feldes „City“ einzigartig sein. Dies zu gewährleisten ist jedoch Aufgabe des Archivars und sprengt den Rahmen dieser Arbeit.

Neben diesen hierarchisch geordneten Datensind Metadaten gesondert zu betrachten, die spezifisch für das Freilichtmuseum Mueß sind, und zwar „Author“, „Rights“, „Copyright“ und „Credit“. Obwohl wir zur Zeit nur eine Quelle der Metadaten, Mueß, spezifizieren, sollen in Zukunft Bildmetadaten weiterer Quellen in das Ortschronikenportal importiert werden. Einem Bild kann genau eine Quelle zugeordnet werden („Source“ hingegen bezieht sich auf den Namen des Bildarchives, welches nicht eindeutig durch die Quelle definiert ist). Die oben untersuchten funktionalen Abhängigkeiten führen uns zu einem Modell mit vier Entities: „Picture“, „City“, „State“ und „Quelle“ (Abbildung 4.3). Änderungsoperationen auf Relationen, die sich nicht direkt von der „Picture“-Entity ableiten, sind stark eingeschränkt – zum einen sollten Ortschronisten die Autorenschaft und Rechteinformationen generell nicht (bzw. nur im Seitentext des Ortschronikenportals) ändern, andererseits sollten Änderungen, die mehrere Bilder betreffen (wie die Zuordnung einer Stadt zu einem Bundesland) nicht durch Änderung der Metadaten individueller Bilder erfolgen. Falls die Änderung dieser Metadaten durch den Ortschronisten in Zukunft gewünscht ist, so wäre eine spezifische Oberfläche (entsprechend der Organisationssicht für Metadaten auf Katalogebene von Media Pro) zu designen.

Der unabhängig vom Ortschronikenportal funktionierende Prototyp der Software benötigt zwei weitere Attribute der „City“, nämlich Längen- und Breitengrad. Diese Informationen sind jedoch in bereits vorhandenen Tabellen des Ortschronikenportals vorhanden und müssen daher nicht redundant abgespeichert werden. Da Bilder auch unabhängig vom zugeordneten Ort lokalisiert werden sollen, sind die Attribute „Längengrad“ und „Breitengrad“ hingegen als Attribute von „Picture“ notwendig, nicht als Attribute von „City“. Die Koordinaten erhalten als Voreinstellung die Koordinaten der zugeordneten Stadt, sind jedoch durch den Ortschronisten anpassbar. Auf diese Weise könnten auch Bilder mit Koordinaten versehen werden, die weit entfernt von der nächstgelegenen Ortschaft entstanden sind.

Kapitel 5

Prototypische Implementierung und Bewertung

5.1 Prototypische Implementierung

Das vorangegangene Konzept wurde in Form eines Prototypen implementiert, wobei zunächst auf eine Einbindung des Ortschronikenportals oder des Mueßer Bildarchives verzichtet wurde. Stattdessen wurden MediaWiki 1.31.0 und Phase One Media Pro 2.3.0.266 lokal installiert.

Während das Ortschronikenportal das Datenbankmanagementsystem PostgreSQL einsetzt, wurde für die prototypische Implementation MySQL5.7 gewählt. Die Komponenten Importer, Exporter, XML-Adapter, SQL-Adapter, PHP-Adapter und Template-Vervollständiger wurden als Java-Klassen in der Java Runtime Environment 10.0.2 realisiert. Die XSLT-Implementation Xalan2.7.0 wurde als XML-Bibliothek in das Projekt eingebunden. Um von den Java-Komponenten aus auf die API des Datenbankmanagementsystem zugreifen zu können, wurde der MySQL Connector Java 8.0.9 verwendet. Da MediaWiki lokal installiert war, wurde kein SSH Adapter eingebunden, um PHP-Skripte ausführen zu können. Das verwendete PHP-Skript war „edit.php“, welches das Erstellen und Editieren von MediaWiki-Seiten ermöglicht.

Die lokale MediaWiki-Installation wurde um die Extensions ParserFunktion, LocalUpload und DynamicPageList erweitert. ParserFunction ermöglicht, Metadaten Infoboxen dynamisch an vorhandene Metadaten anzupassen, sodass Felder mit unbekanntenen Metadatenwerten ausgeblendet werden. LocalUpload vereinfacht das Hochladen mehrerer Dateien. Alle Bilder, die in einem vordefinierten Ordner abgelegt worden waren, werden automatisch in Wiki integriert, ohne das eine Auswahl einzelner Dateien notwendig ist. DynamicPageList erlaubt das automatische Generieren von Bildergalerien für Kategorien. Wir verwenden diese Funktion, um Bildergalerien für Städte zu erzeugen, auf die über die Kartenanwendung zugegriffen werden kann. Für die Kartenanwendung selbst verwenden wir die JavaScript-Bibliothek OpenLayers5.0.0.

Die Steuerung des Prototyps basiert auf einer graphischen Benutzeroberfläche, wobei den einzelnen Aktivitäten des Workflows eigene Schaltflächen zugewiesen wurden.

Der Prototyp kann einen vollständigen Zyklus – bestehend aus Import, Veränderung der generierten Infobox durch den Nutzer, und Export einer von Media Pro akzeptierten XML-Datei – durchführen. Die Erzeugung der interaktiven Karte, welche auf Bildergalerien für Orte verlinkt, ist ebenfalls möglich. Die Kartenanwendung markiert Städte als farbige Kreise. Sowohl Farbe

der Kreise als auch eine beistehende Zahl informieren den Nutzer über die Anzahl der zugeordneten Bilder. Nach Auswahl einer Stadt wird die Bildergalerie in einem neuen Tab des Browsers geöffnet.

Die Infobox zeigt gegenwärtig alle verfügbaren Annotationen eines Bildes an. Dateiname, Label und Bewertung sind jedoch Daten, die für interne Verwaltungsoperationen benötigt werden und nicht dargestellt werden sollten. Es wird dem Anwender bisher nicht verdeutlicht, welche Metadaten in den Media Pro-Katalog re-exportiert werden und welche Metadaten vom Export ausgeschlossen sind (z.B. Angaben zum Bundesland).

Für die prototypische Implementation werden quellspezifische Metadaten nicht gesondert betrachtet - weder wurden ihre Abhängigkeit voneinander durch Repräsentation in einer eigenen Datenbanktabelle berücksichtigt, noch wurde ihre Re-Import in den Media Pro-Katalog ausgeschlossen.

5.2 Vergleich mit existierenden Systemen

Sowohl die eingangs beschriebenen Webseiten Flickr und Fotki, als auch mit MediaWiki generierte Wikis, erlauben auf Wunsch das Anzeigen technischer Metadaten hochgeladener Fotografien als Tabelle. Diese Funktion von MediaWiki bleibt auch im Ortschronikenportal erhalten, zusätzlich erlaubt die Software jedoch die tabellarische Darstellung der Annotationen des Archivars aus Mueß. In den genannten Systemen können Annotationen in Form von „Tags“, „Hashtags“ oder „Kategorien“ in ungeordneter Weise vorgenommen werden. Diese Tags haben gegenüber den tabellarischen Einträgen der vorliegenden Arbeit den Vorteil, nicht nur der Information und Klassifikation, sondern auch der Navigation zu dienen. In der vorliegenden Arbeit beschränken wir uns darauf, hierarchische Schlüsselwörter und die zugeordnete Stadt auf Media Pro-Kategorien abzubilden. Das Einfügen der entsprechenden Ortskategorie dient jedoch nicht in erster Linie der verbesserten Navigation zwischen verschiedenen Bildern eines Ortes, sondern erleichtert das Generieren von Bildergalerien für jeden Ort, die über die Kartenanwendung erreichbar sind.

Bezeichnungen für die Gestaltung der Metadaten von Bildern durch den Nutzer sind neben „Soziale Klassifikation“ auch „Folksonomy“. Kritik an diesem Begriff basiert auf der fehlenden Hierarchisierung der hierfür meist verwendeten Tags. Wie eingangs erwähnt erlaubt das Soziale Netzwerk Fotki eine einfache Hierarchiebildung auf Basis der Albenstruktur. Die vorliegende Arbeit geht über diese einfache Möglichkeit hinaus, indem MediaWiki-Kategorien verwendet werden, um für jedes Bild beliebig viele Hierarchien zu definieren. Tatsächlich basieren diese hierarchischen Schlüsselwörter jedoch nicht auf den Prinzipien der „Folksonomie“ oder der „Sozialen Klassifikation“, sondern bilden ausschließlich die Ordnungsstruktur des Freilichtmuseums Mueß ab. Vergleicht man die Erwartungen des Freilichtmuseums mit dem Konzept der Sozialen Klassifikation fallen klare Diskrepanzen der Erwartungen auf. Potentielle Vorteile der Sozialen Klassifikation – wie die mögliche Mehrsprachlichkeit, die Aktualität des Wortschatzes und die Verwendung eines Vokabulars, das auch für die Suche nach Bildern vom Anwender verwendet würde – stören die streng kontrollierte Metadatensystematik des Freilichtmuseums. Das vorliegende Arbeit löst dieses Problem teilweise, indem besonders strikt regulierten Metadaten – wie die hierarchischen Schlüsselwörter – nicht nach Mueß re-exportiert werden.

5.3 Alternativansatz: Importieren und Exportieren in Echtzeit

Der in dieser Arbeit vorgestellte Workflow setzt sich aus zwei Phasen – Import und Export – zusammen, die einander abwechseln und durch einen Vertreter des Ortschronikenportals oder ein Skript initiiert werden. Interaktionen des Ortschronisten und des Archivars mit dem jeweiligen System – dem Ortschronistenportal bzw. Media Pro – erfolgen zwischen den beiden Phasen.

Alternativ könnten Import und Export durch Nutzereingaben in Echtzeit ausgelöst werden. Das Erstellen eines neuen Export-Dokumentes durch den Archivar und die Änderung der Metadaten-Infoboxen durch den Ortschronisten würde serverseitig unmittelbar (bzw. nach Auswählen einer entsprechenden Schaltfläche durch den jeweiligen Nutzer) zum Parsen der Seite und zur Aktualisierung der Metadaten-Relationen führen. Diese Interaktion könnte clientseitig durch die JavaScript-Bibliothek JQuery realisiert werden. Serverseitig würde z.B. ein Java Servlet die Anfrage des Nutzers verarbeiten. Die in dieser Arbeit beschriebene Software ließe sich auch nachträglich um entsprechende Komponenten erweitern.

Der Nachteil dieser alternativen Softwarearchitektur besteht in der Notwendigkeit eines mit Servlets kompatiblen zusätzlichen Servers, z.B. eines Apache Tomcats. Die vorliegende Arbeit hat die Konzeption einer Software zum Ziel, die auf den Architekturelementen von MediaWiki beruht, anstelle nachträglich aufwändig mit dieser abgestimmt werden zu müssen. Es gibt am Lehrstuhl Datenbank- und Informationssysteme der Universität Rostock zwar bereits ein vergleichbares Projekt – das Flurnamenarchiv (Abbildung 2.2) – das die genannte Verarbeitung von Nutzereingaben in Echtzeit über ein Servlet realisiert, dieses wurde bisher jedoch nicht in das Ortschronikenportal integriert. Eine weitere Alternative stellt die Konzeption einer vollständigen MediaWiki Extension auf PHP-Basis dar. Der hiermit verbundene Arbeitsaufwand hätte jedoch den Rahmen der vorliegenden Arbeit gesprengt.

Da Metadaten vom Archivar des Freilichtmuseums in Form von Katalogen exportiert und auch wieder importiert werden, ist eine Reaktion der Software auf Veränderungen einzelner Bildmetadaten ohnehin nicht relevant. Für den Ortschronisten ist die fehlende Reaktion der Plattform in Echtzeit nur bei Zuordnung eines Bildes zu einer Stadt erkennbar: Die Kartenanwendung zeigt für jeden Ort die Anzahl verfügbarer Bilder an, aktualisiert diese Angaben jedoch nur in der Importphase.

Kapitel 6

Zusammenfassung und Ausblick

6.1 Zusammenfassung

Um Bildinhalte digitaler Bildarchive such- und nutzbar zu machen, ist die Auszeichnung der Bilder mit deskriptiven Metadaten von zentraler Bedeutung. Im Rahmen dieser Arbeit wurde ein Konzept erarbeitet, das Bildarchiv des Freilichtmuseums Mueß online bereit zu stellen und dabei Bilder zusammen mit ihren Metadaten auf Seiten des Ortschroniken-Portals zu präsentieren. Um die Expertise der Nutzer des Portals im Rahmen von Citizen Science zu nutzen, soll ihnen die Möglichkeit gegeben werden, die bereitgestellten Metadaten anpassen zu können. Nach eingehender Prüfung werden die Änderungen in das Bildarchiv des Freilichtmuseums übernommen. Das vorgestellte Konzept wurde in Form eines Prototyps implementiert, welcher Metadaten zwischen lokal installierten Instanzen von Media Pro und MediaWiki austauscht.

6.2 Ausblick

Die vorliegende Arbeit stellt keine allgemeine Schnittstelle zwischen Bildkatalogisierungs-Programmen und auf MediaWiki basierenden Web-Portalen vor, sondern ist spezifisch auf das Freilichtmuseum Mueß und die dort verwendete Software Media Pro abgestimmt. Wie im Kapitel „Datenbankarchitektur“ erwähnt, soll die hier konzipierte Software künftig aber auch den Import aus anderen Quellen als dem Freilichtmuseum Mueß zulassen. Da die im Datenbankentwurf verwendeten Attribute auf Metadatenfeldern des IPTC Core Standards abgebildet werden können, ist prinzipiell eine Kompatibilität mit anderen Bildquellen, die IPTC Core verwenden, gegeben. Im Einzelfall ist jedoch zu klären, ob die Felder von jeder Quelle in gleicher Weise genutzt werden. Im Gegensatz zum Entwurf der Metadaten-Relationen entspricht das XML-Export-Dokument keinem Metadaten-Standard, sondern ist abhängig von der Katalogisierungssoftware Media Pro. Tatsächlich hat Phase One bereits vor Abschluss dieser Arbeit die Weiterentwicklung von Media Pro eingestellt und einige seiner Funktionen in die Bildaufnahmesoftware Capture One integriert. Unglücklicherweise schließt dies nicht die umfangreichen Import- und Exportfunktionen von Media Pro ein. Da Media Pro funktionstüchtig bleibt und Capture One die Katalogdateien von Media Pro verarbeiten kann, spricht kurzfristig nichts dagegen, die Import- und Export-Funktionalität von Media Pro weiterhin und parallel zu Capture One zu nutzen. Langfristig sollte jedoch eine andere Katalogisierungssoftware gefunden werden, die ebenfalls XML-Dateien (gege-

benenfalls serialisierte XMP-Dateien) importieren und exportieren kann. Da diesen XML-Dateien voraussichtlich eine andere DTD zugrunde liegt als den Media Pro Export-XML-Dokumenten, ist die XML-Java-Adapterkomponente der Software anzupassen.

Adobe Photoshop Lightroom stellt eine weit verbreitete Katalogisierungssoftware dar, die mit Capture One Bilder und Metadaten austauschen kann. Im Gegensatz zu Capture One speichert Lightroom Metadaten in leicht zugänglichen XMP Dateien (als „side car“ bzw. Filial-Dateien des Dateityps xmp). Capture One kann diese Dateien zwar lesen, aber nicht erzeugen. DigiKam und Darktable stellen Open Source Alternativen dar, welche ebenfalls das xmp-Format für Filial-Dateien verwenden.

Im Abschnitt „Online Bildarchive“ wurde bereits CONTENTdm erwähnt. Im Gegensatz zu Media Pro legt CONTENTdm seiner Metadatenkategorisierung den Dublin Core Standard zu Grunde. Obwohl CONTENTdm vielversprechende Export- und Importfunktionalitäten aufweist, spricht der unterschiedliche Metadatenstandard gegen einen Umstieg auf dieses System.

Ein weiteres Beispiel für ein Katalogisierungssystem, das vor allem von Museen und Archiven genutzt wird, stellt FAUST von Land Software-Entwicklung dar. Das System bietet von Haus aus Möglichkeiten an, auf Datenbankinhalte online über OAI-PMH zuzugreifen. Um diese Möglichkeiten effizient ausnützen zu können, wäre jedoch eine völlige Umstrukturierung der vorgelegten Software nötig.

Obwohl zum Zeitpunkt der Vollendung dieser Arbeit noch nicht vorhergesehen werden kann, welche Katalogisierungssoftware das Freilichtmuseum Mueß in Zukunft verwenden wird und welche Kompatibilitäten diese ausweisen wird, ist durch die in dieser Arbeit beschriebenen Software zumindest sichergestellt, dass auf die bisher definierten Metadaten dauerhaft zugegriffen werden kann.

Literaturverzeichnis

- [BCD⁺09] BONNEY, RICK, CAREN B COOPER, JANIS DICKINSON, STEVE KELLING, TINA PHILLIPS, KENNETH V ROSENBERG und JENNIFER SHIRK: *Citizen science: a developing tool for expanding science knowledge and scientific literacy*. *BioScience*, 59(11):977–984, 2009.
- [CJ16] CAMERA & IMAGING PRODUCTS ASSOCIATION und JAPAN ELECTRONICS AND INFORMATION TECHNOLOGY INDUSTRIES ASSOCIATION: *Exchangeable image file format for digital still cameras: Exif Version 2.31*, 2016.
- [Coh08] COHN, JEFFREY P: *Citizen science: Can volunteers do real research?* *AIBS Bulletin*, 58(3):192–197, 2008.
- [Cou18] COUNCIL, INTERNATIONAL PRESS TELECOMMUNICATIONS: *IPTC Photo Metadata Standard*. <https://iptc.org/standards/photo-metadata/iptc-standard/>, zuletzt aufgerufen am 03.11.2018, 2018.
- [Gro10] GROUP, METADATA WORKING: *Guidelines For Handling Image Metadata*, 2010.
- [iML] MULTIMEDIA LIMITED iVIEW: *Manual iView MediaPro 2.6 (Windows, Mac)*. iView Multimedia Limited.
- [IN14] INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL und NEWSPAPER ASSOCIATION OF AMERICA: *IPTC-NAA Information Interchange Model Version No. 4.2*, 2014.
- [Inc12] INCORPORATED, ADOBE SYSTEMS: *Extensible Metadata Platform (XMP) Specification Part 1: Data Model, Serialization, and Core Properties*, 2012.
- [Inc16] INCORPORATED, ADOBE SYSTEMS: *Extensible Metadata Platform (XMP) Specification Part 2: Additional Properties*, 2016.
- [Int18] INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL: *IPTC Photo Metadata Standard 2017.1*, 2018.
- [JH16] JANKE, VOLKER und THOMAS HELMS: *Fotoalben und Ihre Digitale Erschließung — Ein Praxisbericht*. *Rundbrief Fotografie*, 23(2):33–42, 2016.
- [LCH09] LI, YUNPENG, DAVID J CRANDALL und DANIEL P HUTTENLOCHER: *Landmark classification in large-scale image collections*. In: *Computer vision, 2009 IEEE 12th international conference on*, Seiten 1957–1964. IEEE, 2009.

- [Mat06] MATUSIAK, KRYSZYNA K: *Towards user-centered indexing in digital image collections*. OCLC Systems & Services: International digital library perspectives, 22(4):283–298, 2006.
- [Ort18] ORTSCHRONIKEN: *Ortschroniken Mecklenburg-Vorpommern*. https://www.ortschroniken-mv.de/index.php?title=Ortschroniken_Mecklenburg-Vorpommern&oldid=11706, zuletzt aufgerufen am 03.11.2018, 2018.
- [Ril18] RILEY, JENN: *Seeing Standards: A Visualization of the Metadata Universe*. <http://jennriley.com/metadatamap>, zuletzt aufgerufen am 03.11.2018, 2018.
- [SJ09] STVILIA, BESIKI und CORINNE JÖRGENSEN: *User-generated collection-level metadata in an online photo-sharing system*. Library & Information Science Research, 31(1):54–65, 2009.
- [WC11] WIGGINS, ANDREA und KEVIN CROWSTON: *From conservation to crowdsourcing: A typology of citizen science*. In: *System Sciences (HICSS), 2011 44th Hawaii international conference on*, Seiten 1–10. IEEE, 2011.

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Vorlage der angegebenen Literatur und Hilfsmittel angefertigt habe.

Rostock, den 7. November 2018