Leibniz Universität Hannover

Forschungsdatenmanagement, Backup & Archiv

Hans Georg Krojanski

- **Forschungsdaten**
- **Beispiel**
  - Backup oder Archiv oder Forschungsdatenarchiv?
- **Backup, Archiv, Forschungsdatenarchiv**
  - Speicherung (Art, Dauer, Formate)
  - Zugangsmöglichkeiten
- **Services des Rechenzentrums**
- **Datensicherheit**
- **Zukunftssicherheit**
- **Zugänge**
- **Kosten**

Forschungsdaten

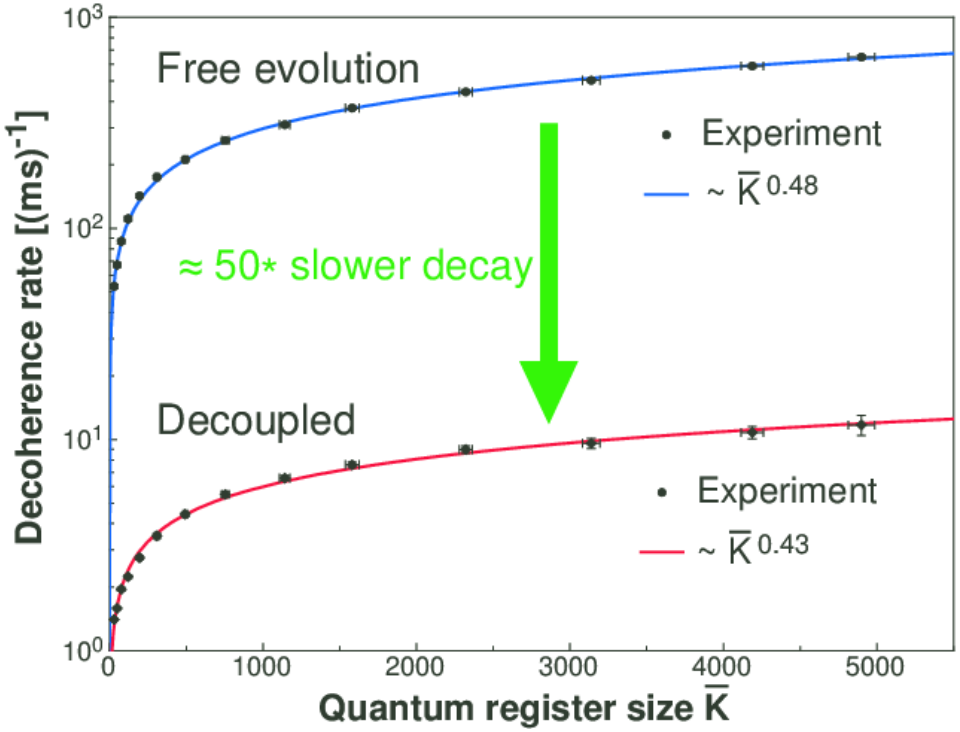
- „Unter **Forschungsdaten** sind [...] digitale und elektronisch speicherbare, Daten zu verstehen, die im Zuge eines wissenschaftlichen Vorhabens z.B. durch Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen.“

  DFG : Ausschreibung “Informationsinfrastrukturen für Forschungsdaten” (2010)

- Keine feste Definition; stark abhängig vom Fachbereich
- Roh-/Primärdaten → verarbeitete Daten → Veröffentlichung
- **Forschungsdatenmanagement**:
  - (Nach-) Nutzung sicherstellen
  - Datenzugriff und -auswertung unabhängig vom Produzent
  - Transparent und nachvollziehbar

Beispiel

- **Veröffentlichte Daten** sind normalerweise stark „komprimiert“
  Hier Ausgangsdaten:
  N * 2048 * 1024
- Was benötigt man zur **Nachnutzbarkeit**?
  - Ausgewertete Daten
  - Primärdaten
- Was benötigt man zur **Nachprüfbarkeit**?
  - Details zum Versuchsaufbau
  - Erfahrungswissen



Krojanski and Suter, Phys. Rev. Lett. **97**, 150503 (2006)

Backup oder Archiv oder Forschungsdatenarchiv?

- Rohdaten => Primärdaten => verarbeitete Daten
- Benutzte Methode (paper: Beschreibung; zus.: RF-Pulsfolgen)
- Kalibrationsmessungen (nur indirekt im paper)
- Parameter des Versuchsaufbaus
  (Filterbandbreiten, Stabilität der Zeitbasis, Genauigkeit, ...)
  - paper: nur Ergebnisse
  - Rohdaten?
  - Designinformationen (Bilder, Texte, interne Berichte)
- Spektrometersteuerung (Programmcode? Steuerungssequenzen aus programmierten RF-Pulsfolgen?)
- Datenauswertungssoftware? (Kommerziell; Eigener Code)
  zum relevanten (damaligen) Zeitpunkt!
- Datenauswertungsprozess (im paper beschrieben)
- Laborbuch, Skizzen, Notizen → **Erfahrungswissen**

Backup, Archiv, Forschungsdatenarchiv

- **Backup**
  - Sollte automatisiert erfolgen
  - Versionierung; Kurzzeitig (einige Wochen)
  - Original-Datenformate (Primärformate)
- **Archiv**
  - Ausgewählte Informationen (→ Kapazität)
  - Bestimmter Zeitpunkt; Langfristig (z. B. 10 Jahre lang)
  - Archivformate (PDF/A, ...) statt Primärformate?
- **Forschungsdatenarchiv**
  - Auswahlkriterien müssen nachvollziehbar sein
  - Archivformate, offene Standards (Formate, Schnittstellen)
  - Dauerhafte Auswertungsmöglichkeit? (Primärdaten OK?)

Niemand möchte Backup, alle möchten Restore...

- **Backup**
  - Restore: bei Bedarf (zu definiertem Zeitpunkt)
  - Spezielle Backup-Client-Software
  - Durch Admin
- **Archiv**
  - Retrieval: bei Bedarf (selbst, auf Nachfrage, ...)
  - Standard-Zugänge (FTP, ...)
  - Nutzerin selbst
- **Forschungsdatenarchiv**
  - Offenes Datenrepo oder „Dark Archive“?
  - Web-Zugang; SOAP, REST?
  - Authentifizierung? (Bei Einstellung und/oder Retrieval)

Service: Backup & Restore

- 2 TB pro Server; insgesamt 4 TB
- Auf Festplattensystem + Tape Library

Backup & Restore

Kurzbeschreibung

Der Service Backup & Restore bietet für alle Institute und zentralen Einrichtungen der Leibniz Universität Hannover kostenlos die regelmäßige Erstellung und kurzfristige Aufbewahrung von Sicherungskopien bestimmter Serverdaten zum Schutz vor ungewolltem Datenverlust. Die Daten werden dabei in einem automatisierten Verfahren über das LAN zum zentralen Dienstleister übertragen und dort als Sicherungskopie für einen begrenzten Zeitraum gespeichert (Backup). Bei Bedarf werden die gesicherten Daten wiederhergestellt (Restore). Die eingesetzte Software ist Bacula Enterprise.

Ihre Vorteile

- Ihre Daten werden kontinuierlich in automatisierter Weise gesichert

Wer kann den Service nutzen?

- ✔ Institute / zentrale Einrichtungen
- ✘ einzelne Beschäftigte
- ✘ Studierende

Kontakt

Ansgar Giesker, 762-3848

Oliver Heimbrock, 762-7919083

support@rrzn.uni-hannover.de

Bacula Enterprise

http://www.baculasystems.com/

Service: Archivierung

- Bandroboter (Quantum Scalar 10K, Migration → Quantum i6K)
- 500 GB pro Nutzer; 8 Jahre Aufbewahrung
- Derzeit **Neugestaltung**

Archivierung

Kurzbeschreibung

Sie haben die Möglichkeit uns mit der Archivierung Ihrer Daten zu beauftragen. Archivierung bedeutet das Speichern und Auslagern von Daten für einen längeren Zeitraum auf Magnetband im Roboter. In der Regel werden die Daten bis zu acht Jahre gespeichert.

Ihre Vorteile

- Wir stellen sicher, dass die Daten, die archiviert wurden, nach Jahren zurückgespielt werden können.
- Es werden zwei Kopien geschrieben (auf unterschiedlichen Bändern).

Wer kann den Service nutzen?

- ✔ **Institute / zentrale Einrichtungen**
- ✔ **einzelne Beschäftigte**
- ✘ **Studierende**

Kontakt

Bei Fragen zum Archiv-Dienst erreichen Sie uns am einfachsten und schnellsten unter:

- datensicherung@rrzn.uni-hannover.de

Diese Adresse wird von mehreren Mitarbeitern gelesen und daher in der Regel schneller beantwortet verglichen mit einer direkten Anfrage an einen unserer Mitarbeiter.

Hardware

Datensicherheit

- **Technische Integrität**
  (Festplatten, Controller, RAM, Netzwerke)
- **Unabhängige Kopien**
- **Prüfsummenverfahren** (ISBN, LUH-ID, .iso, ...)
  - Nur Fehlererkennung (Einzelbitfehler, Blockfehler)
  - Erkennung der Fehlerposition
- **Fehlerkorrekturverfahren**
  - Rückwärtsfehlerkorrektur (z. B. TCP)
  - Vorwärtsfehlerkorrektur
    (z. B. Reed-Solomon-Codes → par2, CD, Mobilfunkstd., ...)
- Kryptographische Hash-Funktionen
- Zeitstempel, Audit-Trails, ...

Zukunftssicherheit

- **Dateisysteme** (resource fork HFS → ._<Name> „AppleDouble“)
- **Dateityp** (→ PRONOM)
- **Datenformate** (XML, csv, asciidoc, ...)
- **Containerformate**
  - Fehleranfälligkeit beachten
  - Zusätzliches Daten- und Dateiformat
- **Sourcecode** (Versionskontrolle), **Libraries**
- **Binaries**
  - Welche Version? (Versionsspezifische Bugs mit Folgen...)
- **Prozesswissen**
- **Haltbarkeit** von Speichermedien
  Steintafeln, Papier, CD/DVD, Festplatten, Magnetbänder, ...

Zugänge

- **Anlieferungsschnittstellen**
- **Auslieferungsschnittstellen** (Web-GUI; SOAP, REST, ...)
- Ggf. **Zugangskontrolle**
  - „Dark Archive“: keine öffentliche Benutzeroberfläche
    - Keine gleichzeitige Publikation
    - Ablage und Zugang einzig gesteuert durch Nutzerin
    - Geeignet für Restriktionen oder hohe Sicherheitsstd. (Datenschutz, Urheberrecht, Patentrecht)
  - Authentifizierung, Autorisierung
- Bereitstellung von **Zugriffsfunktionen**
- Bereitstellung von **Suchfunktionen**
- Zugang nach Ausscheiden des Mitarbeiters?

Kosten

- Je nach Art der Speicherung erheblich
- **Hardware, Software**
  - Datengröße (3D-Modelle vs. Textdateien)
  - Gesamtmenge
  - Integritätsanforderungen (HW, Kopien, Fehlerkorrekturen)
  - End-of-Life Zeitraum der Daten?
  - End-of-Life der HW und SW (Migrationskosten)
  - Regelmäßig (alle 5 bis 10 Jahre je nach HW)
- **Personalkosten**
  - Betrieb, Konzeption
  - Beratung
- **Infrastruktur** (Räume, Strom, Kühlung, ...)