



UNIVERSITÄTS-
BIBLIOTHEK



OMERO ALS
INSTITUTIONELLER
FORSCHUNGSDATEN-
MANAGEMENTSERVICE

**REPORT 4 DES FORSCHUNGSPROJEKTS
„COMMUNITY-SPEZIFISCHE FORSCHUNGSDATEN-
PUBLIKATION (CS-FDP)“**

KOMPETENZZENTRUM
FORSCHUNGSDATEN
DER
RUPRECHT-KARLS-UNIVERSITÄT
HEIDELBERG

17.03.2020

Ziel des Projekts „Community-spezifische Forschungsdatenpublikation“ ist die Weiterentwicklung des Serviceangebots des Kompetenzzentrums Forschungsdaten der Universität Heidelberg durch den Aufbau eines Pools von generischen Softwarewerkzeugen zur Erstellung fach- und Community-spezifischer Publikationsportale für Forschungsdaten, die Entwicklung eines Konzepts zur nachhaltigen Integration der Heidelberger Forschungsdaten in übergreifende Archivierungskonzepte sowie die Professionalisierung des Datenmanagements an der Universität durch die Verankerung des Themas in Forschung und Lehre.

Kern des Projekts ist die Erprobung der fachspezifischen Erweiterung und projektspezifischen Individualisierung universitärer Repositorienangebote (typischerweise Volltextrepositorien bzw. Dokumentenserver, teilweise mittlerweile auch generische Repositorien für Forschungsdaten) um ein Dienstportfolio, mit welchem individuelle und fachspezifische Anforderungen erfüllt werden können.

In diesem sowie den weiteren Projektreports werden entsprechende Softwarelösungen vorgestellt, die im Rahmen des Projekts implementiert und evaluiert wurden.

OMERO: ANWENDUNGSBEREICHE, FUNKTIONALITÄTEN, TECHNISCHE ANFORDERUNGEN

OMERO ist ein Produkt der „Open Microscopy Environment“ (OME), einer Initiative verschiedener staatlicher und kommerzieller Forschungseinrichtungen, die es sich zum Ziel gesetzt haben, Software für die Prozessierung und Verbreitung von Mikroskopie-Daten in den Lebenswissenschaften zu fördern. OMERO (OME Remote Objects) dient hierbei mehreren Zwecken: Es repräsentiert sowohl ein Speichersystem als auch ein Datenmanagement und Analyse Tool (Goldberg et al., 2005). Das zugrundeliegende Datenmodell besteht aus zwei Hauptkomponenten: zum einen verschiedene Open-Source-Tools, die relationale Datenbanken nutzen, zum zweiten ein XML-basiertes Datenformat (OME XML) zum Export der Informationen aus den Datenbanken. Die Flexibilität von OMERO ermöglicht seinen Einsatz in der Lichtmikroskopie, im High Content Screening, der Elektronenmikroskopie und sogar bei nichtbildlichen Genotypdaten (Allan et al., 2012). OMERO stellt daher eine wesentliche Stütze im Portfolio der generischen Softwaretools des Kompetenzzentrums für Forschungsdaten (KFD) im Rahmen des CS-FDP dar, da es eine breite Palette von Imaging-Daten vor allem in den Lebenswissenschaften, aber auch darüber hinaus verwerten kann (<https://www.openmicroscopy.org/omero/>).

Pilotpartner an der Universität Heidelberg ist das Institut für Anatomie und Zellbiologie der Medizinischen Fakultät Heidelberg, Abteilung Funktionelle Neuroanatomie.

An der Universität Heidelberg wird eine Instanz von OMERO als Software zur Verarbeitung sehr großer Mikroskopie-Datensätze eingesetzt, zunächst als Pilotprojekt in der Zellbiologie (Apel et al., 2018). Die wesentliche Neuerung und Hauptstoßrichtung dieses Projektes ist es, die OMERO Instanz, die bereits an diversen nationalen und internationalen Einrichtungen weite Verbreitung findet, an den Hochleistungsspeicher der Universität Heidelberg anzubinden. Der Scientific Data Storage (SDS@hd) ist ein Service des Universitätsrechenzentrums Heidelberg für das Land Baden-Württemberg, welches die sichere Speicherung großer Datenmengen sowie deren schnelle Prozessierung erlaubt. Der Service dient der Ablage von vitalen Forschungsdaten, auf die naturgemäß regelmäßige und schnelle Zugriffe erforderlich sind („hot data“). SDS@hd bzw. LSDF2 stellen wesentliche Elemente der universitären Cloudinfrastruktur (heiCLOUD) dar. Die Einbindung von SDS@hd

ermöglicht das Rendering der Bilddaten für verschiedene Zoomstufen über die Cloudinfrastruktur der Universität. Das heißt, die zum Teil sehr aufwendigen Rechenoperationen erfolgen nicht mehr am Institutsserver oder im Endgerät des Forschenden, sondern werden in der Cloud prozessiert und nur das Ergebnis der Verarbeitung erscheint am Endgerät des Nutzers. Dieser Ansatz führt zu einer drastischen Erhöhung der Prozessierungsgeschwindigkeit der Bilddaten. Aus gesamtuniversitärer Sicht bringt dieser zentralisierte Dienst erhebliche Effizienzsteigerungen, da die Notwendigkeit für das Vorhalten von Recheninfrastruktur an den Instituten und Arbeitsgruppen für die Prozessierung der Daten entfällt. Gleichzeitig wird durch die zentrale Verarbeitung eine wesentliche Voraussetzung für die spätere Publikation der Daten geschaffen. Ursprünglich wurde daher darauf hingearbeitet, OMERO nicht nur als Arbeitsplattform, sondern auch als Publikationsorgan an der Universität Heidelberg zu etablieren. Da zum Zeitpunkt des Projektendes die Nachfrage von wissenschaftlicher Seite noch verhalten ist, wurde dies nicht umgesetzt. Es ist jedoch im Zuge der zunehmenden wissenschaftspolitischen Entwicklung in Richtung Open-Data-Plattformen durchaus wahrscheinlich, dass dies in Zukunft eine Option sein wird.

BEWERTUNG HINSICHTLICH LANGFRISTIGER NUTZUNG

Insgesamt ergeben sich über die Laufzeit des Projektes klare positive Effekte durch die Nutzung von OMERO für die Forschenden. Die schnelle Anbindung an die Cloudinfrastruktur ermöglicht eine deutlich höhere Bearbeitungsgeschwindigkeit der Bilddaten. Zukünftig wäre eine zusätzliche Anbindung an High-Performance Computing Cluster (HPC) denkbar. Das integrierte Datenmanagementsystem bietet erhöhten Schutz vor Datenverlust, diverse Exportmöglichkeiten der Daten (Darwin Core Standard), verbesserte Datenkuration und die Möglichkeit, Bilder und Textdateien mit Musterdatensätzen zu verknüpfen. Der Dienst wurde über die gesamte Laufzeit durchgehend genutzt und weitere Arbeitsgruppen der Universität haben Ihr Interesse bekundet zukünftig OMERO zu nutzen.

Aus Sicht der Administratoren am KFD besticht OMERO durch eine ausgereifte und gut gepflegte Software. Die Dokumentation der Software ist ebenfalls sehr anwenderfreundlich und bedingt durch die weltweite Verbreitung, existiert eine vergleichsweise große Entwicklercommunity, die sehr aktiv in Entwicklerforen ihre Erfahrungen teilen. Im laufenden Betrieb verursacht OMERO daher einen vergleichsweise überschaubaren Aufwand für die Administratoren und empfiehlt sich aus Sicht des KFD für eine weitere Nutzung.

WEITERFÜHRENDE LITERATUR

Allan, C., Burel, J.-M., Moore, J., Blackburn, C., Linkert, M., Loynton, S., MacDonald, D., Moore, W.J., Neves, C., Patterson, A., Porter, M., Tarkowska, A., Loranger, B., Avondo, J., Lagerstedt, I., Lianas, L., Leo, S., Hands, K., Hay, R.T., Patwardhan, A., Best, C., Kleywegt, G.J., Zanetti, G., Swedlow, J.R., 2012. OMERO: flexible, model-driven data management for experimental biology. *Nat. Methods* 9, 245–253. <https://doi.org/10.1038/nmeth.1896>

Apel, J., Gebhart, F., Maylein, L., Wlotzka, M., 2018. Offene Forschungsdaten an der Universität Heidelberg: von generischen institutionellen Repositorien zu fach- und projektspezifischen Diensten. *O-Bib Offene Bibl.* Herausgeber VDB 5, 61–71. <https://doi.org/10.5282/o-bib/2018H2S61-71>

Goldberg, I.G., Allan, C., Burel, J.-M., Creager, D., Falconi, A., Hochheiser, H., Johnston, J., Mellen, J., Sorger, P.K., Swedlow, J.R., 2005. The Open Microscopy Environment (OME) Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging. *Genome Biol.* 6, R47. <https://doi.org/10.1186/gb-2005-6-5-r47>

WEITERFÜHRENDE WEBLINKS

<https://www.openmicroscopy.org/omero/>

<https://www.urz.uni-heidelberg.de/en/sds-hd>

<https://doi.org/10.25504/FAIRsharing.t6srnm>