4. Digitalisierung historischer biologischer Literatur – BioLib: http://biolib.de (Stüber, K.)

Die Digitalisierung historischer biologischer Literatur dient mehreren Zwecken. Zum einen werden durch die Verfügbarkeit über elektronische Medien wertvolle Originalwerke geschont und für den Benutzer bequem zugänglich gemacht. Zum anderen lassen sich über Schlagworte und Volltextrecherchen schwer zugängliche Textstellen auffinden, die für Zitate oder als Belege ermittelt werden müssen. Durch den leichten Rückgriff auf Originalbeschreibungen können Unsicherheiten bei der Festlegung von Artnamen leichter vermieden werden. Die Artbeschreibungen sind aber oft bereits im 17ten oder 18ten Jahrhundert erfolgt und die entsprechenden Werke nur schwer zugänglich.

Nicht zuletzt sind sehr viele historische biologische Abbildungen künstlerisch wertvoll und sprechen auch den heutigen Menschen durch ihre Detailtreue und Ästhetik an, was sich an dem großen Interesse zeigt, das den bereits im Netz verfügbaren botanischen Werken auch vom allgemeinen Publikum entgegengebracht wird. Derzeit werden täglich circa 30-40.000 Zugriffe registriert.
Bislang wurden hauptsächlich klassische Werke der Biologie auf unterschiedliche Weise digitalisiert. Die Werke wurden zum Teil im Volltext erfasst oder gescannt und als Graphik den Internet-Benutzern zur Verfügung gestellt. Z.Z. sind 53000 gescannte Seiten aus 237 Werken verfügbar.
Zur Erleichterung und Automatisierung der Arbeit wurden eine Reihe PERL-Programme entwickelt, die sowohl unter Windows, UNIX/LINUX oder MacOS benutzt werden können. Das Ziel ist eine möglichst rasche Digitalisierung bei gleichzeitig höchster Qualität. Das Einscannen der Rohdaten kann mit Hilfe einer hochauflösenden Digitalkamera sehr rasch erfolgen. Bei guten Vorlagen können etwa 6 Seiten pro Minute eingescannt werden. Diese werden dann im Computer gespeichert und auf CDs gebrannt. Die meisten weiteren Arbeitsschritte sind weitgehend automatisiert. Sind die Bilddaten aufbereitet, können die Webseiten mit den oben genannten Programmen automatisch erstellt werden. Danach erfolgt eine manuelle Qualitätskontrolle und -korrektur bei Seiten, die Fehler aufweisen. Als letztes erfolgt die Verschlagwortung, d.h. es werden den Webseiten, die für das Dokument wichtigen Schlüsselwörter (Themen, Art- und Gattungsnamen, Stichwörter etc.) beigefügt, ein Vorgang, der ebenfalls nur manuell erfolgen kann, aber in Zukunft teilweise durch OCR (optical character recognition) ergänzt werden soll. Erste Tests mit verschiedenen OCR-Programmen ergab eine sehr gute Lesegenauigkeit mit Hilfe des Programmes ABBYY-FineReader.
Insgesamt muss für ein Buch mit 300 Seiten mit einem Aufwand von 3-4 Arbeitsstunden gerechnet werden. Alle digitalisierten Bücher sind über die Webseite http://www.biolib.de öffentlich zugänglich. Die einzelnen Werke sind z.Z. hauptsächlich auf dem Webserver des MPI für Züchtungsforschung, Köln gespeichert, weitere Spiegelungen sind in Vorbereitung.