



## Keeper Projekt

Das skalierbare und zuverlässige Langzeitarchiv

Schnell, zuverlässig und skalierbar – Mit dem KEEPER Projekt entstand bei der Max-Planck-Gesellschaft ein Langzeitdatenarchiv auf Basis der private Cloud Lösung Seafile. Es garantiert die Speicherung sämtlicher Daten für alle Mitarbeiter für mindestens ein Jahrzehnt.

# Das skalierbare Langzeitarchiv

## Seafire bei der Max-Planck-Gesellschaft

### Die Herausforderung großer Datenmengen

Universitäten, Forschungsinstitute sowie kleine und mittelständische Unternehmen benötigen die Fähigkeit, schnell und unkompliziert Daten speichern, tauschen und langfristig archivieren zu können. Diese Fähigkeit wird zunehmend zu einer Grundvoraussetzung um sich im immer schnelleren Wettbewerb behaupten zu können.

Dementgegen stehen immer größer werdende Datenmengen, komplizierte Organisationsverflechtungen, Internetkriminalität und strenge Datenschutzbestimmungen. Dies ist nur ein Bruchteil der täglichen Herausforderungen mit denen IT Abteilungen täglich konfrontiert werden und die es zunehmend kompliziert machen eine performante und zuverlässige Lösung anzubieten.

Im Rahmen des KEEPER Projektes hat die Max-Planck-Gesellschaft mit Hilfe der private Cloud Software Seafire ein einzigartiges Speicher- und Archivierungssystem für sämtliche Mitarbeiter und Forschungspartner der Max-Planck-Institute geschaffen. KEEPER garantiert jedem Mitarbeiter einen Terabyte Speicherplatz, die für mindestens 10 Jahre bereitgestellt werden.

### Key Facts

- 1 TB Speicherplatz und 10 Jahre garantierte Aufbewahrung für jeden Mitarbeiter der Max-Planck-Institute
- Geschwindigkeit und Zuverlässigkeit durch lokales Arbeiten und zentrale Datenspeicherung in der Max-Planck-Infrastruktur
- Erfüllt die Archivierungsvorschriften der Max-Planck-Gesellschaft und die Regeln zur Sicherung "guter wissenschaftlicher Praxis"

## Ein zentraler Datenspeicher für alle Max-Planck-Institute

### Wieso Seafile sich am Ende gegen alle Alternativen durchsetzen konnte

KEEPER ist ein Service der „Max Planck Digital Library (MPDL)“. Die MPDL ist eine zentrale Einheit der Max-Planck-Gesellschaft, mit dem Ziel die bestmögliche Unterstützung in den Bereichen der Informations-bereitstellung, dem Datenmanagement und der Publikationsunterstützung zu bieten. KEEPER ist dabei eine zentrale Speicher- und Synchronisations- und Archivierungslösung um den riesigen Datenmengen der Forscher und Forscherinnen eine Heimat zu geben.

Eine zentrale Anforderung neben dem Datenmanagement war die nahtlose Integration in bestehende Arbeitsprozesse. Nur wenn der Arbeitsfluss der Mitarbeiter nicht gestört oder verändert wird, erreicht man die größtmögliche Akzeptanz der Anwender.

Mit KEEPER hat man dieses Ziel erreicht und bietet nun jedem Mitarbeiter ein umfangreiches Datenspeicherkontingent von bis zu einem Terabyte. Die Mitarbeiter können wie gewohnt lokal an ihren Forschungsdaten arbeiten, diese jedoch zentralisiert in der Max-Planck-Infrastruktur speichern.

Natürlich ist eine der wichtigsten Aufgaben von KEEPER die Langzeitarchivierung von Dateien und Informationen, sobald ein Projekt abgeschlossen ist. Gleichzeitig können durch die Synchronisationsfähigkeiten Daten im Team oder mit externen Forschungspartnern unkompliziert geteilt werden.

Darüber hinaus erspart KEEPER seinen Nutzern eine Menge Sorgen: KEEPER wurde so aufgebaut, dass sich die Mitarbeiter sicher sein können, dass sie durch die Nutzung sämtliche Archivierungsvorschriften der Max-Planck-Gesellschaft erfüllen und auch automatisch die Regeln zur Sicherung „guter wissenschaftlicher Praxis“ einhalten. Gleichzeitig schützt KEEPER sämtliche Daten dank seiner integrierten

Dateiversionierung und die Volltextsuche erleichtert das Auffinden der gesuchten Informationen.

Obwohl KEEPER bisher primär den internen Mitarbeitern zur Verfügung stand, startete Anfang 2018 der Rollout von KEEPER an alle Mitarbeiter der Max-Planck-Institute. Gleichzeitig können kooperierende Unternehmen und Institute auch Zugang zum KEEPER Speichersystem erhalten. Es ist geplant das Datenvolumen von KEEPER kontinuierlich zu vergrößern um in der finalen Ausbaustufe Petabyte an Daten aufnehmen zu können.

Wenn man ein Langzeitdatenarchiv für die nächsten Jahrzehnte plant, kann die Evaluationsphase gar nicht umfangreich genug sein. Neben der Prüfung der geforderten Funktionen gilt es auch Sicherheitsaspekte, Zuverlässigkeit und die voraussichtlichen Folgekosten abzuwägen.

So haben Mitarbeiter der „Max Planck Digital Library“ und des KEEPER Projektes neun Softwarelösungen mehrere Wochen unter die Lupe genommen und auf Herz und Nieren geprüft. Am Ende konnte sich „Seafile“ gegenüber allen anderen private Cloud-Lösungen klar durchsetzen und Konkurrenten wie owncloud, github, pydio und powerfolder hinter sich lassen.

Seafile überzeugte seine Tester mit der besten Performance und Stabilität im gesamten Testfeld, wobei die gute Integration in die vorhandenen Arbeitsprozesse eine Grundvoraussetzung war.

Hinzu kamen die umfangreichen Möglichkeiten der API Schnittstelle, die ansprechende und individuell anpassbare Oberfläche, die ausgefeilten Algorithmen und eine gute Unterstützung für verschiedene Backendclustersysteme.

# Technische Details

## Der Aufbau des Datenarchivs erfolgt in Phasen

Die Entwicklung des KEEPER Projektes erfolgte bisher in zwei aufeinander aufbauenden Phasen.

In der ersten Phase genügte noch ein einzelner „Seafile Server“ mit 4 CPU-Kernen und 32 GB RAM für die ca. 200 Anwender. Als Storage Backend kamen zwei weitere virtuelle Maschinen und ein gespiegeltes IBM Cluster-Speichersystem zum Einsatz. Dieses hochverfügbare Backend bewältigte mit Leichtigkeit die durchschnittlich 30 Gigabyte an Datenbewegungen pro Tag.

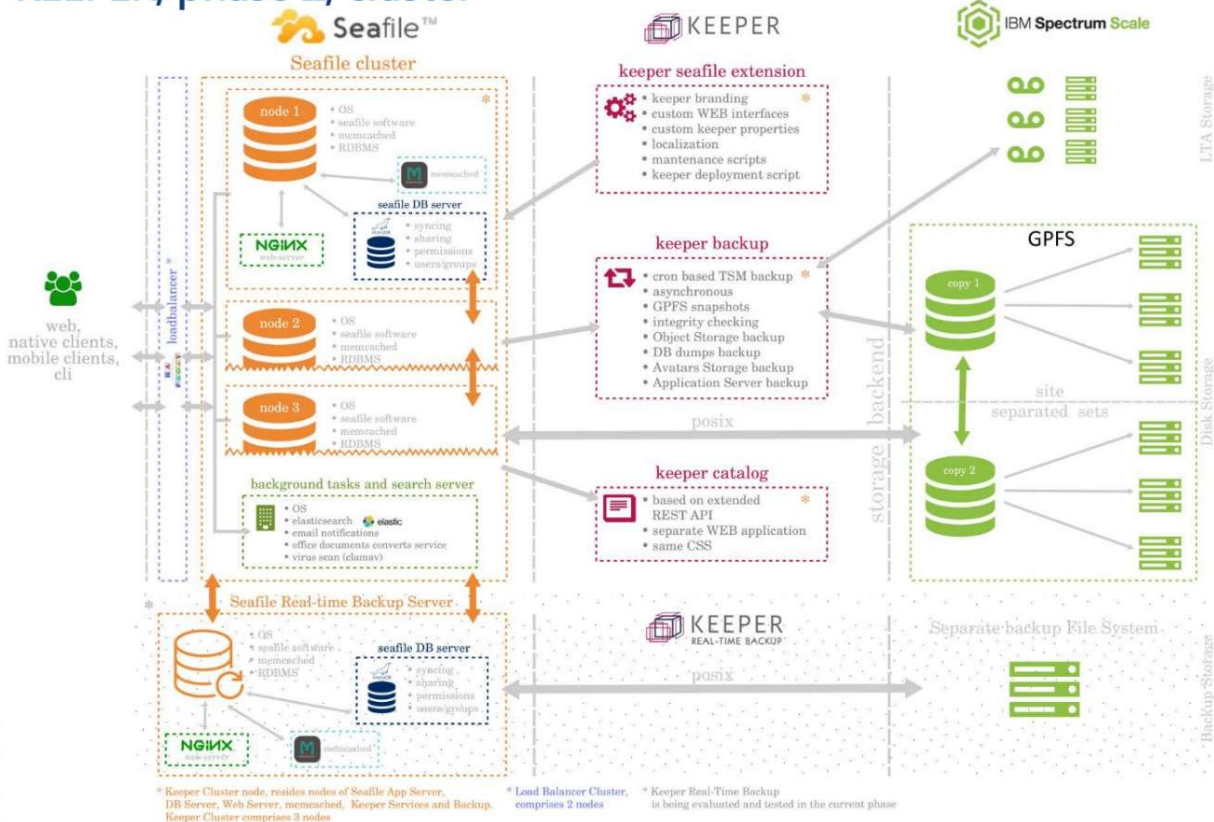
In Phase 2 des KEEPER Projektes wurde der Seafile Server durch weitere Server zu einem Applikation-Cluster ergänzt. Mit Hilfe von vorgeschalteten Loadbalancern werden die Nutzerzugriffe verteilt und ein separater Server

für Hintergrundarbeiten entlastet die Applikationsserver. Ein weiterer Server für Real-Time Backups ist in Planung.

Das Storage Backend basiert weiterhin auf einem IBM GPFS, wurde jedoch auf zwei Standorte verteilt. Weitere Ergänzungen gab es im Clustermanagement. So wurden zusätzliche Softwarelösungen wie Icinga und Kibana zur Überwachung und zum Monitoring der VMs und Server integriert.

Die nächsten Schritte, die das KEEPER Projekt im Jahr 2018 angehen möchte, ist die Integration mit internen und externen IT-Services. Weitere Phasen werden in den kommenden Jahren mit Sicherheit folgen.

### KEEPER, phase II, cluster



Quelle: „Seafile as long term archiving tool for Max-Planck institutions“ von Vladislav Makarenko, Max Planck Digital Library

# Seafile und die Max-Planck-Gesellschaft

Eine gute und langfristige Verbindung



## Seafile Ltd.

Seafile ist eine on-premise File Hosting und Sharing Lösung, die seit 2012 von der Seafile Ltd. mit Sitz in Peking entwickelt wird. Seafile wird weltweit von über 500.000 Nutzern täglich genutzt. Unternehmen und Organisationen wie Kaspersky Lab oder die Humboldt-Universität zu Berlin setzen auf die Vorteile der Seafile-Cloud.

Den Europavertrieb von Seafile hat die Mainzer Firma datamate übernommen. Mit der Seafile Community Edition bietet der Hersteller einen Seafile Server mit eingeschränkten Funktionen unter einer Open Source Lizenz an.



MAX-PLANCK-GESellschaft

## Max-Planck-Gesellschaft

Die Max-Planck-Gesellschaft zur Förderung der Wissenschaften e. V. (MPG) ist eine der führenden deutschen Institutionen im Bereich der Grundlagenforschung. Der gemeinnützige Verein mit Sitz in Berlin unterhält 84 rechtlich unselbstständige Forschungsinstitute und -einrichtungen in ganz Deutschland.

Die Max-Planck-Gesellschaft genießt weltweite Anerkennung. Nicht-universitäre Forschungsinstitutionen haben sie 2006 im Times-Higher-Education-Supplement-Ranking zur weltweit besten nicht-universitären Forschungseinrichtung gekürt.

## datamate

datamate ist Seafiles offizieller Support- und Vertriebspartner in Europa.

datamate GmbH & Co. KG  
117er Ehrenhof 5  
55118 Mainz / Deutschland

Telefon +49 (0)61313270777  
Email [seafile@datamate.org](mailto:seafile@datamate.org)

<https://de.seafile.com>  
<https://datamate.org/seafile>