

Auch Daten brauchen eine Infrastruktur

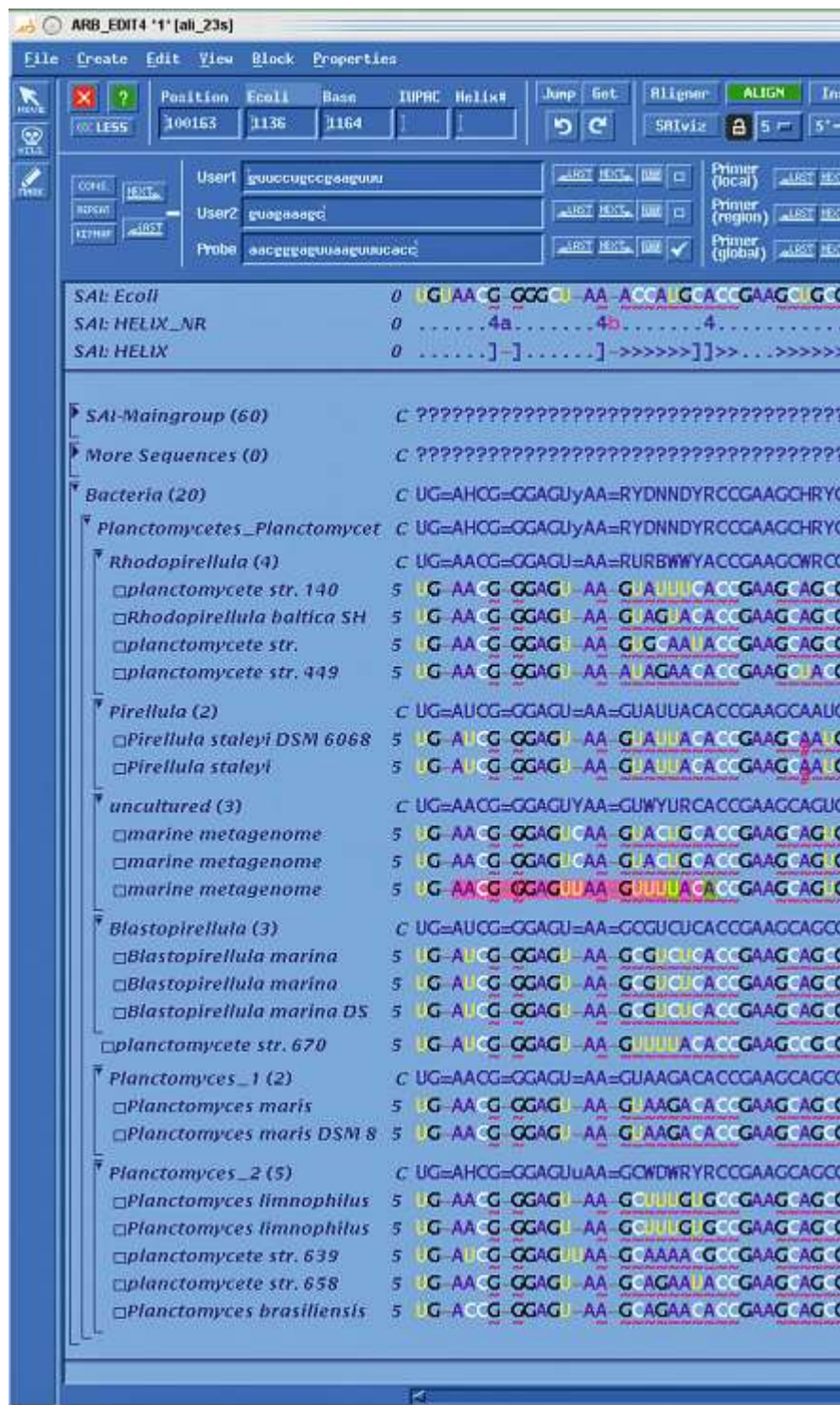
Die DFG will Forschungsdaten für die wissenschaftliche Nachwelt erhalten und zugänglich machen. Zehn Millionen Euro stellt sie dafür bereit. Ein Teil davon geht an die TUM.

Wie digitale wissenschaftliche Daten nachhaltig zu behandeln sind, ist in sehr vielen Fachgebieten noch unklar. Die meisten dieser Daten werden weder systematisch archiviert noch für eine Nachnutzung verfügbar gehalten. Dazu müsste die Wissenschaft vertrauenswürdige Datenrepositorien etablieren und klare Verantwortlichkeiten für das Management von Forschungsdaten definieren.

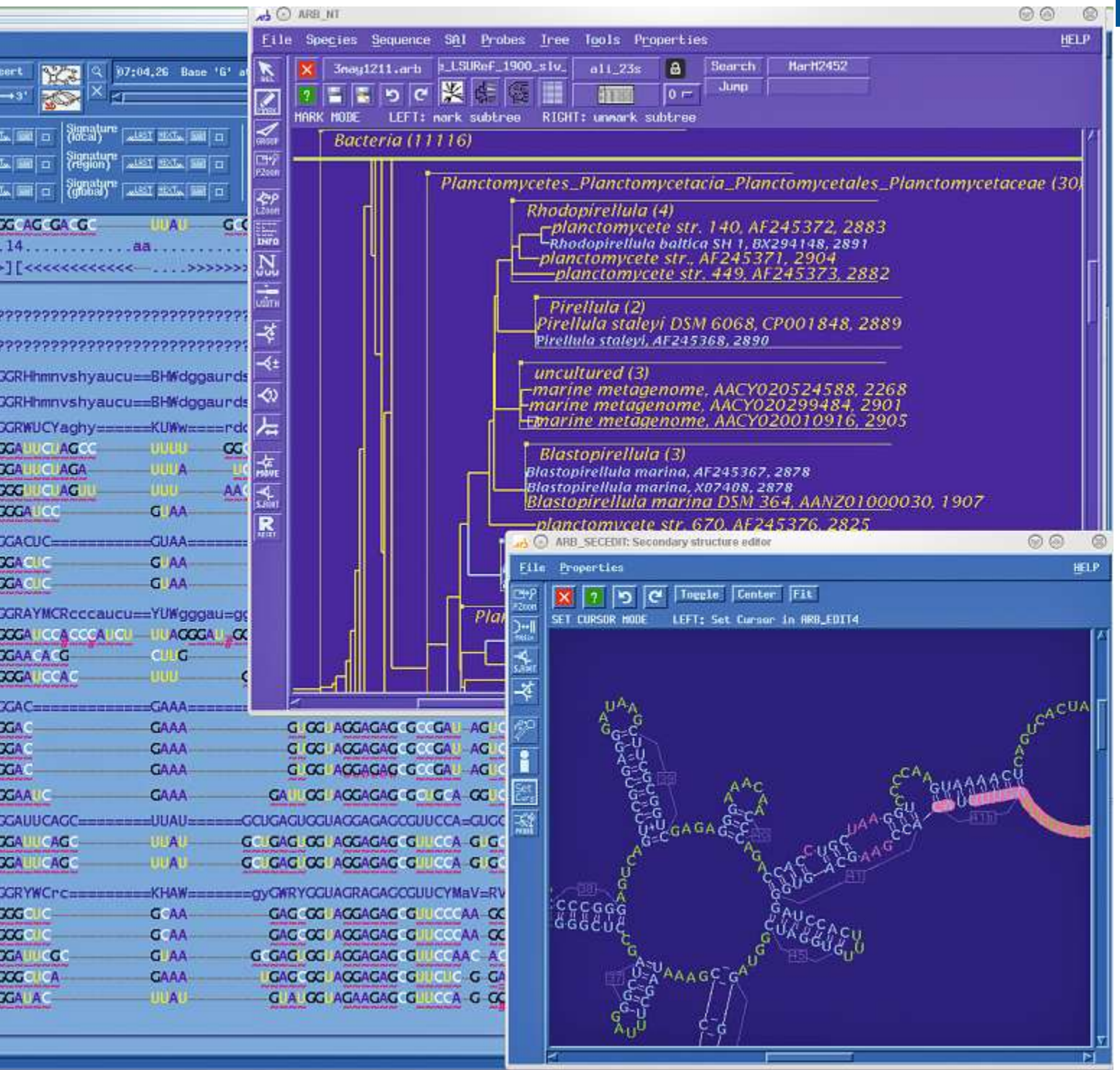
Deshalb startete die DFG Anfang 2010 die Ausschreibung »Informationsinfrastrukturen für Forschungsdaten«: Wissenschaftler sollten gemeinsam mit Bibliotheken, Archiven oder Rechenzentren Projekte initiieren zu Entwicklung und Aufbau fachspezifischer Organisationsformen sowie zur Professionalisierung bereits bestehender Forschungsdatenrepositorien. Ziel sind bedarfsgerechte und nachhaltige Informationsdienstleistungen, die den Zugang zu Forschungsdaten ermöglichen.

Unter den 27 ausgewählten und mit 9,9 Millionen Euro geförderten Projekten ist das Vorhaben von Dr. Wolfgang Ludwig. Der wissenschaftliche Mitarbeiter am TUM-Lehrstuhl für Mikrobiologie erhält 175 000 Euro für sein Projekt »ARB in the age of high throughput sequencing: adaption to the requirements of large scale environmental and metagenomic studies and maintenance of the respective databases«. Er gründete dieses Projekt – der Name ARB leitet sich vom lateinischen Wort für Baum, arbor, ab – 1993 als interdisziplinäre Initiative gemeinsam mit dem Lehrstuhl für Rechnertechnik und Rechnerorganisation, Parallelrechnerarchitektur der TUM.

Das umfangreiche Softwarepaket ARB dient dem Erstellen, Warten und Analysieren integrativer Datenban-



ken bekannter DNS-Sequenzen. Seit Jahren wird es weltweit genutzt. Im Zentrum stehen Alignment, Profile und Qualitätsanalyse von Sequenzdaten, denen jede Art von – die entsprechenden Organismen oder Sequenzen betreffenden – Zusatzdaten lokal oder via Internet zugeordnet werden können. Eine grafische Benutzeroberfläche bietet Softwarewerkzeuge etwa für Datenimport und -export, Visualisierung oder Qualitäts-, Struktur- und phylogenetische Analysen. Alle diese Werkzeuge interagieren untereinander und mit der zentralen Datenbank. Auch stehen ARB-Datenbanken für Evolutionsrekonstruktion, Taxonomie und Identifizierung relevanter Sequenzen zur Verfügung, wobei Datenbanken ribosomaler RNA- (rRNA-)Sequenzen von zentraler Bedeutung sind.



Im Rahmen der DFG-Förderung soll das ARB-Softwarepaket weiterentwickelt und den Erfordernissen der enormen Datenmengen angepasst werden, die bei der modernen Hochdurchsatzsequenzierung anfallen. Zudem will man die Prozesse so weit als möglich automatisieren. Ferner sollen die genannten Onlinedatenbanken weitergeführt werden, die rRNA-Datenbanken in Zusammenarbeit mit dem Max-Planck-Institut für Marine Mikrobiologie in Bremen.

www.arb-home.de
www.arb-silva.de

Im Hintergrund der Primärstruktureditor des ARB-Softwarepakets; rechts oben Visualisierung eines Baums; rechts unten der Sekundärstruktureditor: ribosomale RNA-Sequenzen von Bakterien.