Annotating Scientific Collection data using AnnoSys – A Generic Online Annotation System and Repository

Okka Tschöpe, Lutz Suhrbier, Anton Güntsch, Walter G. Berendsohn
https://annosys.bgbm.org | annosys@bgbm.org

Botanic Garden and Botanical Museum Berlin
Freie Universität Berlin, Königin-Luise Str. 6-8, 14195 Berlin (Germany)

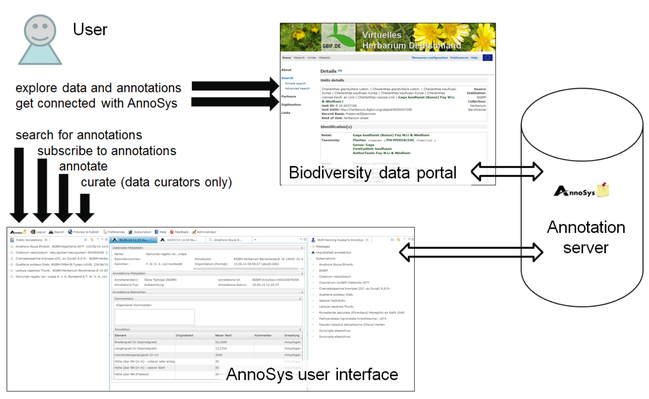

Fig. 1. **User interactions between biodiversity data portals and AnnoSys** (double arrow = data flow, single arrow = user action)

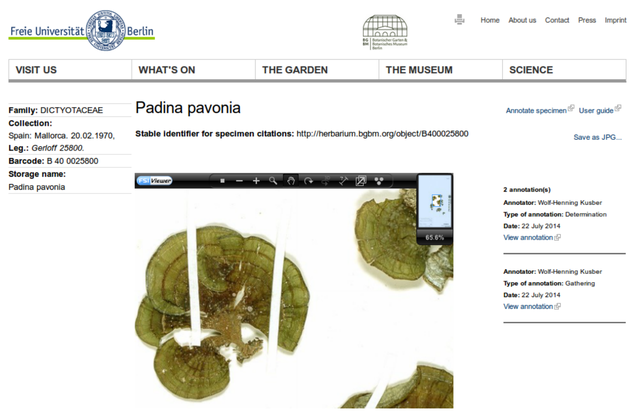

Fig. 2. **Digital specimen with two AnnoSys annotations** (Virtual Herbarium of Herbarium Berolinense (B), http://ww2.bgbm.org/Herbarium/specimen.cfm?Barcode=B400025800)

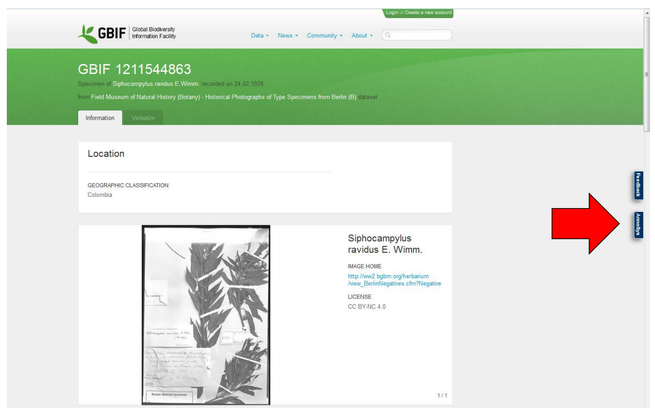

Fig. 3. **AnnoSys integration in the GBIF data portal** (Red arrow indicating the interface to the AnnoSys system), http://www.gbif.org)

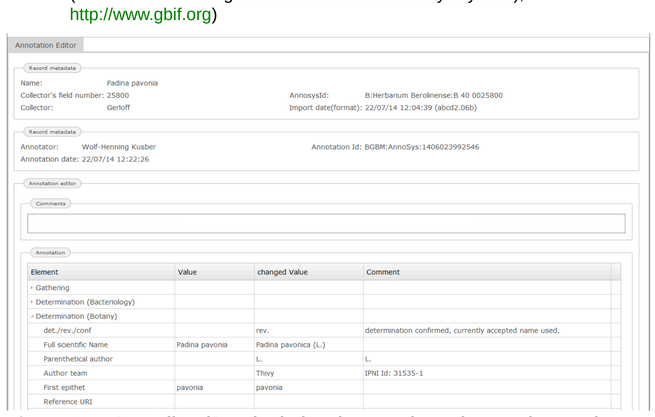

Fig. 4. **AnnoSys Editor (Standard View, improved version as of November 2016): Annotation of a scientific name**

Fig. 5. **AnnoSys Editor (Expert View): Annotation of a scientific name and its classification** (showing an added ABCD element; marked line).

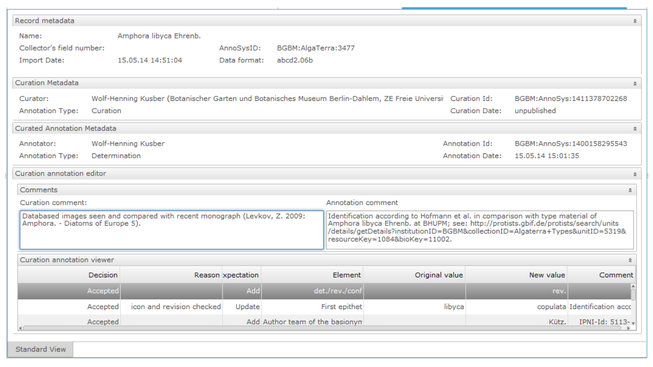

Fig. 6. **AnnoSys Curation: Annotation edited by a curator** (left rows show acceptance and evaluation of annotated values)

Introduction

AnnoSys is a web-based system for correcting and enriching biodiversity data in publicly available data portals. Using the example of digital specimen and observation data the current release of *AnnoSys* establishes a number of workflows which enable online annotations. Currently, it is integrated into a dozen biodiversity data portals including the Global Biodiversity Information Facility (GBIF http://www.gbif.org) and provides means to comfortably publish, search and manage annotations referring to these data.

The *AnnoSys* System

AnnoSys manages a repository for annotations and annotated records. Its desktop oriented web interface enables annotators and collection managers to work with annotations using community-specific criteria, such as scientific name, identification, or nomenclatural type. Fig. 1 depicts *AnnoSys*' basic annotation workflow.

Data portals that are connected to *AnnoSys* display available annotations concerning their records (Fig. 2) and redirect their users to the *AnnoSys*' web interface (Fig. 3), where existing annotations to a given specimen ca be examined or new annotations can be created (Fig. 4, 5). Collection managers can receive curatorial rights, which enable them to communicate decisions concerning the acceptance or rejection of annotations referring to data objects in their collections. (Fig. 6). A subscription based message system permits registered *AnnoSys* users and curators to get informed by email (Fig. 7) about any kind of annotation event via a self-defined subset of e.g. specific taxon names, countries, collections, or institutions.

An analysis of the existing annotations in *AnnoSys* shows that a wide range of use cases is covered (Fig. 8) and allows a first insight in the relevance of the system (Fig. 9) which is currently adapted to the user's needs.

While the *AnnoSys* approach is generic with regard to the usage of biodiversity related data formats and standards, the current implementation focuses on data based on the ABCD standard provided by the GBIF/BioCASE protocol interfaces and the DarwinCore standard.

Additional Features

- Integrated workflow for "batch" annotations
- Linked Data, REST and SPARQL based web services provide public access to any annotations including the annotated record data stored within the *AnnoSys'* repository
- Interfaces for rights management and system administration

2nd Funding Phase - Objectives

- Significant performance improvement
- Synchronising annotations with other platforms using the W3C Open Annotation standard like Filtered Push
- Supporting more data formats and use cases (e.g. checklists)
- Input verification through integration of results from test services (e.g. georeferencing, taxonomic thesauri)
- Evaluating the support of image annotations

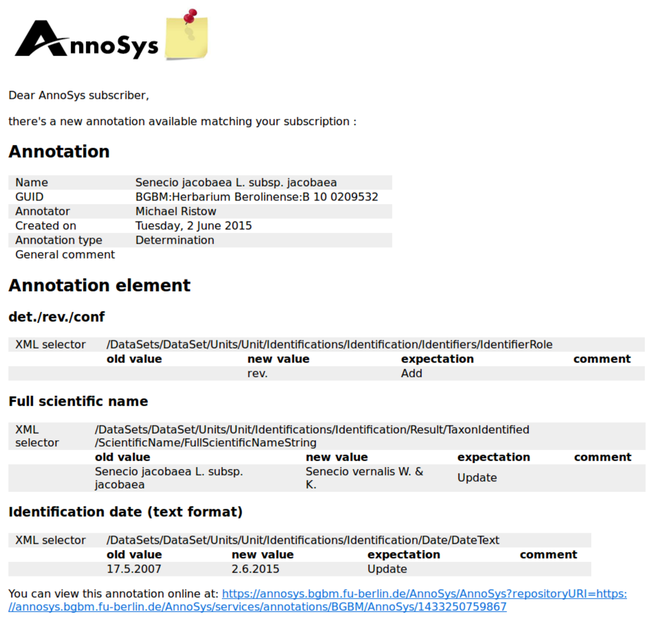

Fig. 7. **Annotation notification email to AnnoSys subscribers**

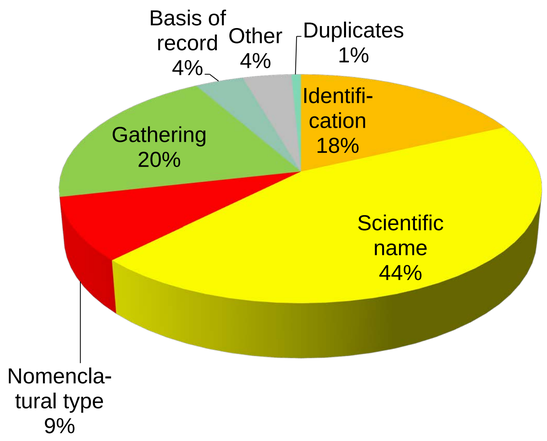

Fig. 8. **Percentages of annotation types, chosen by AnnoSys users** The main use cases are corrections and/or additions according to label data, user's expertise, external data sources or own scientific work

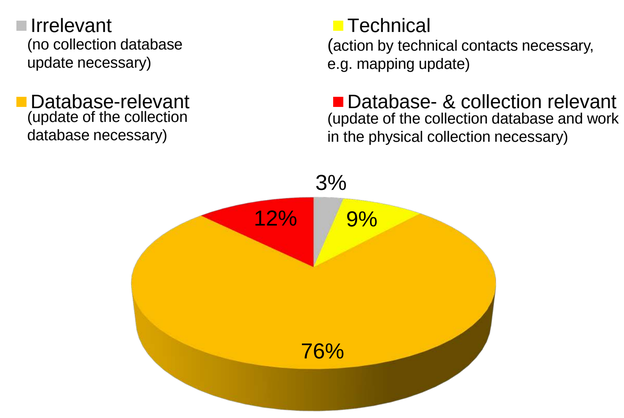

Fig. 9. **Analysis of the relevance of annotations in the AnnoSys system for data curators, curators of natural history collections and technical contacts of dataset**

Literature

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., Kusber, W.-H. & Suhrbier, L. 2015: *AnnoSys User Guide*. Botanic Garden and Botanical Museum Berlin-Dahlem (BGBM). Published on the Internet https://annosys.bgbm.org/sites/default/files/annosys-user-guide_v1-4_2015-06-30.pdf

Suhrbier, L.; Kusber, W.-H.; Tschöpe, O.; Güntsch, A. & Berendsohn, W. G. (accepted): *AnnoSys - Implementation of a generic annotation system for schema based data using the example of biodiversity collection data.*

Acknowledgement

The authors thank the German Research Foundation for financial support of the project "Ein generisches Annotationssystem für Biodiversitätsdaten" project numbers BE 2283/4-2 (2011-2014), BE 2283/4-2 (2015-2017).