und ZIM) die sogenannten „Research Data Services“ an (https://www.uni-due.de/rds). Hilfreich ist vor allem die Übersicht, welche Backuplösung in der Universität sich für welche Art von Daten eignet (https://www.uni-due.de/rds/speicher_matrix.php.)

Kommerzielle cloudbasierte Backuplösungen Es gibt zahlreiche kommerzielle Angebote für Backups, die auf Servern in einer Cloud liegen. Es gibt Angebote für mehrere Terabyte. Solche Angebote werden monatlich oder jährlich abgerechnet. Die Preise variieren, aber man kann von ca. 100 Euro pro Jahr für einen großen Arbeitsplatzrechner ausgehen. Zu solchen Diensten gehören z.B. *Amazon Web Services (AWS) – S3* oder *Backblaze*. Es sollte bedacht werden, dass man keinerlei Kontrolle über den Speicherort der Daten oder die Art des Zugriffs durch Dritte besitzt. Ohne vollständige Verschlüsselung unter eigener Kontrolle sollten solche Dienste nicht genutzt werden.

Beispiele für Backupstrategien

Beispiel anhand einer Arbeitsgruppe

Die Arbeitsgruppe betreibt einen größeren Fileserver und vier größere Rechner für Simulationen und Berechnungen, jeweils unter Linux. Dazu kommen mehrere Arbeitsplatzplatzrechner, die meisten unter Windows. Die Arbeitsplatzrechner und die Rechenserver synchronisieren ihre Dateien mit dem Fileserver der Arbeitsgruppe (*Nextcloud*). Dieser Rechner wird von einen lokalen Backup-Server sowie durch ein NAS in einem anderem Gebäude gesichert. Der Backupserver wird zusätzlich durch einen zentralen Dienst der Universität auf einem anderen Campus gesichert.

Beispiel anhand eines Mac-Laptops

Das Laptop wird über das zum Betriebssystem gehörende Programm *Time Machine* verschlüsselt auf ein NAS gesichert, sobald sich das Laptop im lokalen WLAN befindet. Das Programm erstellt stündlich eine Kopie geänderter Dateien. Parallel dazu werden die geänderten Dateien kontinuierlich durch *Nextcloud* auf dem Server der Arbeitsgruppe synchronisiert, dessen Platten ebenfalls verschlüsselt sind. Zusätzlich läuft auf dem Rechner ein kommerzielles Backup-Programm (*Carbon Copy*), das wöchentlich ein bootfähiges Backup des Rechners auf das NAS speichert.

Neben diesen automatischen Sicherung erfolgt am Ende eines Tages ein manuelles Backup (mit Carbon Copy), das durch Anstecken einer externen SSD (ca. 200 Euro) ausgelöst wird. Die SSD befindet sich nicht auf demselben Tisch wie das Laptop. Das Backup sichert nur die jeweils neuen Dateien und dauert 2 Minuten. Die SSD enthält ein bootfähiges Image des Betriebssystems einschließlich aller Nutzerdaten. Die SSD verbleibt stets im Homeoffice.

Literatur

Deutsche Forschungsgemeinschaft (DFG) (2022): Leitlinien zur Sicherung guter wissenschaftlicher Praxis, Bonn.
Preston, W. C. (2021): Modern data protection. Sebastopol: O’Reilly.
Rat für Sozial- und Wirtschaftsdaten (RatSWD) (2023): Forschungsdatenmanagement in kleinen Forschungsprojekten, Berlin.

Anleitungen

Video, Verschlüsselung mit Cryptomator https://www.uni-due.de/gesellschaftswissenschaften/videoanleitungen

Videos, Syncback+Windows https://www.uni-due.de/gesellschaftswissenschaften/videoanleitungen#backups

Step-by-Step, Duplicati https://www.urz.uni-heidelberg.de/de/support/anleitungen/arbeitsplatzrechner-sichern-mit-duplicati

Step-by-Step, Timemachine (Mac) https://www.psychologie.uni-heidelberg.de/it/knowledgebase/timemachine-backup-fuer-die-datensicherung-auf-dem-mac

Checkliste

- ☑ Backup aller relevanten Dateien erfolgt dreifach.
- ☑ Backups erfolgen auf zwei verschiedenen Medien.
- ☑ Backups sind an zwei verschiedenen Orten hinterlegt.
- ☑ Mindestens zwei Backups sind voll automatisiert.
- ☑ „Bare-Metal-Backup“ ist sichergestellt.
- ☑ Rückspielen der Backups wurde getestet.
- ☑ Backups sind verschlüsselt.
- ☑ Aufbewahrung der Passwörter/Backups dokumentiert.
- ☑ Backup-Dokumentation erstellt und jährlich geprüft.
- ☑ Dokumentation dem Projektleiter/Betreuer übergeben.

Backups für wissenschaftliche Arbeiten

Prof. Dr. Rainer Schnell
Research Methodology Group
Universität Duisburg-Essen

14. Januar 2024

Die Deutsche Forschungsgemeinschaft (DFG 2022, Leitlinie 17) fordert von allen wissenschaftlich Tätigen die Sicherung der Forschungsdaten, der Forschungsergebnisse und – gegebenenfalls – der Forschungssoftware, in der Regel für einen Zeitraum von zehn Jahren. Daher gehört eine systematische Backup-Strategie zwingend zur jedem Forschungsprojekt. Das gilt auch für jede Qualifikationsarbeit.

Notwendigkeit von Backups

Die scheinbare Zuverlässigkeit moderner IT-Systeme lässt die prinzipielle Möglichkeit von Hardwareausfällen, Softwarefehlern (auch durch Updates) sowie Fehlbedienungen, die zum Verlust von Forschungsdaten oder Texten führen können, häufig vergessen.

Zusätzlich gefährden Brände und Wasserschäden nicht nur private Rechner, sondern auch Universitätsinstitute und Rechenzentren. Weiterhin werden Rechner und Laptops vergleichsweise häufig gestohlen oder gehen auf anderen Wegen verloren.

Ransomware-Angriffe auf mehrere Universitäten in Deutschland haben vor allem 2022 zu erheblichen Datenverlusten geführt, auch an der Universität Duisburg-Essen.

Backup-Pläne

Für jeden Laptop, Arbeitsplatzrechner und Server einer wissenschaftlichen Arbeitsgruppe oder eines einzelnen Wissenschaftlers benötigt man einen schriftlichen Backup-Plan, der auch als gedrucktes Dokument Teil der Dokumentation eines Forschungsprojekts oder eine Qualifikationsarbeit sein sollte. Ein Backup-Plan sollte eine eindeutige Bezeichnung der Rechner, die Art der Sicherung, die Häufigkeit der Sicherung, die Verschlüsselung samt Hinweis auf die Aufbewahrung der Passwörter, die verwendete Software und den Aufbewahrungsort der Sicherung enthalten.

3-2-1-Regel

Als Standard gilt die 3-2-1-Regel:

3 unabhängige Kopien jeder Datei auf
2 unterschiedlichen Medien, wobei
1 Medium an einem anderen Ort aufbewahrt wird.

Zu sichernde Dateien

Hierzu gehören neben den Arbeitsdokumenten (Textdateien), administrativen Dateien, Literaturdatenbanken, statistischen Datensätzen, Programmlogs, Quelldateien eigener Programme auch eventuell spezielle Forschungssoftware auf den eigenen Rechnern.

Falls man webbasierte Anwendungen wie z.B. *Overleaf* oder *Limesurvey* betreibt, müssen auch die Anwenderdateien dieser Server gesichert werden, wenn man sich nicht auf die Betreiber dieser Server verlassen möchte.

Dateien des Betriebssystems oder Standardsoftware wie Textverarbeitungsprogramme müssen im Regelfall selten gesichert werden, falls eventuelle Lizenzcodes getrennt gesichert werden. In jedem Fall empfiehlt sich auch eine regelmäßige Sicherung eines Rechners derart, dass das Backup zum Booten eines Rechners genutzt werden kann („bare metal backup“).

Test des Backups

Ein Backupsystem muss getestet werden. Wichtig ist nicht nur die korrekte Speicherung der richtigen Dateien, sondern auch das erfolgreiche Zurückspielen gesicherter Daten. Dies wird oft fälschlich als trivial angenommen. Es müssen die entsprechende Software und Hardware sowie die erforderlichen Passwörter zur Verfügung stehen. Häufig wird übersehen, dass zur Wiederherstellung der Daten eines Backups ein Rechner mit installiertem Betriebssystem erforderlich werden könnte.

Sicherstellung eines „bare metal backups“

Beginnt man mit einem Rechner im Auslieferungszustand kann die Wiederherstellung des Backups mehrere Arbeitstage in Anspruch nehmen. Es empfiehlt sich daher, zumindest eine zusätzliche Art der Sicherung zu wählen, die auch die Wiederherstellung eines Rechners samt vollständigem Betriebssystem erlaubt („bare metal backup“). Je nach Betriebssystem gibt es dafür zusätzliche (Linux: *Clonezilla*), zumeist kostenpflichtige (Mac: *Carbon Copy, SuperDuper*) oder sogar nur als Abonnement verfügbare Programme (z.B. Windows: *Acronis*).

Verschlüsselungen

Es empfiehlt sich fast immer, Backups zu verschlüsseln. Moderne Betriebssysteme erlauben dabei Laufwerke vollständig zu verschlüsseln (z.B. Windows: *BitLocker*, Mac: *FileVault*, Linux: *LUKS*). Ebenso können zusätzlich einzelne Dateien oder einzelne Ordner verschlüsselt werden. Ein frei verfügbares Programm hierzu ist z.B. *Cryptomator*. Verschlüsselungen einzelner Dateien sollten vor allem für Sicherungen auf Netzlaufwerken oder Cloudspeichern verwendet werden. Verschlüsselungen erfordern sichere (lange) Passwörter (und/oder Wiederherstellungsschlüssel). Diese Schlüssel bzw. Passwörter gehen erstaunlich oft verloren. Daher sollten diese immer zu der schriftlichen Dokumentation einer Backup-Strategie gehören und an mehr als einem Ort (z.B. in einem Passwort-Manager wie *Keepass*, auf einem Handy und zusätzlich in einem Aktenordner auf Papier) aufbewahrt werden.

Speichermedien

Für automatische Backups empfehlen sich zunächst klassische Festplatten (HDDs), „Solid State Disks“ (SSDs) und lokale Netzlaufwerke (z.B. NAS). Ebenso sind Sicherungen in der Cloud möglich. Für Archivierungen kommen zusätzlich CDs, DVDs und vor allem M-Disks (BDXL; Kapazität bis 100 Gigabyte) in Frage. Die Haltbarkeit hängt stark von den Bedingungen der Lagerung ab. Die Lagerung sollte kühl und trocken erfolgen. Auch unter optimalen Bedingungen sind alle Speichermedien prinzipiell fehleranfällig. Bei langfristiger Speicherung über mehr als ein Jahrzehnt muss die mögliche Inkompatibilität zukünftiger Hard- oder Software mit dem Speichermedium oder der verwendeten Software bedacht werden. Es empfiehlt sich daher, sowohl verschiedene Medientypen als auch mehrfache Kopien pro Medium zu erstellen. Weiterhin sollten Backupmedien an mindestens zwei räumlich weit getrennten Orten aufbewahrt werden.

Zentralisierte Lösungen

Es gibt universitätsinterne Backupdienste und kommerzielle Lösungen für ein Backup in der Cloud. Trotz der Bequemlichkeit sollte eine zentralisierte Lösung auf keinen Fall das einzige Backup sein. Die Wiederherstellung dezentraler Backups hat bei den Ransomware-Angriffen der Jahre 2022 und 2023 an einigen Universitäten Monate gedauert und einige Forschungsdaten sind für immer verloren gegangen. Ein dezentrales Backup in eigener Verantwortung ist daher immer unabdingbar.

Universitätsinterne Sicherungen für Arbeitsplatzrechner Das ZIM bietet auch zur Sicherung von Arbeitsplatzrechnern interne Netzlaufwerke an, die selbst gesondert gesichert werden. Einzelheiten finden sich unter https://www.uni-due.de/zim/services/fileservice/.

Datensicherung.NRW Die Universitäten in NRW arbeiten an einem gemeinsamen Datensicherungsprojekt (*Commvault*). Dieser Service steht allen Angehörigen der Hochschule für dienstlich genutzte Rechner zur Verfügung. Einzelheiten finden sich unter https://www.uni-due.de/zim/services/backup.

Sicherung von Forschungsdaten an der UDE Speziell zur Sicherung und Archivierung von Forschungsdaten bieten drei Einrichtungen der Universität (SSC, UB