

## **AnnoSys: Annotationen als Evaluationswerkzeug für die Nutzung virtueller Belegdaten in der Umweltforschung**

*Wolf-Henning Kusber, Okka Tschöpe, Lutz Suhrbier, Sascha Gennrich, Anton Güntsch & Walter G. Berendsohn*

Botanischer Garten und Botanisches Museum Berlin-Dahlem, Freie Universität Berlin, Königin-Luise Str. 6-8, 14195 Berlin, annosys@bgbm.org

**Keywords:** Annotationen, Biodiversitätsdaten, BioCASE, Datenstandards, Datenqualität

### **Einleitung**

Annotationen sind ein wichtiger Qualitätskontroll-Mechanismus in der Biodiversitätsforschung. Dies gilt insbesondere im Bereich der sammlungsbezogenen Forschung, die sich v.a. mit der Umgrenzung und Verbreitung von Arten befasst. Traditionell besuchen Wissenschaftler naturkundliche Archive, Sammlungen und Herbarien, revidieren und bestimmen Belege aus drei Jahrhunderten, entziffern und ergänzen handschriftliche Etiketten, notieren ihre Recherche-Ergebnisse auf Annotations-Etiketten und hinterlegen diese Zettel bei den Sammlungsbelegen.

### **Das Problem**

In einer zunehmend digital vernetzten Welt stehen Biodiversitätsdaten in großen Mengen über Biodiversitätportale wie GBIF (Global Biodiversity Information Facility, <http://www.gbif.org>) und BioCASE (Biological Collection Access Service for Europe, [www.biocase.org](http://www.biocase.org)) zur Verfügung. Das Problem bei der wissenschaftlichen Nachnutzung dieser Daten besteht in der notwendigen Qualitätsprüfung dieser sehr inhomogenen Daten. Bisher blieben fehlerhafte Daten häufig vergleichsweise lange im Internet unkorrigiert zugänglich, da es in der Vergangenheit kaum befriedigende Feedback-Mechanismen vom Datennutzer zum Datenhalter gab. Idealerweise sollten die Datenhalter jedes Feedback in die Datenbank zurückschreiben können, die Änderungen mit den physischen Belegen verknüpfen und die geänderten Daten an die Biodiversitäts-Netzwerke weitergeben. Dies ist aber oft nicht möglich, da den Datenhaltern dafür das Personal fehlt, oder da es sich um statische, nicht mehr veränderbare Datenbestände handelt (z.B. solche, die aus der Literatur extrahiert wurden).

### **Die Problemlösung**

Als Reaktion auf den dringenden Bedarf der Scientific Community und aufbauend auf den Erfahrungen mit einem Pilot-Annotationssystem (Güntsch et al. 2009, Kusber et al. 2009) wurde 2014 nach 3-jähriger Entwicklungszeit die erste Ausbaustufe des AnnoSys-Annotationssystems für digitalisierte Biodiversitätsdaten veröffentlicht und vollständig dokumentiert (Tschöpe et al. 2012-, 2013, 2014).

Kernstück ist ein Annotationsdatenrepositorium für vernetzte, hochkomplexe wissenschaftliche Biodiversitätsdaten (Kusber et al. 2014), das am BGBM Berlin-Dahlem vorgehalten wird. AnnoSys

ist seit April 2014 in mehreren Sammlungs- und Datenportalen für alle Biodiversitätsdaten, die im aktuellen XML-basierten ABCD-Standard (Access to Biological Collection Data, Berendsohn 2007, Holetschek et al. 2012) zur Verfügung stehen, im Einsatz. In den Portalen wird bei den dargestellten Daten ein Annotations-Button angezeigt, der die Nutzer zur Benutzeroberfläche von AnnoSys leitet (Tschöpe et al. 2012-). Nach Anmeldung wählt der Nutzer zunächst einen Annotationstyp aus (Tabelle 1) und kann dann den ausgewählten Datensatz annotieren, d.h. für die Werte jedes Eintrags können individuell Aktualisierungen und Ergänzungen vorgeschlagen und Kommentare eingefügt werden (Abb. 1).

Element	Original value	New value	Comment	Expectation
det./rev./conf		rev.	ex icon	Add
Family		Catenulaceae	BRM classification	Add
Full scientific name	Amphora libyca Ehrenb.	Amphora copulata (Kütz.) Schoen	see Hofmann et al.	Update

**Abb. 1: Sechsspaltiger AnnoSys-Editor. Element: ABCD-Element; Original value: Datenbank-Eintrag zu diesem Element; Spalte mit Icons zur Auswahl der Aktion „löschen des Original-eintrags“ bzw. „korrigieren der Annotation“; New value: annotierter Eintrag zu diesem Element; Comment: Kommentar zum neuen Wert, der Referenzen und Begründungen enthalten kann; Expectation: automatisch von AnnoSys generierte Erwartung an den Datenhalter (Hinzufügen, Aktualisieren oder Löschen des Elementes).**



**Abb. 2: Ausgedrucktes Annotations-Etikett für die Hinterlegung der revidierten Daten in der Sammlung. Die AnnoSys-Annotation bezieht sich auf die Revision eines AlgaTerra-Datensatzes zu *Amphora libyca* Ehrenb. (siehe Abb. 1), der mit Bilddaten im GBIF.DE Algae & Protists Datenportal zugänglich gemacht wurde.**

Nutzer des Annotationssystems erhalten per E-Mail eine Bestätigung ihrer Annotation über das bilinguale Benachrichtigungs-System (DE/EN) von AnnoSys (siehe Gennrich 2014). Zusammen mit den Originaldaten im ABCD-Format werden alle Annotationen unter Verwendung des RDF-basierten W3C Open Annotation Data Model (Sanderson et al. 2013) im Annotationsdatenrepositorium gespeichert und dort sowie im entsprechenden Datenportal veröffentlicht.

Sollte bei mehreren Datensätzen derselbe Annotationsbedarf festgestellt werden (z. B. orthographische Fehler bei wissenschaftlichen Namen oder systematisch vertauschte Koordinaten), können Massenannotationen vorgenommen werden.

Ebenfalls über das AnnoSys-Benachrichtigungs-System können Datenhalter und technisch Verantwortliche des Datensatzes informiert werden. Interessierte Nutzer suchen Annotationen im AnnoSys User-Interface oder abonnieren Annotationen ihres Interessengebietes, z.B. zu Taxa oder bestimmten Institutionen. Kuratoren sind aufgerufen, über eine spezielle Kuratoren-Nutzeroberfläche vorliegende Annotationen zu bewerten, ihrerseits zu kommentieren und durch die Verknüpfung mit

den physischen Belegen für ihre Sammlung nutzbar zu machen (AnnoSys Annotations-Etikett siehe Abb. 2).

Das AnnoSys-Annotationssystem ist gegenwärtig über das Digitale Herbar des BGBM (Abb. 3) sowie mehrere Biodiversitätspportale der BioCASE-Portalfamilie zugänglich, z.B. das Portal GBIF-D Algae & Protists (<http://protists.gbif.de/protists>, siehe Kusber et al. 2013) und Portale zu weltweiten bzw. zu europäischen Datenbeständen sowie zur Flora Deutschlands. Die gegenwärtige Ausrichtung des Annotationssystems auf den ABCD-Standard (Holetschek et al. 2012) liegt darin begründet, dass die meisten Datenprovider aus Deutschland diesen Standard seit vielen Jahren erfolgreich nutzen und hierdurch nahezu alle Sammlungsdaten, alle Daten aus der floristischen Kartierung Deutschlands und neuerdings über 4 Millionen Datensätze aus dem Citizen-Science Bereich für Annotationen zugänglich werden. Der im Rahmen von GBIF-D entwickelte BioCASE-Monitor gibt einen tagesaktuellen Überblick über annotierbare BioCASE-Datensammlungen aus Deutschland, die dem ABCD-Standard folgen (Abb. 4, GBIF-D 2014, Glöckler et al. 2013).

**Tab. 1: Beschreibung derzeit unterstützter Annotationstypen im AnnoSys Annotations-Editor und ihre Relevanz für die ökologische Forschung.**

<b>AnnoSys Annotations-Typ</b>	<b>Beschreibung des Annotations-Typs</b>	<b>Relevanz für die ökologische Forschung</b>
Scientific Name	Wissenschaftlicher Name	Zeichengenaue Korrektheit wichtig für die Verknüpfung von Namens- und Faktendaten
Determination	Wissenschaftlicher Name mit Metadaten der Bestimmung (det.), Revision (rev.) oder Bestätigung (conf.): Person, die die Bestimmung vorgenommen hat, Bestimmungsdatum, Quelle für die Bestimmung, Kommentierung des Artkonzeptes	Kenntnis der taxonomischen Konzepte ist für vergleichende ökologische Auswertungen essentiell
Nomenclatural Type	Information zum Beleg, der das Taxon definiert, meist bei Auffindung von Typusmaterial in Sammlungen	Bei Abbildung des Typus für die Evaluation von Bestimmungen hilfreich
Gathering	Aufsammlungsdaten mit Sammelort, Georeferenz, Sammler, Sammelnummer, Sammeldatum	Georeferenz und Datum sind wesentliche Daten für das Vorkommen von Arten in Zeit und Raum
Record Basis	Beleg (Preserved Specimen), Observation, Fossil, Living (ex Situ lebender Organismus)	Die richtige Zuordnung ist notwendiges Such- und Sortierkriterium bei der Auswertung von Massendaten
Storage of Duplicates	Ort, an dem Duplikate von Belegen hinterlegt sind	gering
Other	Allgemeine Kommentare oder gleichzeitige Annotation von mehreren Annotationstypen	Abhängig von der jeweiligen Annotation

Annotationen können schon jetzt bei der Evaluation von Datensätzen für ökologische Auswertungen helfen, da sie zeitintensive Doppelarbeit verschiedener Projekte ersparen helfen. Fehlerkorrek-

turen und Hinweise auf inkonsistente Daten können mit anderen Wissenschaftlern geteilt werden, sodass für Auswertungen ungeeignete Datensätze aus Analysen ausgeschlossen werden können.

## Zusammenfassung und Ausblick

AnnoSys ist ein frei zugängliches Annotationssystem zur Qualitätsverbesserung von digitalisierten Biodiversitätsdaten, das auf dem „W3C Open Annotation“ Standard für die Speicherung und den Austausch von Annotationen basiert. Für die weitere Entwicklung dieses generischen Annotationssystems wurden drei prioritäre Bereiche identifiziert:

- Einbindung von AnnoSys in weitere Portale, z.B. in das globale GBIF-Datenportal für Biodiversitätsdaten
- Erweiterung der Annotationsmöglichkeiten auf weitere Sammlungsdatenstandards (u.a. DwC, DwC-A, ABCDDNA) und andere Typen von Biodiversitätsdaten (z.B. Arten-Checklisten, Rote Listen, Catalogue of Life)
- Synchronisierung der Annotationen mit anderen den W3C Open Annotation Standard nutzenden Plattformen, wie z.B. Filtered Push, zum Aufbau einer globalen Annotations-Infrastruktur

Wesentlich für die weitere Entwicklung und Erweiterung von AnnoSys ist die Rückmeldung aus der Scientific Community zu weiterem Bedarf, da AnnoSys auch für interdisziplinäre Initiativen und Projekte außerhalb des Bereichs von Taxon-basierten Beleg- und Observationsdaten angepasst werden kann.

The screenshot displays the AnnoSys web interface for a specimen record. At the top, there are logos for Freie Universität Berlin and BGHM, along with a navigation menu (Home, About us, Contact, Press, Imprint). Below the menu is a horizontal navigation bar with categories: VISIT US, WHAT'S ON, THE GARDEN, THE MUSEUM, and SCIENCE. The main content area shows the specimen name *Osmunda regalis* var. *crispa* and its family OSMUNDACEAE. It provides collection details: Germany: prope Tegel (Berlin), Leg.: F. W. H. A. von Humboldt s.n., Barcode: B -W 19503 -01 0, and Storage name: *Osmunda regalis* var. *crispa*. A digital image of the specimen is shown in a viewer window, with a label that reads "Mus. Bot. Berol." and "B-W 19503-01 0". To the right of the image, there is an annotation by Olka Tschöpe, dated 30 June 2014, with the type of annotation being "Gathering". Below the annotation, there is a "Willdenow folder" section with a thumbnail image of the folder.

**Abb. 3:** Digitalisierter Beleg von Alexander von Humboldt aus Berlin-Tegel, der den Typus von *Osmunda regalis* f. *crispa* Willdenow zeigt. Zu diesem Beleg existiert in der Datenbank des Herbarium Berolinense eine Nutzerannotation, die im Portal angezeigt wird und über das AnnoSys Annotationsrepositorium abrufbar ist.

Zu Abb. 3: Röpert D. (ed.) 2000-: Digital specimen images at the Herbarium Berolinense. - Published at <http://ww2.bgbm.org/herbarium/> (Barcode: B -W 19503 -01 0 / ImageId: 261335) [2014-09-25].

BioCASE Monitor Service 1.2  
developed by LIP! and others



Title	# Records Dec. 2010	# Current Records	Difference	FullScientificNameString	Useful Links
AlgaTerra (cached at 2014-10-16 12:51:00) renew	949	3127	+2178	date: 740 tax: 3127	View mapping 
Algaterra Types (cached at 2014-10-16 12:51:00) renew	7618	7867	+249	date: 337 tax: 185	View mapping 

Abb. 4: BioCASE Monitor-Service (GBIF.DE 2014).

## Danksagung

Die Deutsche Forschungsgemeinschaft fördert das Projekt „Ein generisches Annotationssystem für Biodiversitätsdaten“ im Rahmen des LIS-Programms, GBIF-D wurde unter dem BMBF-Förderkennzeichen 01 LI 1001 A-F unterstützt.

## Literatur

- Berendsohn, W.G. (ed.) (2007): Access to Biological Collections Data – ABCD. Published on the Internet <http://www.tdwg.org/activities/abcd/> [2014-10-22]
- GBIF.DE (2014): GBIF-D Statistik. Published on the Internet <http://www.gbif.de/statistik> [2014-09-25]
- Gennrich, S. (2014): Benachrichtigungen im semantischen Annotationssystem AnnoSys. - Bachelorarbeit am Institut für Informatik der Freien Universität Berlin, Arbeitsgruppe Software Engineering, Berlin, 40 S. Published on the Internet <http://www.inf.fu-berlin.de/inst/ag-se/theses/Gennrich14-anno-sys.pdf> [2014-09-25]
- Glöckler, F., Hoffmann, J. & Theeten, F. (2013): The BioCASE Monitor Service - A tool for monitoring progress and quality of data provision through distributed data networks. *Biodiversity Data Journal* 1: e968. DOI: 10.3897/BDJ.1.e968
- Güntsch, A., Berendsohn, W.G., Ciardelli, P., Hahn, A., Kusber, W.-H. & Li, J. (2009): Adding content to content – a generic annotation system for biodiversity data. *Studi Trent. Sci. Nat.* 84: 123-128.
- Holetschek, J., Dröge, G., Güntsch, A. & Berendsohn, W.G. (2012): The ABCD of rich data access to Natural History Collections. *Plant Biosystems* 146(4): 771-779. DOI: 10.1080/11263504.2012.740085
- Kusber, W.-H., von Mering, S. & Holetschek, J. (2014): Netzverfügbarkeit primärer Biodiversitätsdaten: Schritt für Schritt zur BioCASE-Anbindung limnologischer Datenquellen an das GBIF-Netzwerk. – Pp. 211-215 in: Deutsche Gesellschaft für Limnologie (DGL) ed.: *Erweiterte Zusammenfassungen der Jahrestagung 2013 (Potsdam-Berlin)*, Hardegsen.
- Kusber, W.-H., von Mering, S. & Jahn, R. (2013): GBIF-Dateninfrastruktur: Limnologische Beobachtungs- und Belegdaten publizieren, abfragen und analysieren. – Pp. 485-489 in: Deutsche Gesellschaft für Limnologie (DGL) ed.: *Erweiterte Zusammenfassungen der Jahrestagung 2012 (Koblenz)*, Hardegsen.
- Kusber, W.-H., Zippel, E., Kelbert, P., Holetschek, J., Güntsch, A. & Berendsohn, W.G. (2009): From cleaning the valves to cleaning the data: Case studies using diatom biodiversity data on the Internet (GBIF, BioCASE). *Studi Trent. Sci. Nat.* 84: 111-122.
- Sanderson, R., Ciccacese, P., Van de Sompel, H. (ed.) (2013): W3C Open Annotation Data Model - Community Draft, 08 February 2013. Published on the Internet <http://www.openannotation.org/spec/core/> [2013-11-26]
- Tschöpe, O., Kusber, W.-H. & Suhrbier, L. (2014): AnnoSys User Guide. - Botanischer Garten und Botanisches Museum Berlin-Dahlem (BGBM), Zentraleinrichtung der Freien Universität Berlin, Biodiversity Informatics Section. Published on the Internet <https://annosys.bgbm.fu-berlin.de/sites/default/files/Users%20Instructions.pdf> [2014-09-25]

Tschöpe, O., Macklin, J.A., Morris, R.A., Suhrbier, L. & Berendsohn, W.G. (2013): Annotating Biodiversity Data via the Internet. *Taxon* 62(6): 1248-1258. DOI: <http://dx.doi.org/10.12705/626.4>

Tschöpe, O., Suhrbier, L., Güntsch, A. & Berendsohn, W.G. (2012-): AnnoSys Website. Published on the Internet <https://annosys.bgbm.fu-berlin.de> [2014-09-25]