

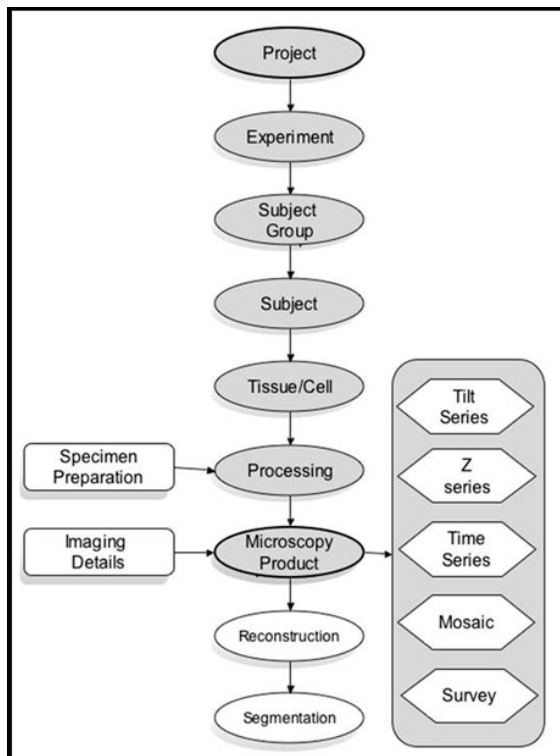
Datenbanken: CCDB und SAO

Datenbanken sind mittlerweile integraler Bestandteil der Datenverwaltung in der Biologie. Sie helfen bei der Suche nach relevanten Informationen, dienen dem Austausch in der Forschung und dem Data Mining. Da die Menge biologischer Daten stetig anwächst, insbesondere aufgrund der mannigfaltigen Abbildungstechniken und großen Datenmengen der Elektronentomographie, ist eine besondere Handhabung dieser Daten notwendig. Die Herausforderung dabei liegt in der Datenaufbereitung und Speicherung in geeignet indizierten und verknüpften biologischen Datenbanken, um eine einheitliche Struktur, leichtere Durchsuchbarkeit und Automatisierbarkeit von Analysen durch Software zu ermöglichen.

Die CCDB - **Cell Centered Database** ist ein Beispiel für einen Versuch in diese Richtung: Sie ist eine der ersten Online-Datenbanken für zelluläre Bilddaten und wurde im März 2002 vom National Center for Microscopy and Imaging Research (NCMIR) erstellt. Die CCDB macht hochauflösende 3D-Licht- und elektronenmikroskopische Aufnahmen für die Forschung zugänglich und enthält Daten auf der mesoskopischen Ebene, wie zelluläre Netzwerke oder molekulare Strukturen. Daten können über eine Web-Interface abgelegt und aufgerufen werden. Hierfür stellt die CCDB unter <http://ccdb.ucsd.edu> ein Webformular zur Verfügung.

Aufbau der CCDB: In der Datenbank werden die Daten mit Oracle 8i¹ (ab 2008 mit Oracle 10g + Java-Applikationen²) in einem objekt-relationalen Datenmodell dargestellt. Abbildung 1 zeigt die Hauptklassen von Informationen, die in den CCDB-Tabellen enthalten sind. Das Schema folgt sozusagen dem Vorbild eines Rekonstruktionsprozesses von 2-dimensionalen Mikroskopiebildern und ist in den Schritten vom Experiment bis zur Analyse organisiert.

Im Mittelpunkt des CCDB-Datenmodells liegt das „Microscopy Product“, das bezieht sich zum Beispiel auf ein Set zusammengehörender zweidimensionaler Bilder, die durch ein Licht- oder Elektronenmikroskop aufgenommen wurden.



Jedes „Microscopy Product“ stellt eine eindeutige Kennung für die Datenbank dar, deren Bilder systematisch zueinander zugeordnet werden. Zum Beispiel enthält eine so genannte „Tilt Series“ Bilder, die durch den Neigungswinkel zugeordnet werden. Die „Z series“ wird über den Wert der Tiefe zugeordnet, etc.

Jedes Mikroskopie Produkt kann dann eine oder mehrere Arten von Daten zusammen mit ihm speichern, wie zum Beispiel: 2D Bilder, Rekonstruktion und Segmentierung. Es gibt auch Bilder die anzeigen, wo die Daten zu lokalisieren sind, zum Beispiel mit Hilfe eines anatomischen Atlas oder einer Bild Karte. Das bietet einen breiteren Kontext. Jede der drei Haupt-Datentyp, stellt die CCDB in voller Auflösung als Datei zum Download zur Verfügung, darüber hinaus kann man sich Thumbnails eine Animation anzeigen lassen, was die Anzeige im Web vereinfacht.

Abbildung 1: CCDB-Architektur

¹ Vgl. Martone, Gupta, Wong, Qian, Sosinsky, Ludäscher, Ellisman (2002): S.

² Vgl. Martone, Tran, Wong, Sargis, Fong, Larson, Lamont, Gupta, Ellisman (2008):S.

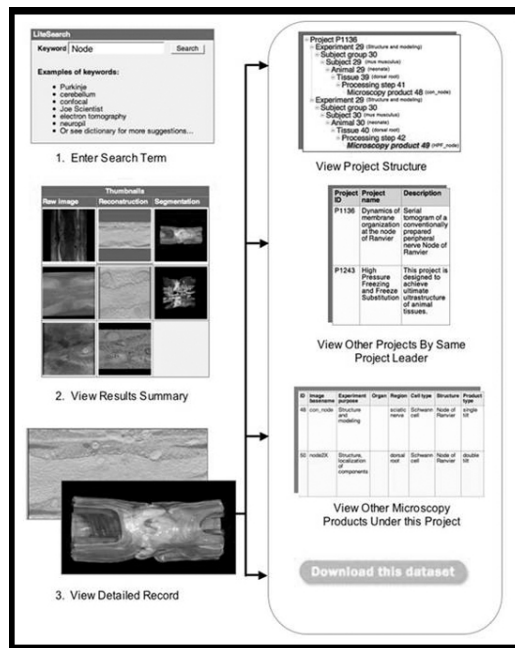


Abbildung 2: CCDB-Interface

kann man die Daten downloaden oder Sie unter den eigenen Dateien in einem Cache für den zukünftigen Download speichern. Darüber hinaus lässt die CCDB einen Datenupload zu: Die Dateneingabe erfolgt in einem sicheren Web-Portal und erfordert ein Konto für den Zugang. Sobald der Prozess der Datenerfassung abgeschlossen ist, wird der Datensatz zur Prüfung bei der CCDB eingereicht. Das CCDB prüft, ob die Eingaben richtig sind und gibt die Daten für den öffentlichen Zugang frei. Sollte der Datensatz fehlerhaft sein, gibt ihn die CCDB zur Korrektur an den Urheber zurück. Werden die Datensätze einmal veröffentlicht, bleiben die Rechte zwar beim Urheber, dieser ist jedoch nicht in der Lage die Datensätze zu löschen. Stellen sich Datensätze nach der Veröffentlichung doch als fehlerhaft heraus, werden sie entsprechend von den CCDB Kuratoren gekennzeichnet.

Die CCDB entwickelte auch eine weitere Datenbank namens **SAO - Subcellular Anatomy Ontology**. Das ist eine Ontologie-Datenbank, die hierarchische Strukturen beschreiben kann. Eines der wichtigsten Ziele des CCDB Projektes war es, die durch verschiedene Techniken und in verschiedenen Auflösungen erfassten Daten, in einem hierarchischen Modell der Zelle darzustellen. Die CCDB allein, kann das nicht leisten. Sie liefert lediglich Bilder der Experimente und Rekonstruktionsdetails der Daten. SAO hingegen kann mit Hilfe eines Sets von verschiedenen Konzepten oder Entitäten („ist ein“ oder „ist Teil von“) innerhalb eines Feldes die Beziehung von Zellkomponenten zueinander darstellen. Zum Beispiel: „Neuron ist eine Zelle“ oder „Zelle ist Teil der Plasma Membran“. Die Ontologien stellen also ein standardisiertes und maschinenlesbares Vokabular zur Verfügung, um Zuweisungen zu formulieren. Der SAO-Ansatz wurde allerdings nicht weiterverfolgt. Stattdessen wurde SAO in ein neues Projekt eingebunden: **The Gene Ontology Cellular Component Ontology (GO-CCO)**. Das GO-CCO integriert die Ontologien aus SAO und CCO in die Gene Ontology (GO) und fügte neue Begriffe hinzu, so dass GO-CCO entstand.

Die Datenbank erfordert für die Explorierung keinen Usernamen oder Passwort. Alle Daten, die eingegeben werden, sind öffentlich zugänglich. Nachdem ein Suchbegriff eingegeben wird, wird eine Zusammenfassung der Suchergebnisse angezeigt. Eine Reihe von Thumbnails demonstriert die Art der Daten (2D-Bild, Reconstruction oder Segmentierung), die für jeden Datensatz gespeichert wurden. Sobald der Nutzer einen Datensatz auswählt, wird eine detaillierte Aufzeichnung angezeigt, hier kann sich der Anwender Bilder mit höherer Auflösung, Animationen, Bildkarten und alle dazugehörigen Metadaten anzeigen lassen. Innerhalb einer detaillierten Aufzeichnung werden dem Benutzer mehrere Optionen für die weitere Suche oder den Download angeboten. So kann beispielsweise die Projektstruktur für das Projekt angezeigt werden, die u.a. zeigt, welche Datensätze von der gleichen Person verfasst wurden. Man kann andere Projekte von dem gleichen Verfasser oder andere Mikroskopie-Produkte unter dem gleichen Projekt betrachten. Und schließlich

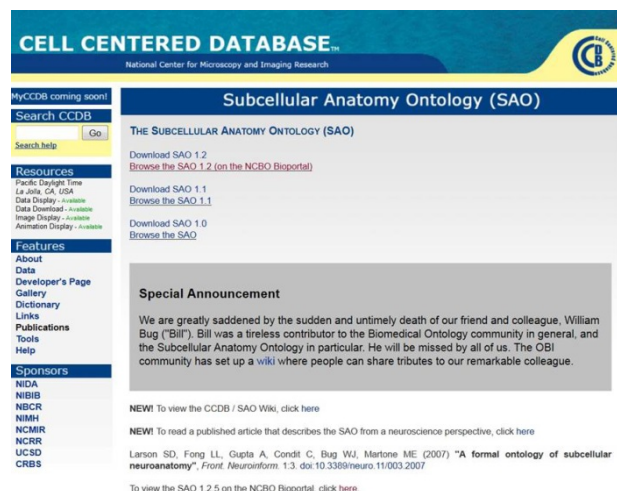


Abbildung 3: SAO-Interface

Die Ontologien stellen also ein standardisiertes und maschinenlesbares Vokabular zur Verfügung, um Zuweisungen zu formulieren. Der SAO-Ansatz wurde allerdings nicht weiterverfolgt. Stattdessen wurde SAO in ein neues Projekt eingebunden: **The Gene Ontology Cellular Component Ontology (GO-CCO)**. Das GO-CCO integriert die Ontologien aus SAO und CCO in die Gene Ontology (GO) und fügte neue Begriffe hinzu, so dass GO-CCO entstand.

Quellen:

Paper:

Martone ME, Gupta A, Wong M, Qian X, Sosinsky G, Ludäscher B, Ellisman MH (2002): A cell-centered database for electron tomographic data. Journal of structural biology. Online verfügbar: <http://www.ncbi.nlm.nih.gov/pubmed/?term=12160711> (12.05.2012)

Martone ME, Tran J, Wong WW, Sargis J, Fong L, Larson S, Lamont SP, Gupta A, Ellisman MH. (2008): The cell centered database project: an update on building community resources for managing and sharing 3D imaging data. Journal of structural biology, online verfügbar: <http://www.ncbi.nlm.nih.gov/pubmed/?term=18054501> (12.05.2014)

Roncaglia P1, Martone ME, Hill DP, Berardini TZ, Foulger RE, Imam FT, Drabkin H, Mungall CJ, Lomax J. (2013): The Gene Ontology (GO) Cellular Component Ontology: integration with SAO (Subcellular Anatomy Ontology) and other recent developments. In: Journal of biomedical semantics. Online verfügbar: <http://www.ncbi.nlm.nih.gov/pubmed/24093723> (13.05.2014)

Internet:

CCDB / SAO Website: <http://ccdb.ucsd.edu/CCDBWebSite/sao.html> (13.05.2014)

CCDB Website: <http://ccdb.ucsd.edu> (13.05.2014)

GO Website: <http://geneontology.org> (13.05.2014)

Abbildungen:

Abbildung 1: CCDB-Architektur <http://www.ncbi.nlm.nih.gov/pubmed/?term=12160711>

Abbildung 2: CCDB-Interface <http://ccdb.ucsd.edu>

Abbildung 3: SAO-Interface <http://ccdb.ucsd.edu/CCDBWebSite/sao.html>