

SPEICHERUNG UND ARCHIVIERUNG VON FORSCHUNGSDATEN AM I3MAINZ

Projektleitung

Prof. Dr. Kai-Christian Bruhn
(Fachbereich Technik / i3mainz -
Institut für Raumbezogene
Informations- und Messtechnik)

Projektmitarbeiter

Thomas Engel M.A.
Dipl.-Ing. (FH) Nikolai Bock M.Eng.

Laufzeit

Seit 2013

Finanzierung

i3mainz – Institut für Raumbezogene Informations-
und Messtechnik
ZIK - Zentrum für Informations- und Kommunikati-
onstechnik

Kooperationspartner

ZIK - Zentrum für Informations- und Kommunikati-
onstechnik

Kontakt

i3mainz@hs-mainz.de

Themenstellung

Seit mehr als einer Dekade formulieren deutsche Wissenschaftsorganisationen Grundsätze für den Umgang mit Forschungsdaten speziell im Hinblick auf deren Nachnutzung¹. Wurden diese zunächst auf freiwilliger Basis zur Disposition gestellt, sind es heute vor allem die großen Wissenschaftsförderer, die auf einer Umsetzung der Grundsätze bestehen. Jeder Antrag auf Förderung muss daher eine ausformulierte Strategie zum Thema Forschungsdatenmanagement beinhalten. Ein zentraler Bereich dieser Strategien ist die Speicherung und Langzeitsicherung von Forschungsdaten. Als erfolgreiches Institut ist das i3mainz aufgefordert, Strategien für den Umgang mit Forschungsdaten zu entwickeln, um

die Anschlussfähigkeit und auch zukünftige Erfolge bei der Drittmittelerwerbungsleistung zu gewährleisten.

In Forschungsprojekten mit Schwerpunkt Geoinformatik und Vermessung entstehen große Datenmengen, sowohl bei der Erfassung als auch bei der Prozessierung und Auswertung. Daten im Terabyte-Bereich sind keine Seltenheit, insbesondere beim Einsatz moderner Messverfahren zur

3-dimensionalen Erfassung, wie dem terrestrischen Laserscanning (TLS) oder „Structure from motion“ (SfM). Auch zur Verwaltung bzw. Bereitstellung von räumlichen Sensordaten oder der Prozessierung umfangreicher Geodatenätze werden große Speicherkapazitäten benötigt. Die Forderung nach zuverlässiger und nachhaltiger Verfügbarkeit stellt dabei auch hohe Anforderungen an einen zentralen Netzwerkspeicher sowie an das



Abb. 1: Schicht 1 – Storage-Array des i3mainz, gehostet im Serverraum des ZIK

¹ Anfängliche Diskussionen kulminierten 2003 in der „Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen“; eine aktuelle Formulierung der Problematik findet sich hier: <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsatz/>

Netzwerk selbst. Eine Sicherungs- und Archivierungsstrategie muss angesichts wachsenden Speicherbedarfs in der Lage sein, adäquat zu skalieren.

In der sich entwickelnden Landschaft der Forschungsdatenzentren und -infrastrukturen nimmt das i3mainz als datenerzeugende, aber nicht langzeitarchivierende Institution eine Position ein, die als prototypisch für die meisten Forschungsinstitute angesehen werden kann. Um eine mittelfristige Verfügbarkeit der relevanten Datenbestände zu gewährleisten, wurde eine zweischichtige Speicher- und Backup-Strategie entwickelt:

1. Schicht – Internes Filesharing des i3mainz

In der ersten Schicht steht ein zentraler Datenspeicher mit derzeit 48TB Bruttokapazität zur Verfügung (erweiterbar auf bis zu 386TB). Technisch handelt es sich um ein Storage-Array, welches auf RAIDSET-Ebene durch Hotspare-disks, wie auf RAID-Ebene durch block-level striping und doppelter Parität (RAID6) Redundanz bietet. Weiterhin werden neben obligatorischer redundanter Stromversorgung auch die RAID-Controller redundant betrieben, was so genanntes „Multipathing“ ermöglicht. Das System wird von zwei Mitarbeitern des i3mainz sowie einem externen Dienstleister gewartet und in den Serverräumen des Zentrums für Informations- und Kommunikationstechnik (ZIK) gehostet. Es basiert auf iSCSI-Technologie, welche sich durch eine

flexible Zuweisung von Festplattenspeicher sowohl an physische als auch an virtuelle Server auszeichnet. Dokumentierte Prozesse ermöglichen potentiell die Administration durch Dritte.

2. Schicht – Langzeitsicherung/ Archivierung via IBM Tivoli-System

Auch redundant ausgelegte Systeme sind nicht unfehlbar, so dass in jedem Fall eine Sicherung der Daten erfolgen muss. Da ein Forschungsinstitut wie das i3mainz nicht in der Lage sein kann, eine mittelfristige Datenarchivierung auf professionellem Niveau zu gewährleisten, wurde eine Kooperation mit dem ZIK geschlossen, welches für diese Zwecke ein IBM Tivoli System betreibt. Sicherungsstrategien sind tägliche Datensicherung (Backup) sowie langfristige Zustandsarchivierungen (Archive). Die Tape-Medien, auf denen gesichert wird, werden räumlich aus dem Hostingbereich entfernt, um durch eine weitere Redundanzebene einem Totalverlust vorzubeugen. Im Normalfall werden im i3mainz also mindestens drei Kopien einer Datei und Dateiversion vorgehalten.

Ausblick

Die Strategie zur Sicherstellung physischer Datenverfügbarkeit ist mit einer institutsinternen Datenpolitik zu koppeln, in der festgelegt wird, welche Daten aus welchen Forschungszusammenhängen in welchen Formaten abzuspeichern sind. Zudem ist eine einheitliche Auszeichnung mit Metadaten not-



Abb. 2: Schicht 2 – IBM Tivoli System für Backup und Archivierung der Forschungsdaten des i3mainz

wendig. Die derzeitige Praxis setzt diese Zielstellung nicht vollständig um und muss daher in den kommenden Jahren angepasst werden. Eine festzulegende Daten-Politik ist sinnvollerweise mit den Datenzentren abzustimmen, die eine Langzeitarchivierung gewährleisten wollen.