

Digitalisierte Bilder und Schriften Agrartechnik in der „AgTecCollection in mediaTUM®“

Hermann Auernhammer, Arne Seifert, Astrid Teichert, Heinz Bernhardt

Lehrstuhl für Agrarsystemtechnik
Technische Universität München
Am Staudengarten 2
85354 Freising-Weihenstephan
{Hermann.Auernhammer;Heinz.Bernhardt}@wzw.tum.de

Universitätsbibliothek
Technische Universität München, München
{Arne.Seifert;Astrid.Teichert}@ub.tum.de

Abstract: Analoges Wissen wird morgen nur dann genutzt werden, wenn es in gut dokumentierter Form digital verfügbar gemacht wird. Mit mediaTUM® wurde dafür ein leistungsfähiges Archivierungssystem geschaffen. Verfügbares agrartechnisches Material in Form von Bildern und Schriften wurde digitalisiert und aufbereitet. In der Kollektion „AgTecCollection“ in mediaTUM erfreut es sich weltweiter Beliebtheit im Hinblick auf Zugriffe und Downloads.

1 Einleitung

Heute wird im wissenschaftlichen Tagesgeschehen ebenso wie in der breiten Gesellschaft nahezu ausschließlich mit digitalen Medien gearbeitet. Nicht in digitaler Form verfügbares Material wird vermehrt als „nicht existent“ betrachtet. Die Nutzung analogen Materials in Form von Bildern und Schriften mit einem unvermeidbar höheren Zeitaufwand und zusätzlichen Maßnahmen bei der Erfassung und Verarbeitung wird so zur Ausnahme oder zu einer „verordneten Verpflichtung“, wenn wissenschaftlich fundierte Recherchen gefordert und dokumentiert werden müssen. Im Interesse aller bisher erbrachten wissenschaftlichen Leistungen mit Dokumentation in analoger Form muss deshalb über Maßnahmen zu einer Überführung dieses „Schatzes“ in die digitale Welt von heute und morgen nachgedacht werden.

2 Archivierungssystem mediaTUM®

Im DFG-geförderten Projekt IntegraTUM entwickelte die Bibliothek der Technischen

Universität München die Medienserver-Software mediaTUM®1 [LPS10]. Im Vordergrund stand neben der Publikation und des Nachweises von Informationen vor allem die Archivierung wertvoller Bestände. Daher wurden die Funktionalitäten und Anforderungen der Langzeitarchivierung bereits bei der Planung der Infrastruktur/Software mit berücksichtigt und das System frühzeitig an den Langzeitspeicher des Leibniz-Rechenzentrums angeschlossen [SE10]. Die Philosophie der Entwicklung war geprägt von der engen Zusammenarbeit mit den späteren Nutzern des Systems. Ausgehend von einer priorisierten Anforderungsliste der Anwender konnte so eine Software gebaut werden, die exakt auf die Bedürfnisse der Benutzer zugeschnitten war und zugleich deren Akzeptanz gegenüber dem System erhöhte.

Großer Wert wurde auf Flexibilität und Erweiterbarkeit gelegt. Es sollte jederzeit möglich sein, ausgehend von einem generischen Ansatz flexibel auf Kundenanforderungen und -wünsche zu reagieren und neue Funktionalitäten einzubauen. So können z.B. nicht nur Bilder in hierarchischen Strukturen verwaltet und archiviert werden, sondern nahezu auch jeder beliebige Inhalt jedes beliebigen digitalen Objektes. Eine Vielzahl an Datentypen ist integrierbar. Dies reicht von Bildern über Dokumente, Videos und Audiomaterial bis hin zu Forschungsprimärdaten. Aktuell in Entwicklung sind weitreichende Funktionalitäten aus dem Bereich eScience, womit wissenschaftliches Primärmaterial einfach und sicher aufbewahrt, dokumentiert und für spätere Forschungen verfügbar gemacht werden kann.

Datensicherheit und Zugriffsschutz: Neben Flexibilität und Langzeitarchivierung nimmt die Sicherheit der Inhalte gegenüber von Zugriffen Dritter einen sehr hohen Stellenwert ein. Wichtiges Material soll nur einem zugeordneten Personenkreis zur Verfügung stehen, wozu ein mächtiges Rechte- und Benutzerkonzept gefragt ist. Durch die Realisierung als Cloud-Service [ST12] und dem damit verbundenen Zugriff über das Internet ist zudem der Schutz der Daten enorm wichtig. Insgesamt zeigt sich aber, dass vor allem der Zugriff über das Internet, ohne die Notwendigkeit der Installation weiterer Software, für die einfache Verbreitung und Bereitstellung von Inhalten sehr sinnvoll ist, um Wissen für einen breiten Interessentenkreis zur Verfügung zu stellen.

Technische Details: Die komplette Software für mediaTUM® besteht aus Open-Source-Komponenten und ist selbst unter einer GNU General Public License² veröffentlicht. Das bedeutet für den Systembetreiber, dass keine Kosten für Software entstehen und jeder an der Weiterentwicklung teilnehmen kann. Als Programmiersprache kommt dabei Python³ zum Einsatz, eine einfache, aber sehr mächtige Sprache mit weiter Verbreitung. Für die annotierten Metainformationen wird standardmäßig die relationale Datenbank MySQL⁴ verwendet. Es wurde auch darauf geachtet, dass das Betriebssystem nicht festgelegt ist, auf welchem der Server betrieben wird.

¹ <http://mediatum.sourceforge.net>

² <http://www.gnu.org/licenses/gpl.html>

³ <http://www.python.org>

⁴ <http://www.mysql.org>

3 Erfassung analoger agrartechnischer Bilder und Schriften

Mit dem Übergang von der Folie zur PowerPoint-Präsentation in der Lehre wurde 2001 am Lehrstuhl für Landtechnik der TUM in Weihenstephan mit der Digitalisierung des verfügbaren Diabestandes aus der Lehre begonnen. In ersten eigenen Ansätzen wurde für einen Testsatz von 1000 digitalisierten Dias ein vereinfachtes EXCEL-basiertes Archivierungssystem entwickelt und getestet. Die Weiterentwicklung dieses Systems wurde Mitte 2003 abgebrochen, als sich die Zusammenarbeit mit der Universitätsbibliothek der TU-München eröffnete [AH05].

Über mediaTUM[®] konnte nun ein leistungsfähiges Dia-Digitalisiersystem beschafft werden. Damit wurde es möglich, neben den Dias aus der Lehre auch alle verfügbaren Dias aus der bisherigen Forschung in das Vorhaben einzubeziehen. Zudem wurde der Gesamtbestand von 3914 Dias vom Lehrstuhl für Landmaschinen der TU-München in Garching übernommen. Systematisch wurde dazu der Arbeitsablauf geplant und über wissenschaftliche Hilfskräfte umgesetzt. In weitgehender Selbstständigkeit wurde von diesen jedes einzelne Dia gereinigt, mit einer fortlaufenden Nummer versehen, verfügbare Daten zum Dia in eine EXCEL-Tabelle übertragen und danach über Stab- oder Rundmagazine der Digitalisierung zugeführt. Bis Mitte 2006 konnten so insgesamt 51632 Dias digital erfasst werden. Die Originale wurden danach dem Archiv der TU-München übergeben.

Parallel dazu wurden alle 6407 Schwarz/weiß-Dias aus dem Institut für Landtechnik der Humboldt-Universität Berlin übernommen, über einen Durchlicht-Flachbettscanner digitalisiert und die verfügbaren Karteikarteninformationen in ein angepasstes EXCEL-Schema übertragen. Das Diamaterial mit zugehörigen Karteikarten befindet sich nun im Deutschen Landwirtschaftsmuseum in Hohenheim.

Schließlich wurden auch alle am Lehrstuhl für Landtechnik angefertigten Dissertationen und Habilitationen gescannt. Hinzu kam die geschlossene Sammlung aller Schriften der ehemaligen „SCHLÜTER-Seminare“ und der Vorträge aus den „SCHLÜTER-Großvorführungen“. Auch die vom Institut für Landtechnik herausgegebenen Lehrbücher für Landtechnik und Bauwesen (Blaue Bücher Landtechnik) wurden digitalisiert. Alle Digitalisate wurden anschließend mit einem OCR-Schriftenerkennungsprogramm bearbeitet. Abschließend wurden von allen Autoren der Bilder und Schriften Rechteübertragungen eingeholt.

4 AgTecCollection in mediaTUM

Begleitend zur Entwicklung von mediaTUM wurde in enger Abstimmung zwischen dem Team mediaTUM und den lehrstuhlverantwortlichen Betreuern der Digitalisierung die Integration der Bilder und Schriften Landtechnik vorangetrieben. Wesentliche Arbeitspakete bestanden in:

- Der Entwicklung eines geschlossenen Schlagwortkatalogs, welcher weitgehend unverändert alle derzeitigen und künftig zu erwartenden Digitalisate (Bilder, Schriften, Prospekte, Beweg-

bilder) repräsentativ beschreiben kann;

- Einem daraus abgeleiteten Klassifikationssystem für den verzeichnisorientierten Zugriff auf die veröffentlichten Objekte;
- Den Metadaten schemata für Bilder und Schriften mit den bibliothekarisch und agrartechnisch erforderlichen Informationen in Form von Wahl- und Pflichtfeldern mit Einfacheingaben oder mit offenen und geschlossenen Indexlisten;
- Einer aussagefähigen Gestaltung der Einstiegsmasken für die agrartechnische Kollektion;
- Wichtigen Zugangsinformationen für den Nutzer im Hinblick auf die Herkunft der Digitalisate, das Copyright und auf die erforderliche Zitation.

Zudem wurden die während der Digitalisierung erfassten Informationen in den EXCEL-Tabellen gründlich überarbeitet. Dann wurden alle in den Digitalisaten enthaltenen Dubletten mit Hilfe eines EDV-Programms isoliert und gekennzeichnet. Anschließend erfolgte die manuelle Ergänzung fehlender Informationen. Schließlich wurde eine Ersetzungsliste erarbeitet, um digitalisierte Zeichnungen durch digital verfügbare CAD-Zeichnungen (AutoCad) ab 1986 zu ersetzen. Erst dann erfolgte die Übergabe aller Digitalisate einschließlich der EXCEL-Tabelle an das Team mediaTUM. Dort konnten die Digitalisate weitgehend problemlos in das System mediaTUM eingespielt und die Informationen aus den EXCEL-Tabellen den jeweiligen Metadaten zugeordnet werden.

Bis Oktober 2012 konnten nahezu 35000 Bilder erschlossen und veröffentlicht werden. Die monatlich erstellten Zugriffs- und Downloadstatistiken zeigen eine ständig zunehmende Nutzung der Kollektion nach absoluten Zahlen und Ländern (Abb. 1).

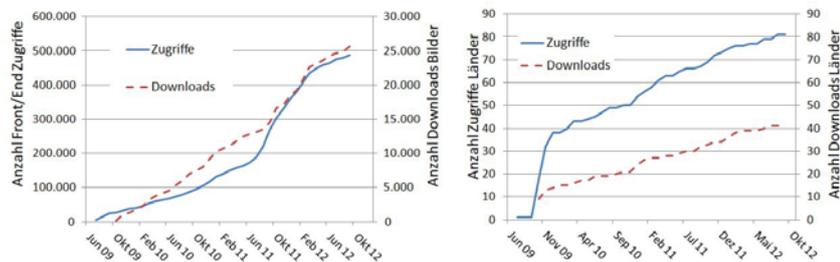


Abbildung 1: FrontEnd-Zugriffe und Downloads (links) und Anzahl der Länder mit Zugriffen und Downloads (rechts) auf Bilder der AgTecCollection in mediaTUM

Literaturverzeichnis

- [AH05] Auernhammer, H.: *Projektbeschreibung AgTecCollection (Bilder und Schriften Landtechnik) in mediaTUM*. AgTecCollection: FG Technik im Pflanzenbau TUM / Auernhammer, TU München 2011, <http://mediatum.ub.tum.de/?id=1078039>.
- [LPS10] Leiss, J.; Pretz, E.; Seifert, A. 2010, mediaTUM: Der zentrale Medienserver der Technischen Universität München, In: Informationsmanagement in Hochschulen; Bode, A.; Borgeest, R. (Hrsg.), Springer, S. 365–377.
- [SE10] Seifert, A.: Langzeitarchivierung bei Medienservern, Dissertation, Universität der Bundeswehr München, <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:706-2185>.
- [ST12] Seifert, A.; Teichert, A.: mediaTUM – Medien in der Cloud, Vortrag auf dem Bayerischen Bibliothekstag 2012, <http://mediatum.ub.tum.de/?id=1109402>.