

„DER KONTINUIERLICHE ANSTIEG VON DATENMENGEN IN DEN UMWELT- UND BIOWISSENSCHAFTEN ERFORDERT EIN UMFASSENDES MANAGEMENT BEREITS IN DER PLANUNGSPHASE VON FORSCHUNGSPROJEKTEN. DIE „GERMAN FEDERATION FOR BIOLOGICAL DATA“ (GFBio) BAUT HIERZU EINE INFRASTRUKTUR MIT PARTNERN IN GANZ DEUTSCHLAND AUF.“



Michael Diepenbroek

Orientierung im Datenmeer – Infrastruktur für Umwelt- und Ökosystemdaten

DOI: 10.1007/s12268-015-0601-z
© Springer-Verlag 2015

■ Innovative Forschung sowie neue Techniken im Bereich der Umweltwissenschaften und Ökosystemforschung generieren kontinuierlich eine Flut neuer Daten und bieten Forschern neue Möglichkeiten von diesen riesigen Datenmengen zu profitieren. Das Internet wird dabei verstärkt für Datensuche und Datenaustausch genutzt.

Die Verfügbarkeit von qualitativ hochwertigen und standardisierten wissenschaftlichen Daten in Datenbanken ist allerdings begrenzt. Ursache sind Defizite im technischen, methodischen, organisatorischen und auch kulturellen Bereich. So ist die meist nicht nachhaltige Finanzierung existierender Archive ein Problem, aber auch die Inkompatibilität vorhandener Standards verhindert die Datenintegration aus verschiedenen Quellen. Darüber hinaus wird die Verfügbarkeit bereits erhobener wissenschaftlicher Daten weiter signifikant verringert, wenn diese nicht in vorhandenen Infrastrukturen archiviert werden. Die Ursachen dafür sind zum einen Unkenntnis über bestehende Datenmanagement-Möglichkeiten, zum anderen die noch immer fehlende Anerkennung der dauerhaften Bereitstellung von qualitativ hochwertigen Daten als wichtiger Bestandteil der wissenschaftlichen Arbeit und Nutzen für die Gesellschaft.

Um sich diesen komplexen Herausforderungen zu stellen, bedarf es einer engen interdisziplinären Zusammenarbeit von Bio- und Umweltwissenschaftlern mit Informatikern und Datenmanagern für innovative und langfristige Lösungen. Das Ziel dabei sollte immer sein, Daten allen Nutzern und Interessenten öffentlich und langfristig zugänglich zu machen. Nur so ist es möglich z. B. Veränderungen in der Zusammensetzung von Tier- und Pflanzengemeinschaften, die etwa durch

den Klimawandel bedingt sind, über Jahrzehnte hinweg zu erforschen.

Um diese Entwicklung aktiv zu unterstützen, fördert die Deutsche Forschungsgemeinschaft (DFG) seit Ende 2013 das Infrastrukturprojekt „GFBio – German Federation for Biological Data“ (GFBio). GFBio ist ein Verbund von 19 Partnern aus ganz Deutschland, koordiniert an der Universität Bremen. Neben Universitäten bilden PANGAEA, sieben deutsche Museums- und Sammlungsarchive, sowie ausgewählte molekularbiologische Archive und Dienste den Verbund. GFBio baut eine Infrastruktur auf, die vorhandene molekulare und ökologische Daten aus bestehenden Dateninfrastrukturen standardisiert und in einem Portal (www.gfbio.org) verfügbar macht.

Vor allem aber will GFBio einen „Rundum-Service“ für die wissenschaftliche Datennutzung bieten. Dazu unterstützt GFBio die Wissenschaftler von der Projektplanung, der Datenerfassung und -visualisierung bis zur anschließenden Langzeitarchivierung in einem der angeschlossenen Datenzentren. Das aktuelle Portal mit dem Ansatz der Zusammenführung von Genom-, Umwelt- und Sammlungsdaten bietet eine Suche über 400.000 Datensätze inklusive über 50.000 Organismen und taxonomische Daten, die die eingebundenen Museen und Sammlungen zur Verfügung stellen. Im Juli 2015 hat die DFG weitere 5,8 Mio. € für die Fortsetzung von GFBio bewilligt. Diese Fördermittel ermöglichen die Weiterentwicklung des Datenportals im Bereich der Datenanalysen und Visualisierung, sowie die Entwicklung einer semantischer Suche, die das Auffinden und Integrieren qualitativ hochwertiger Daten aus den verschiedenen Archiven erleichtert. Um das Bewusstsein für Datenmanagement zu schärfen, bietet GFBio Kurse und einen erweiterten Helpdesk an. Durch die breite internationale Vernetzung der beteiligten

Partner ist eine Einbettung von GFBio in das internationale Umfeld und die dort genutzten Standards und maßgeblichen Datenmanagement-Strategien gegeben.

Mit seinen Aktivitäten will sich der Verband GFBio langfristig als die deutsche Anlaufstelle für Datenmanagement im Bereich Biologie und Umweltwissenschaften etablieren und effektiv zukünftige wissenschaftliche Fragestellungen unterstützen, indem er qualitativ hochwertige Daten schnell und flexibel bereitstellt. ■

Dr. Michael Diepenbroek
Koordinator der „German Federation for Biological Data“ (GFBio)

Korrespondenzadresse:

Dr. Michael Diepenbroek
MARUM – Center for Marine Environmental Sciences
Universität Bremen
Hochschulring 18
POP 330 440
D-28359 Bremen
Tel.: 0421 218-65590
Fax: 0421 218-65505
mdiepenbroek@pangaea.de