AnnoSys – A generic annotation system for biodiversity data [Talk]

Okka Tschöpe
BGBM Berlin; o.tschoepe@bgbm.org

Lutz Suhrbier
BGBM Berlin; l.suhrbier@bgbm.org

Walter G. Berendsohn
BGBM Berlin; w.berendsohn@bgbm.org

Traditionally, natural history collection objects are annotated in written form and the annotations are added directly to the specimen. Specimen annotations, *e.g.* concerning their taxonomic identity, are a quality control mechanism that improves the scientific value of herbarium specimens. However, with data increasingly becoming accessible via the Internet, the traditional flow of annotation information is currently interrupted, because on-line users of data are not able to add annotations to the digital record of the specimen and reporting them back to the original source. Currently missing is a general online annotation system that ensures the continuation of the traditional data sharing and documentation of specimen data. AnnoSys therefore develops an annotation data repository for networked and highly complex biodiversity data available via the internet. It is based on a prototype developed in the context of SYNTHESYS, but uses the W3C Open Annotation Core Data Model and Extension Specification, based on the RDF data model to store annotation information. While AnnoSys principally is open to support any kind of biodiversity data, the focus of the prototype is on collection data in the botanic domain provided by the GBIF/BioCASe system. The Annotation System combines a desktop-like user application interface with the following components: A Repository that persistently saves all annotations together with the original XML-document the annotation refers to. This allows users to compare the original record with the annotated data at the point in time the annotation was created. The repository will also provide efficient query functions to search annotations using specific criteria like specific taxonomic groups or locations. The exchange component provides an interface for downloading the original XML documents from BioCASe data providers or other sources into the repository. It is also intended to provide services allowing external systems to search and access any data stored within that repository. A message system installs information flows to notify annotators, data providers and interested users about the ongoing state of annotation workflow processes. A security component is responsible for the administration of user accounts, including secure user authentication, authorization and access rights management.