

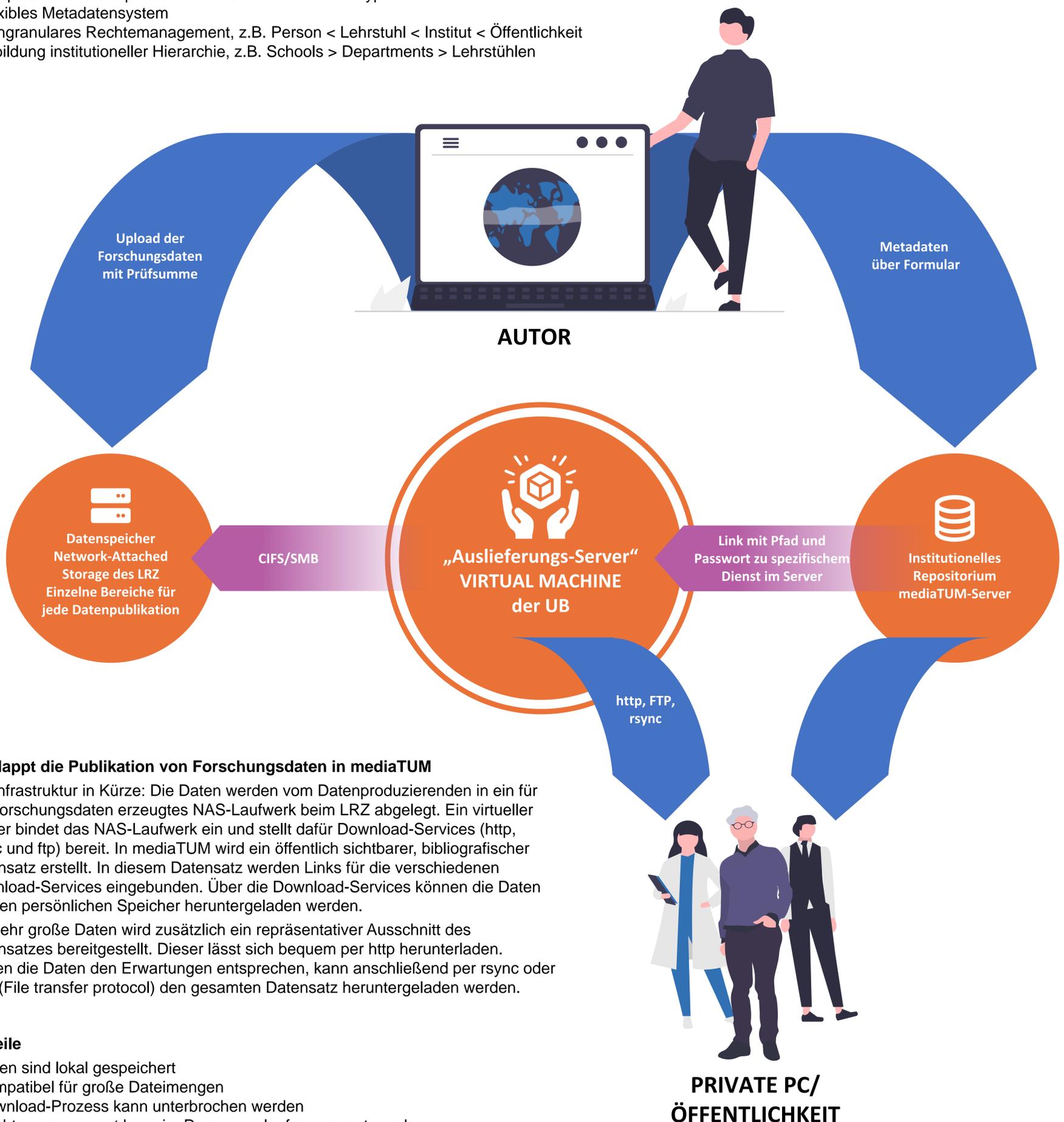
Publikation großer Datenmengen im institutionellen Repository

Christine Wolter

mediaTUM ist der Medien- und Publikationsserver der Technischen Universität München mit mehr als 230.000 öffentlich zugänglichen Datensätzen.

Produktvorstellung mediaTUM

- Open-Source-Software, bei GitHub verfügbar
- Kompatibel für den Upload diverser Datei- und Medientypen
- Flexibles Metadaten system
- Feingranulares Rechtemanagement, z.B. Person < Lehrstuhl < Institut < Öffentlichkeit
- Abbildung institutioneller Hierarchie, z.B. Schools > Departments > Lehrstühlen



So klappt die Publikation von Forschungsdaten in mediaTUM

Die Infrastruktur in Kürze: Die Daten werden vom Datenproduzierenden in ein für die Forschungsdaten erzeugtes NAS-Laufwerk beim LRZ abgelegt. Ein virtueller Server bindet das NAS-Laufwerk ein und stellt dafür Download-Services (http, rsync und ftp) bereit. In mediaTUM wird ein öffentlich sichtbarer, bibliografischer Datensatz erstellt. In diesem Datensatz werden Links für die verschiedenen Download-Services eingebunden. Über die Download-Services können die Daten auf den persönlichen Speicher heruntergeladen werden.

Für sehr große Daten wird zusätzlich ein repräsentativer Ausschnitt des Datensatzes bereitgestellt. Dieser lässt sich bequem per http herunterladen. Sollten die Daten den Erwartungen entsprechen, kann anschließend per rsync oder FTP (File transfer protocol) den gesamten Datensatz heruntergeladen werden.

Vorteile

- Daten sind lokal gespeichert
- Kompatibel für große Dateimengen
- Download-Prozess kann unterbrochen werden
- Rechtemanagement kann im Prozessverlauf angepasst werden
- Öffentlicher Zugang zu ganzem oder repräsentativem Datensatz