

Bayerische Akademie der Wissenschaften

Rundgespräche der Kommission für Ökologie

37

Ökologische Rolle von Pilzen



Verlag Dr. Friedrich Pfeil ISSN 0938-5851 · ISBN 978-3-89937-099-7

Pilzherbarien – Neue Aufgaben im Bereich Biodiversitätsinformatik und Datenmanagement

Dagmar Triebel*

Zusammenfassung

Naturwissenschaftliche Sammlungen wie Pilzherbarien sind Forschungsinfrastruktur- und Wissenschaftseinrichtungen mit großer Bedeutung für Wissenschaft und Gesellschaft. Sie erfüllen traditionell eine Reihe wichtiger Aufgaben, wobei mit Einzug des digitalen Zeitalters neuartige Herausforderungen im Bereich der Biodiversitätsinformatik und des Datenmanagements hinzukamen. Biologische Forschungsdaten müssen digitalisiert, nachhaltig gepflegt sowie zur Analyse prozessiert werden. Im Allgemeinen sind sie komplexer, als es zum Beispiel Daten aus dem Bibliothekswesen sind. Dies verdeutlichen die Modelle der Datenbanken im Rahmen der *Diversity Workbench* sowie die dargestellten Datenflüsse aus den Bereichen taxonomische Daten, Sammlungsdaten, Beobachtungsdaten und ökologische Daten von Pilzen. Innovative Entwicklungen in der Biodiversitätsinformatik und zum Management biologischer Daten können langfristig nur in Kooperationsprojekten von Biologen und Informatikern durchgeführt werden. Entsprechende Projekte müssen daher auf Langfristigkeit angelegt sein. So erscheint es sinnvoll, Abteilungen für Biodiversitätsinformatik zu etablieren und in diesem Kontext eine Institutionalisierung von Datenrepositorien, d.h. digitalen Archiven für biologische Daten, anzustreben. IT-Zentren mit diesem Auftrag sollten an Infrastruktureinrichtungen eingerichtet werden, wie es Naturwissenschaftliche Sammlungen sind.

Summary

Fungal herbaria – new missions concerning biodiversity informatics and data management. Natural History Collections like fungal herbaria are science infrastructure institutions with great relevance for sciences and society. They traditionally have a number of important missions; with the digital era, there are further challenges concerning biodiversity informatics and data management. Biological research data have to be digitalised, sustainably curated and processed. They are in general more complex than, e.g., data in the librarianship. This becomes clear when the software design of the *Diversity Workbench* framework and the exemplary data flows concerning taxonomic data, collection and observation data as well as ecological data of fungi are explained. Innovative developments in the frame of biodiversity informatics and management of biological data should be realized in cooperation between biologists and information scientists. Such projects have to be established on the long run. For these reasons, it makes sense to establish own departments for biodiversity informatics and link them with institutional data repositories, i.e. digital archives for biological data. IT centers with this mission might be best set up in institutions with infrastructure mandate like Natural History Collections.

^{*} Triebel, Dagmar, Dr., IT-Zentrum der Staatlichen Naturwissenschaftlichen Sammlungen Bayerns, Botanische Staatssammlung München, Menzinger Straße 67, 80638 München. E-Mail: triebel@bsm.mwn.de

Einführung

»Wer schreibt, der bleibt. . . . Ob das auch für uns Kinder des elektronischen Zeitalters gilt, hängt davon ab, ob wir die Herausforderungen der digitalen Welt bewältigen oder ob wir unzugängliche Datenfriedhöfe produzieren. « (zitiert aus dem Editorial des Bayerischen Staatsministers für Wissenschaft, Forschung und Kunst Dr. W. Heubisch, Aviso 2, 2009).

Diese Problematik gilt auch für wissenschaftliche Daten an Naturwissenschaftlichen Sammlungen, die im Zusammenhang mit den Sammlungsstücken selbst als Informationsquelle für aktuelle und zukünftige Fragen der Umweltforschung und Ökologie dienen.

Am Beispiel von Pilzherbarien werden im Folgenden neue Aufgaben im Bereich Biodiversitätsinformatik und Datenmanagement dargestellt. Es soll gezeigt werden, dass »was die Zeiten überdauert, [nicht] Spiel des Zufalls« bleiben muss (W. Heubisch, Aviso 2, 2009) und dass das »elektronische Zeitalter« mit seinen digitalen Archiven nicht nur im Bibliotheksbereich, sondern auch für Naturwissenschaftliche Sammlungen neuartige Herausforderungen bereithält.

Traditionelle Aufgaben biologischer Sammlungen

Biologische Sammlungen sind in erster Linie Forschungsinfrastruktur- und Wissenschaftseinrichtungen mit großer Bedeutung für Wissenschaft und Gesellschaft. Expertenwissen und Sammlungsobjekte stehen zur Verfügung, um Fragen der Grundlagenforschung im Bereich Biodiversitätsforschung, Systematik, Evolutionsbiologie, Verbreitungsbiologie, Ökologie, aber auch Fragen zur Änderung der Natur und Umwelt über die Zeit und viele andere gesellschaftsrelevante Fragen zu beantworten (Anonymous 2009). Sammlungen fungieren unter anderem als Referenzsammlungen (Hinterlegung von Standards wie Typusmaterial) und als »Memory-Institutionen«, d. h. Archive für biologische Objekte (Triebel 2007). Es gibt weltweit etwa 450 öffentliche Einrichtungen mit kuratorisch betreuten Pilzsammlungen, die schätzungsweise 850 Wissenschaftler beschäftigen (Thiers 2009). Die größte derartige wissenschaftliche Einrichtung sind die US National Fungus Collections (BPI), Beltsville, Maryland, mit 1,1 Millionen



Abb. 1. Pilzherbarium. Teil der Rollregalanlage an der Botanischen Staatssammlung München.

getrockneter Pilzen. Die Bestände an hinterlegten Pilz- und Flechtenproben der Botanischen Staatssammlung München (M) sind mit etwa 650 000 Sammlungsstücken ebenfalls sehr umfangreich (Abb. 1).¹ Zusätzlich werden Beschreibungen, Zeichnungen und Aquarelle aufbewahrt, die vergängliche Merkmale wie Gestalt und Farben von Pilzfruchtkörpern dokumentieren (Abb. 2).

Zu den traditionellen Aufgaben der Experten an biologischen Sammlungen gehören die präparatorische und kuratorische Betreuung der Bestände. In Bayern geschieht dies seit 1813 mit Gründung des Königlichen Herbariums, wobei die ältesten Pilzobjekte bereits aus dieser Zeit stammen (Abb. 2a). Dabei ging es zu allen Zeiten nicht nur um den Erhalt von Sammlungen bedeutender Forscher nach deren Tod, sondern v.a. auch um eine gezielte Vermehrung und wissenschaftliche Bearbeitung der Bestände, um die zu der jeweiligen Zeit bekannte Vielfalt der Lebensformen zu erforschen und zu dokumentieren. Nur so erlangt eine Sammlung Bedeutung und »lebt«. Die Sammlung in München wächst derzeit um jährlich etwa 5000 Pilzobjekte. Dabei wird darauf geachtet, dass vor allem besonders wertvolles Material wie Referenzmaterial zu aktuellen Forschungen, noch nicht in der Sammlung vorhandene Arten sowie in jüngerer Zeit gesammelte Pilze und Funde aus mykologisch schlecht untersuchten Regionen hinzukommen. Die Bestände werden wissenschaftlich betreut. Dies umfasst auch die Teilnahme der Institution am internationalen Leihverkehr, über den jährlich

¹ http://www.botanischestaatssammlung.de/

rund 1000 Pilzobjekte an über 100 Forschungseinrichtungen weltweit zu den verschiedensten Forschungszwecken versandt werden.

Dazu kommen allgemeine Bildungsaufgaben sowie in den letzten Jahren der Aufbau neuer Formen von biologischen Sammlungen, wie spezielle Kultursammlungen und Sammlungen von DNA-Proben. So wird an der Schwesterinstitution der Botanischen Staatssammlung, der Zoologischen Staatssammlung München, seit 2007 eine so genannte DNA-Bank aufgebaut, in der neben DNA-Proben zoologischer Objekte auch DNA-Proben von Pilzobjekten verfügbar gehalten werden.²

Neue Aufgaben biologischer Sammlungen im 21. Jahrhundert

Im digitalen Informationszeitalter haben sich die geschilderten Aufgabenfelder der Sammlungen beträchtlich erweitert. Die neuen Aufgaben beziehen sich jetzt nicht mehr allein auf die Sammlungsobjekte, sondern zusätzlich auch auf die Digitalisierung der Information zu diesen Objekten. Damit tritt zur Funktion als traditionelle »Objektrepositorien« diejenige von so genannten »Datenrepositorien« bzw. digitalen Datenarchiven hinzu. Dies bedeutet, dass moderne For-

schungssammlungen zunehmend Kompetenzen im Datenmanagement aufbauen und Schwerpunkte in der Biodiversitätsinformatik setzen müssen. Unter Biodiversitätsinformatik versteht man dabei Theorie und Anwendung von Informatikmethoden zur Erfassung, Modellierung, Aufbereitung, Analyse und Archivierung von Biodiversitätsdaten.

An den Staatlichen Naturwissenschaftlichen Sammlungen Bayerns (SNSB), dem Verbund der Bayerischen Naturwissenschaftlichen Sammlungen und Museen, wurde die Bedeutung der Biodiversitätsinformatik in Bezug auf sammlungsobjektassoziierte Daten erkannt und führte 2006 zur Gründung des SNSB IT-Zentrums.⁴ Das SNSB IT-Zentrum versteht sich als IT-Entwicklungsabteilung und Datenrepositorium und übernimmt damit mehrere Aufgaben: (1) den Aufbau eines Datenbank-Frameworks für objekt- und taxonassoziierte biologische Daten (»Diversity Workbench«), (2) die Etablierung nachhaltiger Datenflüsse und virtueller Netzwerke von der Dateneingabe bis zur Datenarchivierung,

- 2 http://zsm.mwn.de/dnabank/
- 3 http://www.botanischestaatssammlung.de/DatabaseClients/BSMwohlfcoll/
- 4 http://www.snsb.info





Abb. 2. Sammlungsobjekte und Dokumente aus dem Pilzherbarium der Botanischen Staatssammlung München. a, Funddokumente zu zwei der ältesten Pilz-Aufsammlungen von 1816 (*Dactylospora athallina*, M-0041389, links) und 1823 (*Craterium minutum*, M-0048158, rechts); b, zwei Schleimpilz-Funde aus der Sammlung Hermann Neubert mit winzigen Fruchtkörpern von 0,5-1 mm Durchmesser (*Lamproderma splendens*, M-0118959, oben; *Lepidoderma tigrinum*, M-0116975, unten); c, Aquarellzeichnungen von Fritz Wohlfarth (oben: *Boletus edulis* (Steinpilz), M-0050801, links; *Lactarius semisanguifluus* (Blutreizker), M-0051451, rechts; unten: *Clitocybe nebularis* (Nebelgrauer Trichterling, Nebelkappe), M-0050908, links; *Suillus plorans* (Zirbenröhrling), M-0052221, rechts)³. – a, b, c, © Botanische Staatssammlung München; b: Makroaufnahmen © Karlheinz Baumann.

Sammlung H. Neubert: 5646 M-0118959

Lamproderma splendens Meyl.

FRANCE. Frankreich, Savoyen, Col du Pré, 1700 m

17.5,1989 leg: M. Meyer, W. Nowotny, H. Neubert



Sammlung H. Neubert: 3662 M-0116975

Lepidoderma tigrinum (Schrad.) Rostaf.

GERMANY. Deutschland, Baden-Württemberg,
Bühlertal, Gertelbachfälle An Moos auf Abies
29.3.1981 leg: H. Neubert













-prozessierung und -präsentation über internationale Netzwerke und (3) den Aufbau eines institutionellen Datenrepositoriums, d.h. eines zentralen digitalen Archivs für biologische Daten.

Diese zusätzlichen neuen Aufgaben der Sammlungen und assoziierter digitaler Archive im Bereich Biodiversitätsinformatik und Datenmanagement betreffen nicht nur Pilzherbarien, sondern sämtliche öffentlichen naturwissenschaftlichen Sammlungen. Im Folgenden sollen die Aufgaben (2) und (3) jedoch anhand von Beispielen aus der Mykologie verdeutlicht werden. Dabei soll herausgestellt werden, dass biologische Daten nachhaltig prozessiert, d.h. dauerhaft gepflegt, ergänzt und geändert werden müssen und dass dazu komplexe und innovative informationstechnologische Werkzeuge und Kompetenzen erforderlich sind.

(1) Aufbau des Datenbank-Frameworks »Diversity Workbench«

Das SNSB IT-Zentrum sieht es als eine seiner zentralen Aufgaben an, die Entwicklung von generischer Software, deren Funktionen nach bestimmtem Regeln allgemein entworfen und daher wiederverwendbar sind, zur Schaffung virtueller Arbeitsumgebungen für das Management biologischer Daten voranzutreiben. Dazu wird seit mehr als 10 Jahren am Aufbau eines Datenbank-Frameworks »Diversity Workbench« gearbeitet. Dies geschieht in Kooperation und Diskussion mit Biologen und Informatikern, die an gemeinsamen Datenkonzepten, Modellierungen und Standards arbeiten. Seit 1999 arbeiteten sieben Softwareent-

wickler an der Weiterentwicklung. Die technische Architektur der *Diversity-Workbench*-Plattform basiert auf dem Grundgedanken einer echten Modularität der entwickelten Datenbanken und Datenbankapplikationen, wobei die einzelnen Komponenten einerseits unabhängig voneinander eingesetzt werden können, andererseits sich über entsprechende Schnittstellen gegenseitig aufrufen lassen. Dieses Konzept garantiert eine immense Flexibilität in der technischen Interaktion. Datenbanken und Software werden ständig weiterentwickelt und über das Internet zum freien Download bereitgestellt.⁵ Inzwischen sind Anwendungen der *Diversity Workbench* weltweit an 28 Institutionen im Einsatz.

Die einzelnen Datenbankanwendungen haben jeweils inhaltlich-strukturell und semantisch umgrenzte Datenbereiche, sog. Datendomänen, zum Gegenstand. Als derartige Domänen im Zusammenhang mit Biodiversitätsdaten wurden einerseits Beobachtungs- und Sammlungsdaten, beschreibende Daten sowie Bild-/Mediendaten erkannt und andererseits taxonomische Daten identifiziert; hinzu treten geografische Daten und Literaturdaten (Abb. 3). Die letzten drei genannten Datendomänen haben für die meisten biologischen Forschungsprojekte eine Thesaurus- oder Backbone-Funktion, sollten also z.B. in einem ökologischen Forschungsprojekt nicht neu erhoben, sondern als Standarddatenguellen genutzt werden können. Beschreibende sowie ökologische, molekulare, Beobachtungs- und Sammlungsdaten werden dagegen in der Regel

⁵ http://www.diversityworkbench.net

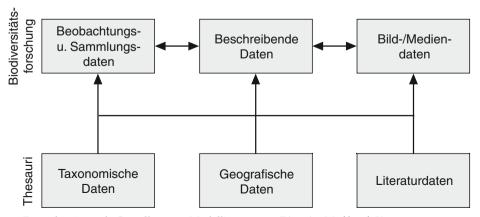


Abb. 3. Datendomänen als Grundlage zur Modellierung von Diversity-Workbench-Komponenten.

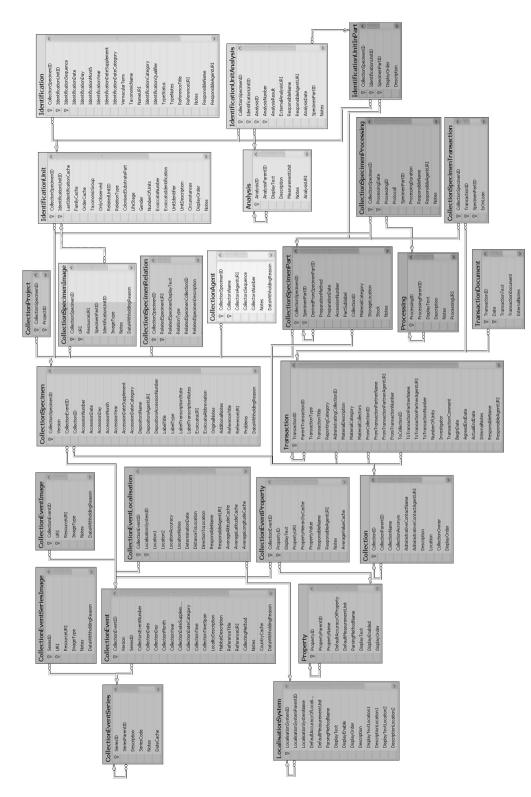


Abb. 4. ER-Schema (Beziehungsmodell) der Datenbankapplikation DiversityCollection.

neu erstellt. Für alle in Abbildung 3 dargestellten Datendomänen und einige weitere werden entsprechende Diversity-Workbench-Komponenten entwickelt, die entsprechend DiversityCollection, DiversityTaxonNames, DiversityDescriptions usw. heißen. Das Datenmodell von DiversityCollection (Abb. 4, Weiss et al. 2007) zeigt die Komplexität einer der derzeit zwölf Datenbanken. Das SNSB IT-Zentrum hat zum Ziel, mit dem Einsatz der Diversity-Workbench-Komponenten möglichst viele Typen von Biodiversitätsdaten strukturiert erfassen und nachhaltig prozessieren zu können.

(2) Etablierung nachhaltiger Datenflüsse und Netzwerke

Die Entwicklung von Software im Bereich Biodiversitätsinformatik steht in engem Zusammenhang mit einer weiteren Aufgabe eines digitalen Datenarchivs an Naturwissenschaftlichen Sammlungen: die Etablierung nachhaltiger Datenflüsse und Netzwerke für Biodiversitätsdaten. Dies soll im Folgenden anhand zweier bereits verwirklichter Netzwerke bzw. eines im Aufbau befindlichen Netzwerkes mit Daten aus der Mykologie, Lichenologie, Botanik und Zoologie dargestellt werden. Die Beispiele beschreiben Entwicklungen am SNSB IT-Zentrum und stellen den Datenfluss und die Prozessierung von taxonomischen Daten, Sammlungsdaten, Beobachtungsdaten und ökologischen Daten dar.

Taxonomische Daten im Informationssystem LIAS

»LIAS names – A Database with names of Lichens, Lichenicolous Fungi and Non-Lichenized Ascomycetes« ist das taxonomische Rückgrat des Informationssystems LIAS.⁶ Es verwaltet als gültig akzeptierte taxonomische Namen sowie Synonyme und Basionyme, d.h. die Taxonomie und Nomenklatur von Flechten- und Pilznamen v.a. aus der Großgruppe der Ascomyceten (Triebel et al. 2007). Dabei umfasst LIAS names als eigenständiges Projekt zurzeit rund 34000 Datensätze, ist seit 2005 online zugänglich und fungiert im Rahmen von »Species 2000, Catalogue of Life« als Global Species Database (GSD).

Der Datenfluss von LIAS names (Abb. 5) basiert

auf DiversityTaxonNames als zentraler Datenbankkomponente (Modell siehe Hagedorn et al. 2008). Über den Rich Client von DiversityTaxonNames, installiert auf den lokalen PCs der Datenerzeuger, werden die Daten in die auf einem Server an den SNSB eingerichtete Serverdatenbank geschrieben. Der Rich Client wird nach den konkreten Erfordernissen taxonomisch arbeitender Arbeitsgruppen seit 2003 technisch ausgebaut (C#-Programmierung im .Net Framework) und erlaubt das Editieren der Daten entsprechend der Vorgaben der verschiedenen nomenklatorischen Codes in der Biologie. Der Datenfluss beginnt also mit der online-Dateneingabe durch den Datenerzeuger am PC-Arbeitsplatz. Die Microsoft SQL-Serverdatenbank dient dabei als Datencontainer und nutzt MS Windows als Plattform.

Um zentrale Dateninhalte für die Allgemeinheit zugänglich zu machen bzw. in anderem Kontext verwenden zu können, werden weitere Software-Implementierungen und technische Mechanismen entwickelt. Inzwischen gibt es für LIAS names JSP-Schnittstellen mit Such- und Browsefunktionen, die auf die Datenbestände zugreifen. Auch der Zugang zu LIAS-names-Daten über SOAP- und REST⁷-basierte Webdienste ist frei. Über diese Mechanismen hat der Anwender die Möglichkeit, den gewünschten Einzeldatensatz nach gezielter online-Anfrage an den SNSB-Server in München in die eigene Datenbank, installiert auf dem eigenen Server oder PC, zu übernehmen. Dies geschieht über einen Programmiercode und kann auch insoweit automatisiert werden, dass Daten über erneute Anfragen an den SNSB Server aktualisiert werden.

LIAS-names-Daten werden zudem über XML-Schnittstellen für spezielle Anwendungen exportiert bzw. auf sie wird über eine nach einem Standard (SPICE) entwickelte Wrapper-Software dynamisch zugegriffen. Dadurch können internationale Projekte und Initiativen wie Species 2000 Anwenderanfragen an ihre Datenportale, wie sie z.B. auf einem Server im U.K. installiert sind, an die angeschlossenen Datenbanken wie LIAS names in München weiterleiten und dem Anwender so die Daten des Partners direkt online zur Verfügung stellen.8

⁶ http://www.liasnames.lias.net/

⁷ http://www.oio.de/public/xml/rest-webservices.

⁸ http://www.catalogueoflife.org/dynamic-checklist/search.php

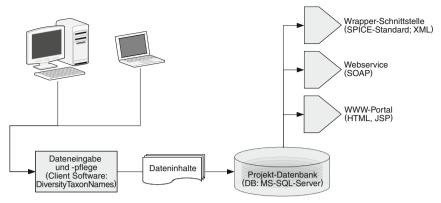


Abb. 5. Schema zum Datenfluss für taxonomische Daten: LIAS names.

Um also eine hohe Flexibilität in der Zusammenarbeit mit internationalen Projekten zu garantieren und gleichzeitig optimale technische Lösungen anbieten sowie technische Neuentwicklungen gegebenenfalls rasch einbauen zu können, sind selbst bei dem relativ einfach strukturierten Netzwerk und Datenfluss im Rahmen von *LIAS names* verschiedene technische Umgebungen und Technologien im Einsatz.

Sammlungsdaten im GBIF-D-Knoten für Mykologie

Der für das Datenmanagement von Sammlungsdaten im Rahmen des Deutschen GBIF-Knotens

für Mykologie (Global Biodiversity Information Facility) seit 2001 aufgebaute Datenfluss ist technisch gesehen ähnlich dem für taxonomische Daten (Abb. 6). Er beruht also auf der dezentralen Dateneingabe über .Net-Clients in eine SQL-Serverdatenbank; hier vor allem Client und Datenbank von DiversityCollection. Darauf folgt allerdings ein Transfer zentraler Tabellen und Felder in eine Cache-Datenbank. Auf die Daten in dieser Cache-Datenbank greifen generische Wrapper (nach ABCD-Standard zur Webanbindung an das GBIF-Portal) einerseits und technisch eigenständige Webschnittstellen-Entwicklungen anderseits zu. Letztere werden am SNSB IT-Zentrum entwickelt und besitzen

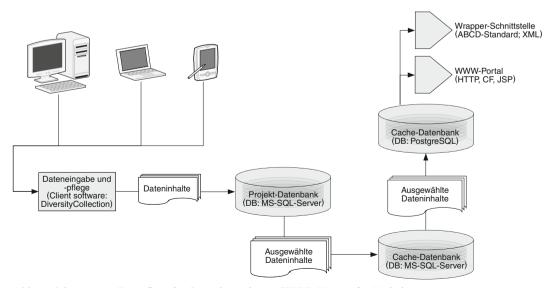


Abb. 6. Schema zum Datenfluss für Sammlungsdaten: GBIF-D-Knoten für Mykologie.

Such- und Downloadfunktionen (über ColdFusion (CF) bzw. Java Server Pages (JSP)). Sie sind an die spezifischen Charakteristika der einzelnen Datensammlungen angepasst. Beispiele für diesen Typ von Webschnittstellen finden sich im GBIF-D-Knoten für Mykologie unter Datenbanken mit Sammlungsdaten.⁹

Bei GBIF handelt es sich um ein internationales Netzwerk, dessen nationale Koordinationsstelle für mykologische Daten am SNSB IT-Zentrum als sog. GBIF-D-Knoten für Mykologie aufgebaut wird (Triebel et al. 2006).10 Der Knoten bietet inhaltlich und technisch Hilfe bei der Erschließung von mykologischen Datenbeständen in Deutschland und deren Anbindung an GBIF. Eine Reihe von Datenerzeugern nutzt dabei den oben geschilderten zentralen Service zum Datenfluss und nimmt an Workshops zur Weiterbildung in diesem Kontext teil. Die SNSB fungieren also seit 2006 als einer von rund 300 GBIF-Datenprovidern für Beobachtungs- und Sammlungsdaten, deren Daten über eine Wrapper-Software an das GBIF-Portal¹¹ in Kopenhagen angebunden sind. So können sie dort abgefragt oder zu Analysezwecken auf den lokalen PC des Anwenders heruntergeladen werden.

Zurzeit sind an den SNSB 133 000 Datensätze von etwa 9500 mykologischen Taxa für GBIF freigegeben. ¹² Etwa 107 000 davon sind georeferenziert, d. h. zumeist mit WGS84-Punktkoordinaten versehen. Etwa 53 000 Datensätze dokumentieren Objekte der Pilzsammlungen an der Botanischen Staatssammlung München, für den restlichen Datenbestand (ca. 80 000) fungieren die SNSB als GBIF-Datenprovider von Daten zu Material aus Sammlungen in Berlin, Halle, Görlitz, Greifswald, Bayreuth und St. Petersburg. Über das GBIF-Portal sind zurzeit 177 448 319 Datensätze (taxonomische Namen, Beobachtungs- und Sammlungsdaten) mittels Abfrageschnittstellen zugänglich (Stand: 8. Juli 2009).

Beobachtungsdaten und ökologische Daten im IBF-Projekt

Ökologische Daten und Daten zum Artenmonitoring im Gelände sind erst seit kurzem im Fokus nationaler und internationaler Projekte zur Etablierung von Standards sowie nachhaltiger Datenflüsse und Datennetzwerke. Daher hat das seit Januar 2009 durch die DFG geförderte Biodiversitätsinformatikprojekt »Aufbau eines Informationsnetzes für Biologische Forschungsdaten von der Erhebung im Feld bis zur nachhaltigen Sicherung in einem Primärdatenrepositorium« (IBF) Pilotcharakter.¹³ Es handelt sich hierbei um ein Verbundprojekt mit vier Partnern, koordiniert am IT-Zentrum der Staatlichen Naturwissenschaftlichen Sammlungen Bayerns (Triebel et al. 2009). Seitens der Universität Bayreuth beteiligen sich der Lehrstuhl für Angewandte Informatik IV sowie das Zentrallabor für DNA-Analytik und Ökoinformatik. Die Universität Regensburg ist mit dem Lehrstuhl für Botanik vertreten.

Die biologischen Daten werden im Forschungsprojekt IBF mittels Personal Digital Assistant (PDA) oder *Smartphone* mit GPS-Funktionalität und der Option der digitalen Bildaufnahme im Gelände erhoben. Optimierte Nutzeroberflächen sollen dazu Standardlisten wie ökologische Faktoren und taxonomische Namensbestände sowie topografische Kartenausschnitte anzeigen. Hierfür wird am Lehrstuhl für Angewandte Informatik IV der Universität Bayreuth eine Software zur Datenerhebung und -anzeige sowie eine Synchronisierungsschnittstelle über mobile Netzwerke (Mobilfunk, WLAN etc.) zur Datensynchronisation und zur Online-Suche in Taxonomiedatenbanken entwickelt (Abb. 7; Jablonski et al. 2009). Die Client-Software *DiversityMobile* versteht sich als Teil einer virtuellen Arbeitsumgebung des Diversity-Workbench-Frameworks und kommuniziert mittels Synchronisationsdatenbank mit einem institutionellen Datenrepositorium, d.h. einem digitalen Archiv, wie es für Forschungsdaten aus der Ökologie und Biodiversitätsforschung am SNSB IT-Zentrum geschaffen wird.

Die zwei biologischen Partner im IBF-Projekt verfolgen sehr unterschiedliche Forschungsansätze in den Bereichen Ökologie und Artenmonitoring. In der Arbeitsgruppe an der Universität Bayreuth werden seit mehreren Jahren ökologi-

⁹ http://www.gbif-mykologie.de/Datenbanken. html

¹⁰ http://www.gbif-mykologie.de

¹¹ http://www.gbif.org

¹² http://data.gbif.org/datasets/provider/158

¹³ http://www.diversitymobile.net



Abb. 7. Mobiler Client DiversityMobile zur Dateneingabe im Gelände (Beispiele).

sche Erhebungen zu spezialisierten herbivoren Insekten auf einer Langzeitversuchsfläche im Ökologisch-Botanischen Garten der Universität Bayreuth durchgeführt (zu der Versuchsfläche siehe Abb. 8; Kehl et al. 2008). Damit steht eine umfangreiche Sammlung von Geländedaten zur Verfügung, die es erlaubt, die Funktionalität des DiversityMobile-Clients und den Datenfluss zum Datenrepositorium, speziell für ökologische Daten, zu testen und entsprechend den Erfordernissen für die Datenerhebung im Gelände zu optimieren. Zukunftsweisend erscheint hier die Möglichkeit, durch den Einsatz von PDAs mit komplex strukturierter Client-Software bereits im Gelände mehrdimensionale ökologische Beziehungen im zeitlichen und räumlichen Kontext zu erfassen.

Die Zentralstelle der floristischen Kartierung Bayerns an der Universität Regensburg koordiniert ehrenamtliche Experten, die das Artenmonitoring von Pflanzen und Pilzen in Bayern durchführen.¹⁴ Über diese Zentralstelle gelangen die Daten an das Landesamt für Umwelt bzw. das Bundesamt für Naturschutz. Ziel des IBF-Projektes ist es, mittels Einsatz des *Smartphones* eine Beschleunigung der Erfassung der Kerndaten zu

den Einzelbeobachtungen im Bereich Monitoring zu erreichen, sowie zusätzlich die Nachhaltigkeit der erfassten Daten zu sichern. Dazu sollen vor allem GPS-Ortskoordination (WGS84 mit Höhenangabe), taxonomische Namen anhand von Referenzlisten, Datum, Beobachter und optional ein oder mehrere Bilder zu einer Beobachtung aufgenommen werden und diese Daten dann über nachhaltige technische Strukturen in einen dauerhaften Datenfluss integriert werden. Dazu werden in den Vegetationsperioden 2009 und 2010 im Raum Regensburg etwa 100 Begehungen mit PDAs durchgeführt, um die Client-Software DiversityMobile bei der Erfassung von Gefäßpflanzen, Pilzen und Flechten zu testen.

Parallel dazu werden am SNSB IT-Zentrum die für das IBF-Projekt relevanten *Diversity-Workbench*-Datenbankkomponenten optimiert, neue Komponenten implementiert sowie relevante Datenbestände wie die taxonomischen Referenzlisten für Pilze, Flechten und Blütenpflanzen in Deutschland oder die Listen der Vegetationseinheiten in Mitteleuropa in die virtuelle Arbeitsumgebung integriert. Abbildung 9 verdeutlicht das Datennetzwerk, das im Rahmen von IBF entstehen soll. Es soll einen durchgängigen Datenfluss von der Datenerhebung im Gelände bis zum Datenanwender sowie dem Datenrepositorium etablieren. Die im Gelände erhobenen Kerndaten können

¹⁴ http://www.bayernflora.de

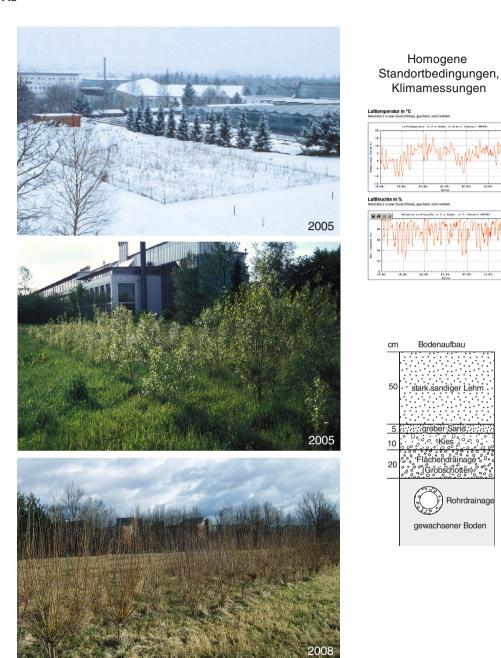


Abb. 8. Dauerversuchsfläche des Ökologisch-Botanischen Gartens der Universität Bayreuth: Die ökologischen Daten werden mit dem DiversityMobile-Client erhoben.

dabei vom Projektwissenschaftler mit Zugangsberechtigung unter Nutzung der verschiedenen lokalen Diversity-Workbench-Datenbankapplikationen ergänzt werden. Um das Projektziel zu erreichen, soll an den Staatlichen Naturwis-

senschaftlichen Sammlungen Bayerns für diese Bereiche ein institutionelles Datenrepositorium, d.h. ein digitales Archiv für Forschungsdaten aus der Ökologie und Biodiversitätsforschung geschaffen werden.

Rohrdrainage

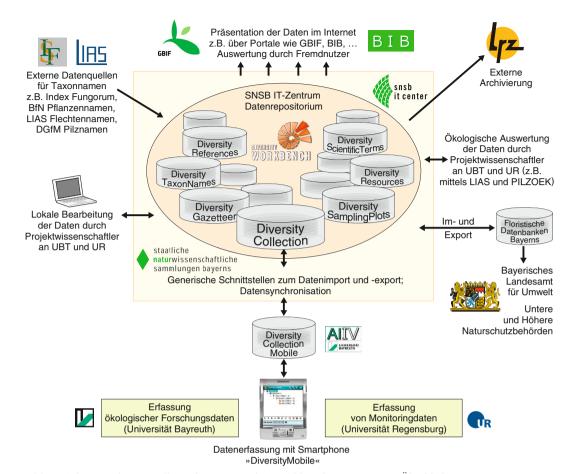


Abb. 9. Schematische Darstellung des IBF-Projektes und beteiligte Partner im Überblick.

(3) Aufbau eines institutionellen Datenrepositoriums

Das SNSB IT-Zentrum als Einrichtung der Staatlichen Naturwissenschaftlichen Sammlungen Bayerns besitzt, wie dargestellt, eine langjährige Erfahrung in der Prozessierung und Archivierung von Primärdaten aus dem Bereich der Biodiversitätsforschung und erfüllt damit die Voraussetzung, eine Rolle als institutionelles Datenrepositorium (d. h. als digitales Archiv) für biologische Forschungsdaten zu übernehmen. Derartige Infrastruktureinrichtungen ähneln digitalen Archiven großer Bibliotheken, jedoch ist die Komplexität der zu verwaltenden biologischen Daten um ein Vielfaches höher. Zu den unmittelbaren Aufgaben eines institutionellen Datenrepositoriums im biologischen Bereich gehören unter anderem: die Beschreibung von

Projektdatenbanken und -systemen, Erfassung und Hinterlegung von Metadaten; die Definition von Datenschemata dieser Datenbanken zur Etablierung von Datenflüssen von und zum Datenrepositorium; der Entwurf und die Implementierung von Schnittstellen; die Bereitstellung von Software und virtuellen Arbeitsumgebungen; die Sicherstellung der dauerhaften Archivierung und Prozessierung von Biodiversitätsdaten und ökologischer Daten aus Sammlungen und Forschung und die Etablierung eines Systems zur Zitierbarkeit von Datensätzen im Internet.

Öffentliche Sammlungen sollten in Zukunft darauf vorbereitet sein, neben den traditionellen Aufgaben zusätzliche Aufgaben im Bereich Biodiversitätsinformatik und Datenmanagement zu übernehmen, wie z. B. digitale Archive für biologische Daten zu etablieren und deren dauerhafte Prozessierung zu gewährleisten. Man denke an

den Einsturz des Historischen Stadtarchivs Köln Anfang März 2009, bei dem wertvolle historische Dokumente, von denen keine Digitalisate existierten, zerstört bzw. beschädigt wurden.

Gerade biologische Daten brauchen institutionelle Datenrepositorien oder digitale Archive, die, ähnlich wie wissenschaftliche Sammlungen, »leben«, und sich ständig technisch wie inhaltlich weiterentwickeln. Nur so kann einmal Erforschtes dokumentiert, datenkuratorisch gepflegt und nachhaltig gesichert, mit neu erhobenen Forschungsdaten zusammen prozessiert und analysiert werden und so für heute noch unbekannte Fragestellungen aus Wissenschaft und Gesellschaft dauerhaft nutzbar bleiben.

Danksagung

Der Aufbau der IT-Strukturen am SNSB IT-Zentrum wurde durch Mittel des Bayerischen Staatsministeriums für Wissenschaft, Forschung und Kunst, des Bundesministeriums für Bildung und Forschung, der Deutschen Forschungsgemeinschaft, der Andrew W. Mellon Foundation, von GBIF-International und der Europäischen Kommission (5. und 7. Rahmenprogramm) gefördert.

Für ihr anhaltendes Interesse am Thema Biodiversitätsinformatik und Datenmanagement an Sammlungen danke ich den Mitarbeitern der Botanischen Staatssammlung München und des SNSB IT-Zentrums. Ohne das große Engagement der beteiligten Entwickler und Forscher ist ein Erfolg von Projekten wie das IBF-Projekt nicht denkbar. Herr K. Baumann (Gomaringen), Frau Dr. A. Kehl, Herr B. Volz (beide Bayreuth), Herr Dr. H. Hilpert und Herr Dr. M. Weiss (beide München) haben Bildmaterial beigetragen bzw. grafisch neu gestaltet.

Abkürzungen

ABCD Standard: Standard zum Common Access to Biological Collections Data [Schema zum Austausch von biologischen Sammlungsdaten]

BfN: Bundesamt für Naturschutz

BIB: Botanischer Informationsknoten Bayern C# (gespr.: "C sharp"): [Programmiersprache]

CF: ColdFusion [Softwarepaket zur Erstellung von Webapplikationen]

DB: Database [Datenbank, organisierte Datensammlung, elektronisches Archiv]

DGfM: Deutsche Gesellschaft für Mykologie GBIF: Global Biodiversity Information Facility

GPS: Global Positioning System HTML: Hypertext Markup Language HTTP: Hypertext Transfer Protocol IBF: Aufbau eines Informationsnetzes für **B**iologische Forschungsdaten von der Erhebung im **F**eld bis zur nachhaltigen Sicherung in einem Primärdatenrepositorium

IT: Information Technology

JSP: Java Server Pages [Software zur dynamischen Erzeugung von HTML- und XML- Ausgaben eines Webservers]

LIAS: A Global Information System for Lichenized and Non-Lichenized **As**comycetes

LRZ: Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften

MS: Microsoft® Corporation

.NET (gespr.: "dotnet"): [Software-Plattform mit Sammlung von Klassenbibliotheken]

PDA: Personal Digital Assistant

PILZOEK: Datenbankanwendung für die Mykologische Standortkunde

REST: ReStructured Text [einfache Markup Language] SNSB: Staatliche Naturwissenschaftliche Sammlungen Bayerns

SOAP: Simple Object Access Protocol [Netzwerkprotokoll zum Austausch von Daten]

SPICE Standard [Standard und Datenaustausch-Schema der Catalogue-of-Life-Initiative]

SQL: Structured Query Language [Datenbanksprache]

UBT: Universität Bayreuth

UR: Universität Regensburg WGS84: World Geodetic System 1984

WLAN: Wireless Local Area Network XML: Extensible Markup Language

Literatur

Anonymous 2009. Scientific collections: mission-critical infrastructure for federal science agencies. A report of the interagency working group on scientific collections (IWGSC). – Office of Science and Technology Policy, Washington, DC, 49 pp. http://www.ostp.gov/galleries/NSTC Reports/Revision_1-22_09_CL.pdf

Bayerisches Staatsministerium für Wissenschaft, Forschung und Kunst (Hrsg.). 2009. aviso 2/2009: Unsterblichkeit. – Zeitschrift für Wissenschaft und Kunst. München: 3.

Hagedorn, G., M. Weiss & D. Triebel. 2008. *DiversityTaxonNames* information model (version 2.1, 20.2.2008). – http://www.diversityworkbench.net/Portal/TaxonNamesModel_v2.1

Jablonski, S., A. Kehl, D. Neubacher, P. Poschlod, G. Rambold, T. Schneider, D. Triebel, B. Volz & M. Weiss. 2009. DiversityMobile – Mobile Data Retrieval Platform for Biodiversity Research Projects. – In: Informatik 2009 – Im Focus das Leben. Beiträge der 39. Jahrestagung der Gesellschaft für Informatik e. V. (GI). GI-Edition: Lecture Notes in Informatics (LNI), Proceedings, 154: 27, 610-624.

- Kehl, A., G. Aas & G. Rambold. 2008. Genotypical and multiple phenotypical traits discriminate *Salix* × *rubens* Schrank clearly from its parent species. Plant Systematics and Evolution, 275 (3-4): 169-179.
- Thiers, B. M. 2009 [continuously updated]. Index Herbariorum: A global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium. http://sweetgum.nybg.org/ih/
- Triebel, D. 2007. II. Biodiversity and conservation. Fungal herbaria knowledge and memory institutions for collecting and processing mycological data. In: Anonymous (ed.): 1st World Conference on the Conservation and Sustainable Use of Wild Fungi. Technical Presentation and Posters. Cordoba, 40 pp.
- Triebel, D., R. Agerer, D. Begerow, H. Boyle, G. Deml, G. Hagedorn, F. Oberwinkler, G. Rambold, M. Schnittler, R. Scholz & H. J. M. Sipman. 2006. Aufbau des deutschen GBIF Knotens für Mykologie: Schlussbericht. München, http://www.gbif-mycology.de/PDFs/Schlussbericht_BMBF01LI0202.pdf

- Triebel, D., D. Peršoh, T. H. Nash III, L. Zedda & G. Rambold. 2007. Chapter 8. LIAS an interactive database system for structured descriptive data of Ascomycetes. In: Currey, G. B. & C. J. Humphries (eds.): Biodiversity databases. Techniques, politics, and applications. The Systematics Association, Special 73: 99-110.
- Triebel, D., W. Ahlmer, A. Bresinsky, O. Dürhammer, S. Jablonski, A. Kehl, J. Klotz, P. Poschlod, G. Rambold, B. Volz & M. Weiss. 2009. Developing a sustainable working platform for gathering biological data in the field. In: Anonymous (ed.): e-Biosphere 09 Conference, 1.-3.6.2009, London, UK, Abstracts. London, http://www.e-biosphere09.org/Final-AbstractVolume.pdf
- Weiss, M., G. Hagedorn & D. Triebel. 2007. *DiversityCollection* information model (version 2.05). http://www.diversityworkbench.net/Portal/Collection-Model v2.05

Diskussion

- G. Nemrava: Sie haben erwähnt, dass die Botanische Staatssammlung mit jährlich etwa 5000 neuen Sammlungsdaten ergänzt wird. Reichen die Kriterien in den Thesauri wie z.B. Taxonomie und Morphologie aus, um Redundanzen zu verhindern? Was sind das für Objekte? Handelt es sich um komplett neu entdeckte Pilze oder um bereits bekannte?
- D. Triebel: Die Botanische Staatssammlung wächst jährlich um ca. 5000 Pilzobjekte an, vielfach neue oder für unsere mykologische Sammlung neue Arten, auf jeden Fall aber mit neuen Funddaten, welche nicht redundant sind. Eine Digitalisierung der Sammlungsobjekte hängt allerdings generell von der Finanzierung durch entsprechende Projekte ab und steht noch am Anfang. Das Pilzherbarium in Görlitz kann als Optimalfall angesehen werden; es hat inzwischen seine gesamten alten Datenbestände unter Verwendung des Sammlungsmanagement-Systems DiversityCollection digitalisiert. In diese Datenbank werden auch, sobald neue Objekte eintreffen, die dazugehörigen Daten unmittelbar aufgenommen. Auch das technische Management wie das Erstellen von Etiketten und Umschlagsbeschriftungen wird über das Datenbanksystem abgewickelt. Erst danach werden die Pilzbelege in die Sammlung
- eingeordnet. An der Botanischen Staatssammlung München werden dagegen immer noch Teilsammlungen aus den bereits integrierten Beständen bearbeitet. Dabei werden thematisch ausgewählte Sammlungsteile, zum Beispiel die Sammlung von Mehltaupilzen, digitalisiert, und dies meist ehrenamtlich oder im Rahmen externer Finanzierung dieser Aktivitäten. Die Botanische Staatssammlung arbeitet an der Digitalisierung ihrer Herbarbestände seit etwa 1999 und hat jetzt ungefähr 2 % der Sammlungen in ihrem Datenbanksystem erfasst.
- **F. Oberwinkler:** Angesichts der enormen Datenflut bin ich froh, dass es Expertise und Kompetenz von Personen gibt, die diese »Daten-Sintflut« bewältigt.
- G. Nemrava: Ich glaube nicht, dass man vor der Datenflut Angst haben muss, denn das hieße ja, dass plötzlich eine große Masse gleichzeitig auf den Menschen eintrifft. Das Merkmal der Datenverarbeitung ist ja, dass man Daten innerhalb der ganzen Komplexität punktuell abfragt. Die Daten sind potentiell ja schon vorhanden und werden nicht in ihrer ganzen Komplexität und Totalität aktuell.

D. Triebel: Bisher sind die meisten mit Sammlungsobjekten verknüpften Daten noch nicht erhoben. Es könnte jedoch durchaus sein, dass in den nächsten Jahren auch noch eine weitere Flut von Daten auf die institutionellen Datenrepositorien zukommt. Es gibt nämlich viele sehr gute, bereits digitalisierte Datensammlungen, sei es von Hobby-Mykologen zu ihren Privatherbarien oder seien es alte Sammlungen von Forschungsdaten, die irgendwo in den Schränken lagern (die Daten von DFG-Projekten müssen zum Beispiel zehn Jahre vorgehalten werden). Wenn diese Daten in den alten Systemen technisch gefährdet erscheinen und an Datenrepositorien geliefert werden, müssen sie in großer Menge und zum Teil mit großem Aufwand erschlossen werden, d.h. in Repositoriendatenbanken importiert, prozessiert und archiviert werden.

F. Oberwinkler: »Erschlossen werden« bedeutet ja, die absolute Kompetenz mitzubringen, dass das, was in die Datenbank eingetragen wird, auch wirklich richtig ist. Wie organisieren Sie es, dass die Kompetenz der Bearbeitenden die Richtigkeit der Daten garantiert?

D. Triebel: Die Originaldaten sollten so lange wie möglich in der Verantwortlichkeit des Forschers bzw. des Datenerzeugers und Bearbeitenden bleiben, bevor und selbst auch nachdem sie in einer Repositoriendatenbank liegen. Das Optimale wäre dazu eine konkrete Zusammenarbeit zwischen institutionellen Datenrepositorien und Forschern während der laufenden wissen-

schaftlichen Arbeiten, um die Datenqualität zu erhöhen und auch um die Metadaten von den Datenbankbesitzern direkt abzufragen und in die Repositoriensysteme zu übernehmen. Die Daten werden auch nach Import in eine Repositoriendatenbank nicht gleich öffentlich gemacht, sondern erst, wenn sie publiziert sind und/oder vom Dateneigner freigegeben werden. Dies kann der Dateneigner zum Beispiel in den Komponenten der Diversity Workbench für jeden einzelnen Datensatz mit dem Setzen eines entsprechenden Flags tun. Die Kooperation mit den Forschern sollte also noch während der Laufzeit eines Forschungsprojektes einsetzen, um die Oualität zu optimieren und den Datenfluss in Hinblick auf die Dauerhaftigkeit der Datenbestände zu etablieren. Auch hier denke ich, dass öffentliche naturwissenschaftliche Sammlungen wie zum Beispiel die Staatlichen Naturwissenschaftlichen Sammlungen Bayerns der richtige Ort für derartige Datenrepositorien in der Biologie sind, weil sie die jahrhundertealte Erfahrung in der Kooperation mit den Forschenden aus dem Bereich Biodiversitätsforschung und Ökologie besitzen, allein schon wegen der wissenschaftlichen Praxis, die Forschungsgegenstände, biologische Objekte, in öffentlichen Sammlungen zu hinterlegen.

F. Oberwinkler: Es geht dabei auch um die weitere Ausbildung der neu hinzukommenden, jungen Forschergeneration im Bereich der Mykologie, sodass sie die Qualitäten mitbringen, mit den biologischen Objekten qualitativ richtig umgehen zu können.

Rundgespräche der Kommission für Ökologie

Herausgegeben von der Bayerischen Akademie der Wissenschaften

Bestellungen an:

Verlag Dr. Friedrich Pfeil, Wolfratshauser Straße 27, D-81379 München Tel.: +49 (0)89 742827-0 · Fax: +49 (0)89 7242772 · E-Mail: order@pfeil-verlag.de www.pfeil-verlag.de