



## Lessons learned

Konzeption und Implementierung von Infrastruktur  
und Workflows vor dem Hintergrund erhöhter  
Datenschutzanforderungen

Berlin, 23.09.2019  
**Felix Lange, Dr. Felix Schäfer**  
Projekt-IT

# Gliederung

- Forschungsprogramm GMPG
- Rahmenbedingungen
- Technische Umsetzung
- Lessons Learned

---

# Forschungsprogramm GMPG

## Allgemein

- „Geschichte der Max-Planck-Gesellschaft“
- Untersuchungszeitraum 1948 – 2002
- Wissenschaftshistorische Fragestellungen mit einer breiten Quellenbasis
- Insgesamt ca. 50 Personen:  
Wiss. Mitarbeiter\*innen – SHKs – Gäste

---

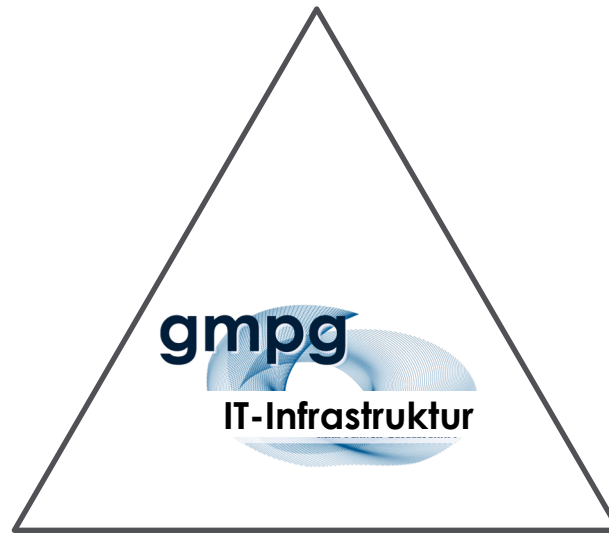
# Forschungsprogramm GMPG

## IT-Aufgaben

- Datenbanken (Archivalien, Personen, Kommissionen, Finanzen)
- Volltextsuche über digitalisierte Archivalien
- Auswertung mit DH-Methoden/Tools
- Infrastruktur für Teil-Projekt zur Massendigitalisierung („M4“)

# Rahmenbedingungen

## Datenschutz & Sicherheit



**Big Data**

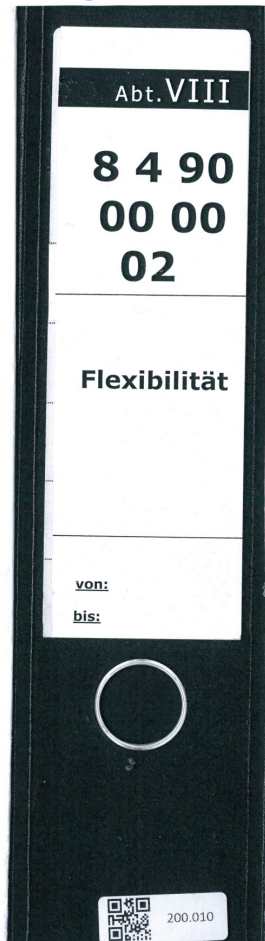
**Projektcharakter**

# Big Data

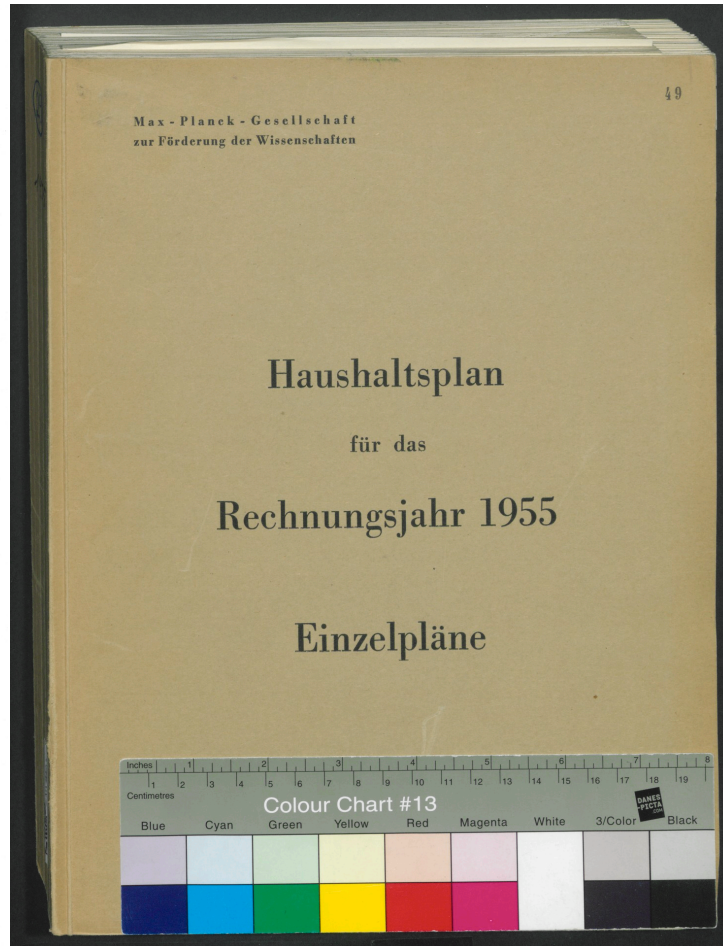
- Massendigitalisierung von  
7.500 Aktenordnern der Registratur der GV  
33.000 Archivalien aus AMPG
- Insgesamt: 9,8 Mio. Scans und ½ PB Speicher
- Wöchentlicher Datenzuwachs zw. 8-15 TB
- Formate:
  - TIFF (unkomprimiert, 400 dpi) für OCR
  - JPEG (150 dpi) für Qualitätssicherung
  - PDF für Volltextsuche/Web-Anwendung



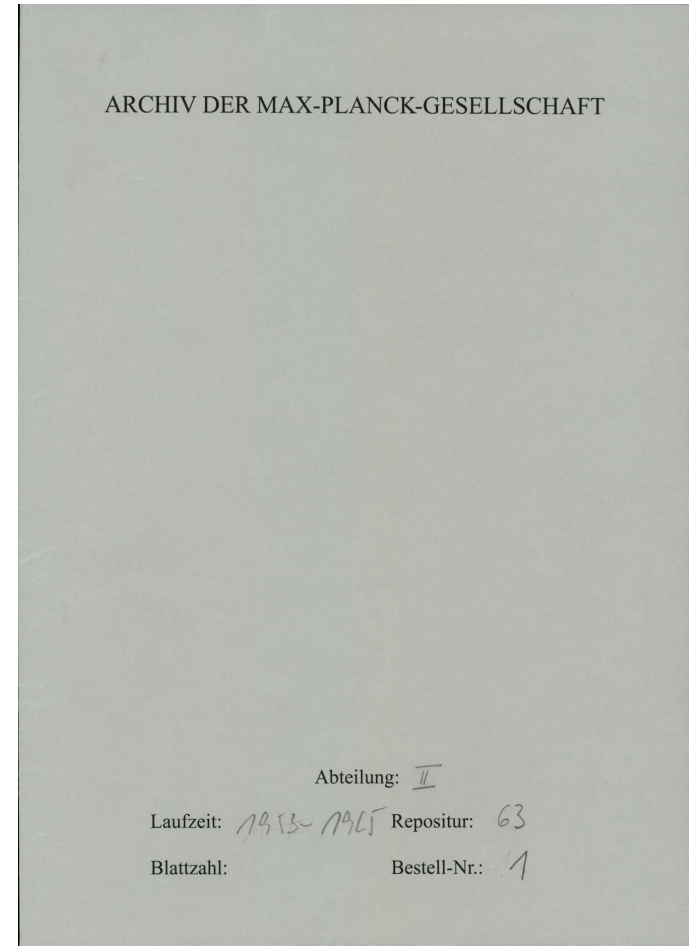
# Big Data



Quelle: Registratur der GV, MPG, München



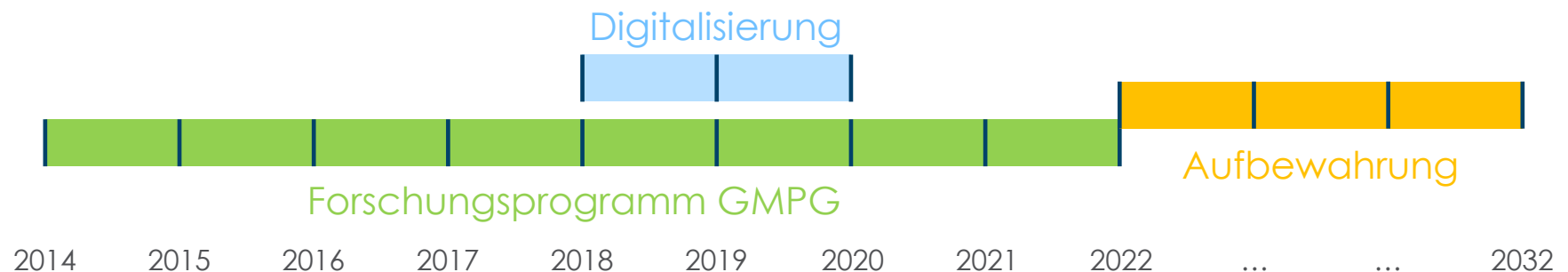
Quelle: Archiv der MPG, Berlin – II. Abt., Rep. 66, Nr. 29



Quelle: Archiv der MPG, Berlin – III. Abt., Rep. 63, Nr. 1

# Projektcharakter

- Aufbau und Betrieb einer projektspezifischen Infrastruktur für befristete Zwecke und Prozesse
- Zeitliche Befristung

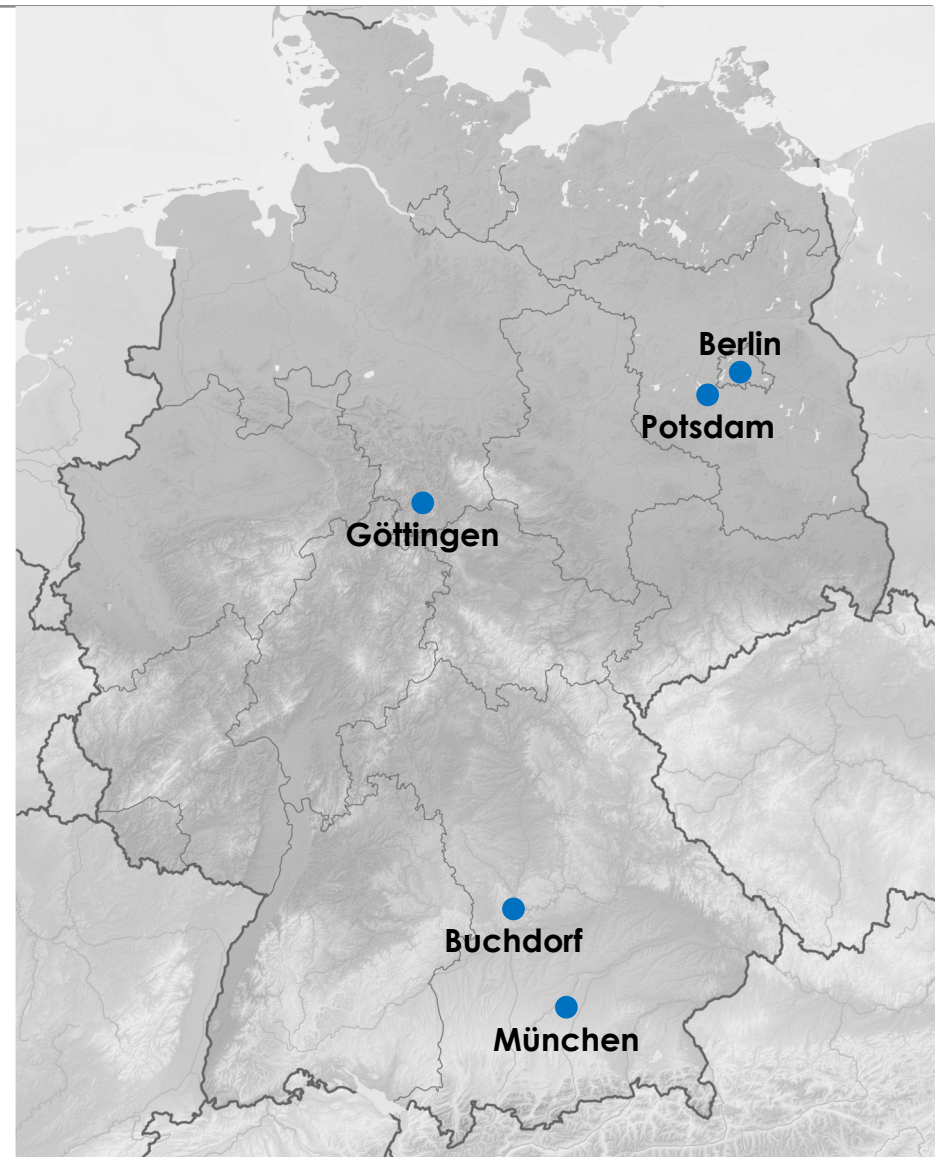


- Mittel-/Langfristige Nachnutzung der Digitalisate durch AMPG



# Projektcharakter

- Verteilte Akteure
  - Datengeber:  
GV & AMPG
  - 2 Scan-Dienstleister  
(inhouse, extern)
  - Qualitätssicherung  
(extern)
  - IT-Infrastruktur:  
GNZ, GWDG
  - Softwarefirma



Quelle: [https://commons.wikimedia.org/wiki/File:Relief\\_Map\\_of\\_Germany.png](https://commons.wikimedia.org/wiki/File:Relief_Map_of_Germany.png)

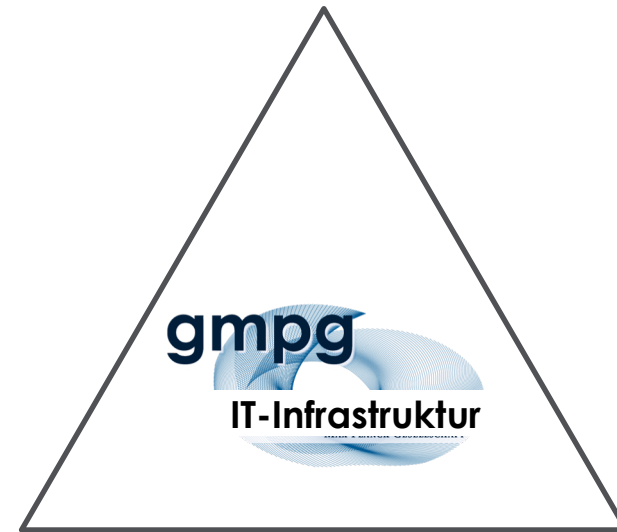
# Datenschutz & Sicherheit

- Rechtliche Vorgaben (u.a DSGVO, Archivrecht)
- Schutzstufenkonzept der MPG:  
normal / hoch / sehr hoch
- Hauptsächlich Daten der Stufe II & III
- Besondere Aktenzugangsregelung für GMPG
- Eigenes GMPG-Datenschutzkonzept



# Herausforderung

- Ziel ist es,  
so **viele** Daten wie möglich  
so **schnell** Zeit wie möglich  
so **sicher** wie notwendig  
zu prozessieren



**Big Data**

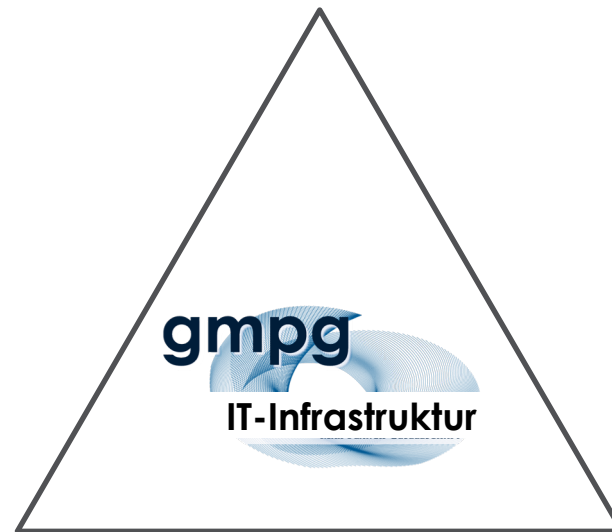
**Projektcharakter**

**Datenschutz & Sicherheit**

# Technische Umsetzung

Anforderungen

**Datenschutz & Sicherheit**



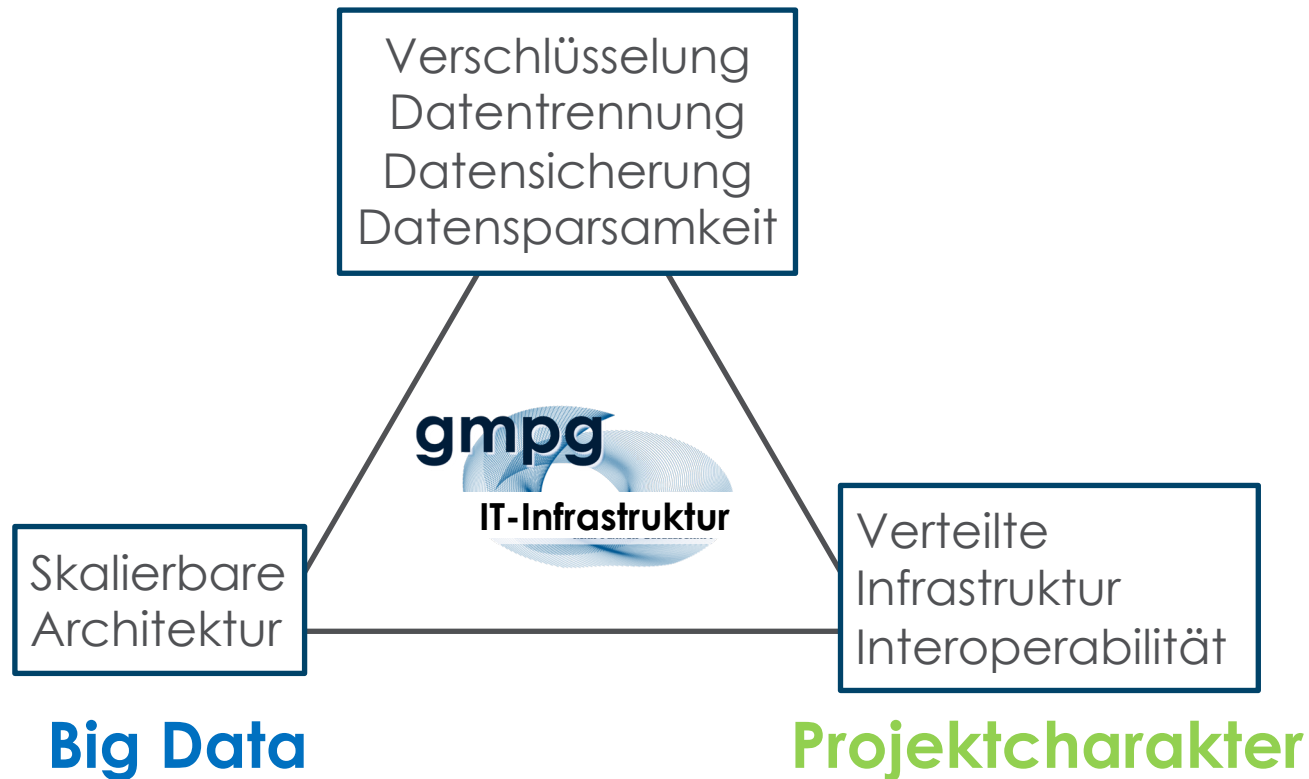
**Big Data**

**Projektcharakter**

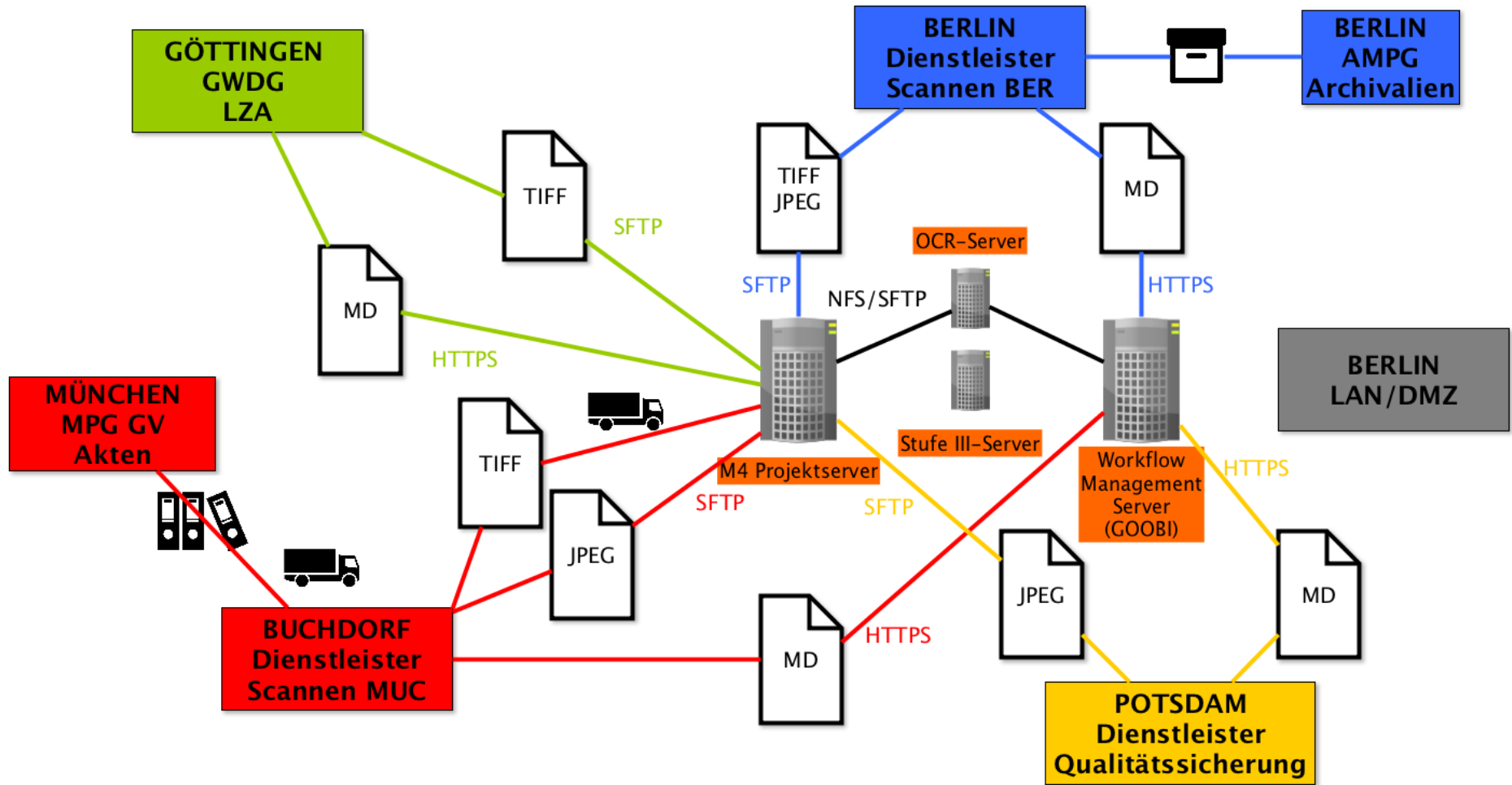
# Technische Umsetzung

## Anforderungen

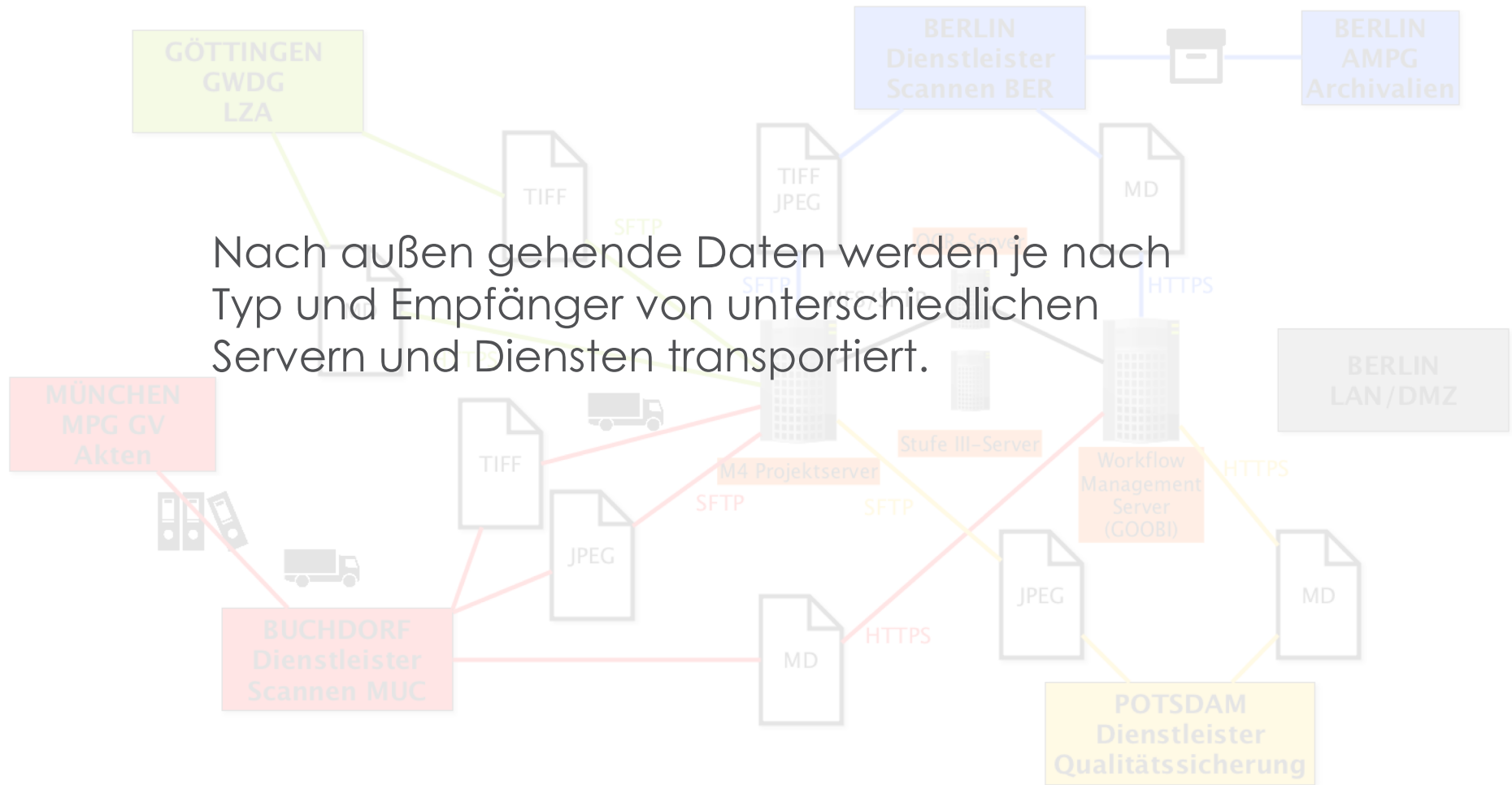
### Datenschutz & Sicherheit



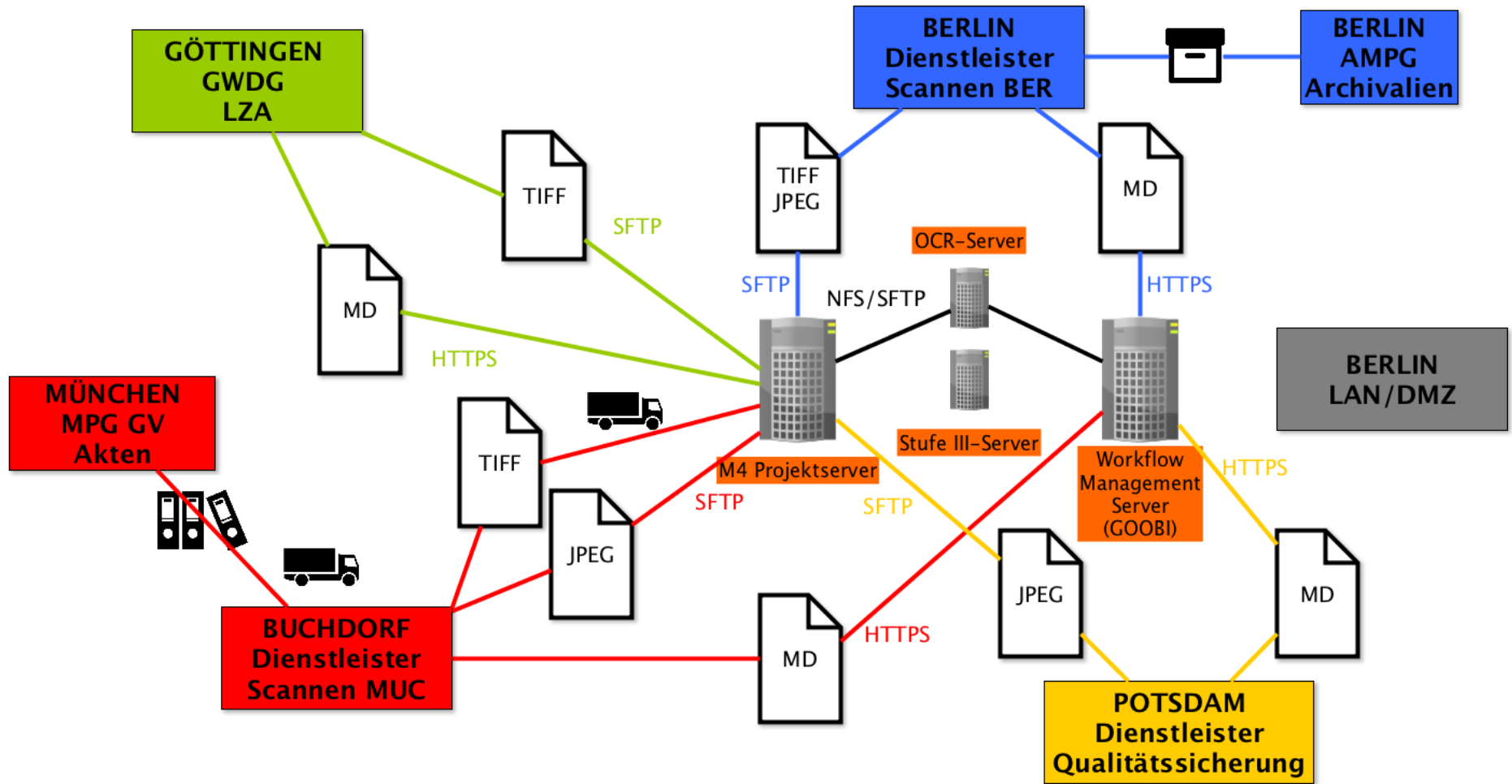
# Workflow-Topologie



# Netzwerktopologie



# Netzwerktopologie





# Softwarearchitektur

Entwurfsentscheidungen

Gretchenfrage  
*Make or buy?*

# Softwarearchitektur

## Buy: Prozesssteuerung

The screenshot shows the GOOBI-Workflow web interface. The browser address bar indicates the URL is localhost:8888/goobi/ui/process\_all.xhtml. The interface has a blue header with navigation tabs: Startseite, Meine Aufgaben, Workflow, Administration, and Controlling. The user is logged in as 'Goobi Administrator' at 00:14:50 on Saturday, 01. Juni 2019.

The main content area is titled 'Vorgänge' and shows a list of 28 items. The list has columns for 'Vorgangstitel', 'Status', 'Projekt', and 'Aktionen'. The status is represented by a progress bar with green, yellow, and red segments. The projects listed are 'Archive\_Project' and 'Manuscript\_Project'.

Vorgangstitel	Status	Projekt	Aktionen
> adolherr_612900487		Archive_Project	
> alphrudeh_591416573		Manuscript_Project	
> bergsphi_625017145		Manuscript_Project	
> bescdeaud_523689713		Manuscript_Project	
> bianjani_618686231		Manuscript_Project	
> biannoti_523090374		Manuscript_Project	
> bojaanat_620650583		Manuscript_Project	
> dissinsis_626789478		Archive_Project	
> fisclett_609336819		Manuscript_Project	
> forsindi_600894029		Manuscript_Project	

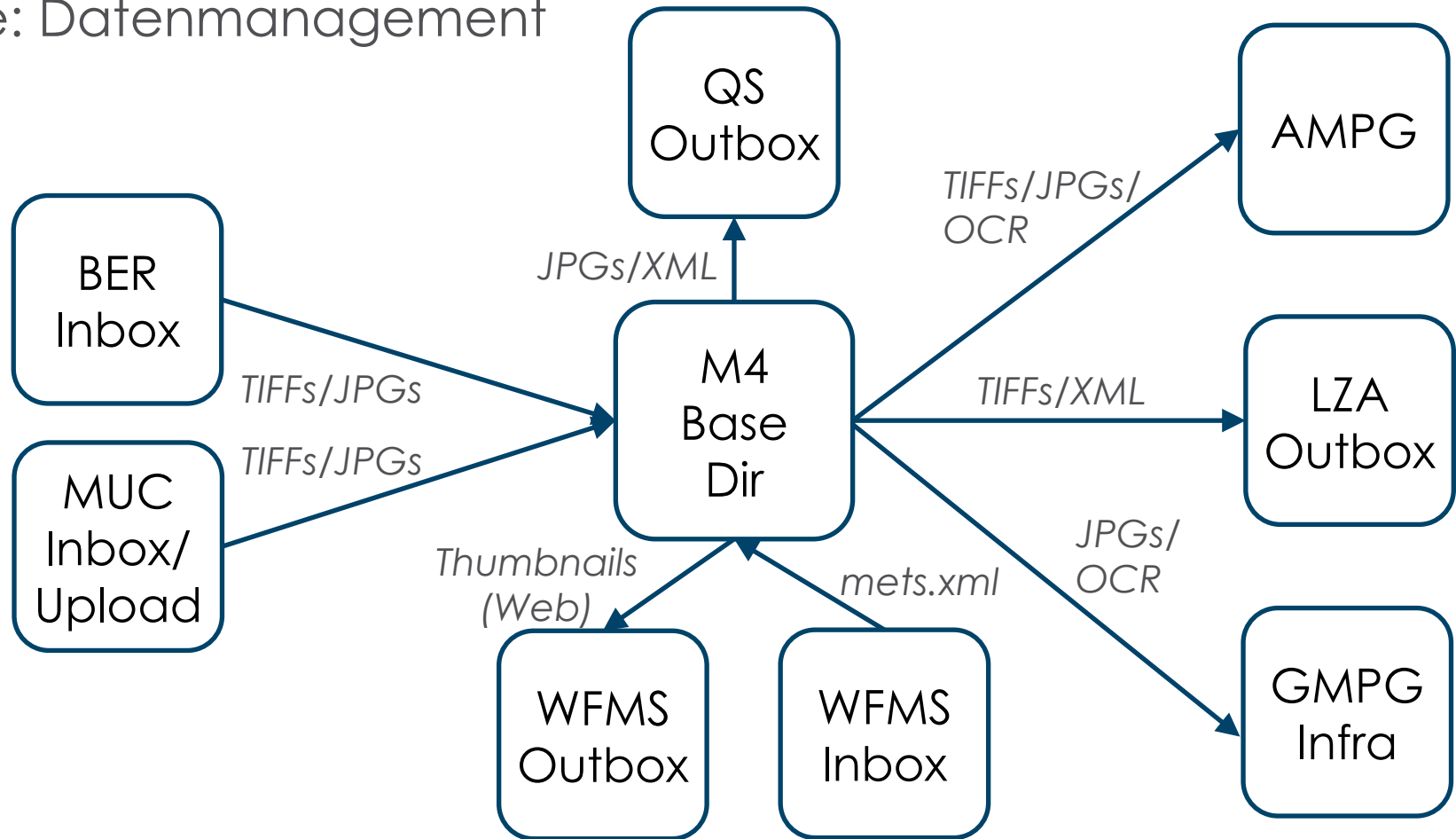
At the bottom, there is a section for 'Mögliche Aktionen' (Possible Actions) with the following options:

- Massenexport durchführen
- Bearbeitungsstatus heruntersetzen?
- Bearbeitungsstatus hochsetzen?
- GoobiScript ausführen
- Suchergebnis exportieren

Screenshot GOOBI-Workflow,  
Quelle: <https://docs.intranda.com/goobi-workflow-de/manager/7>

# Softwarearchitektur

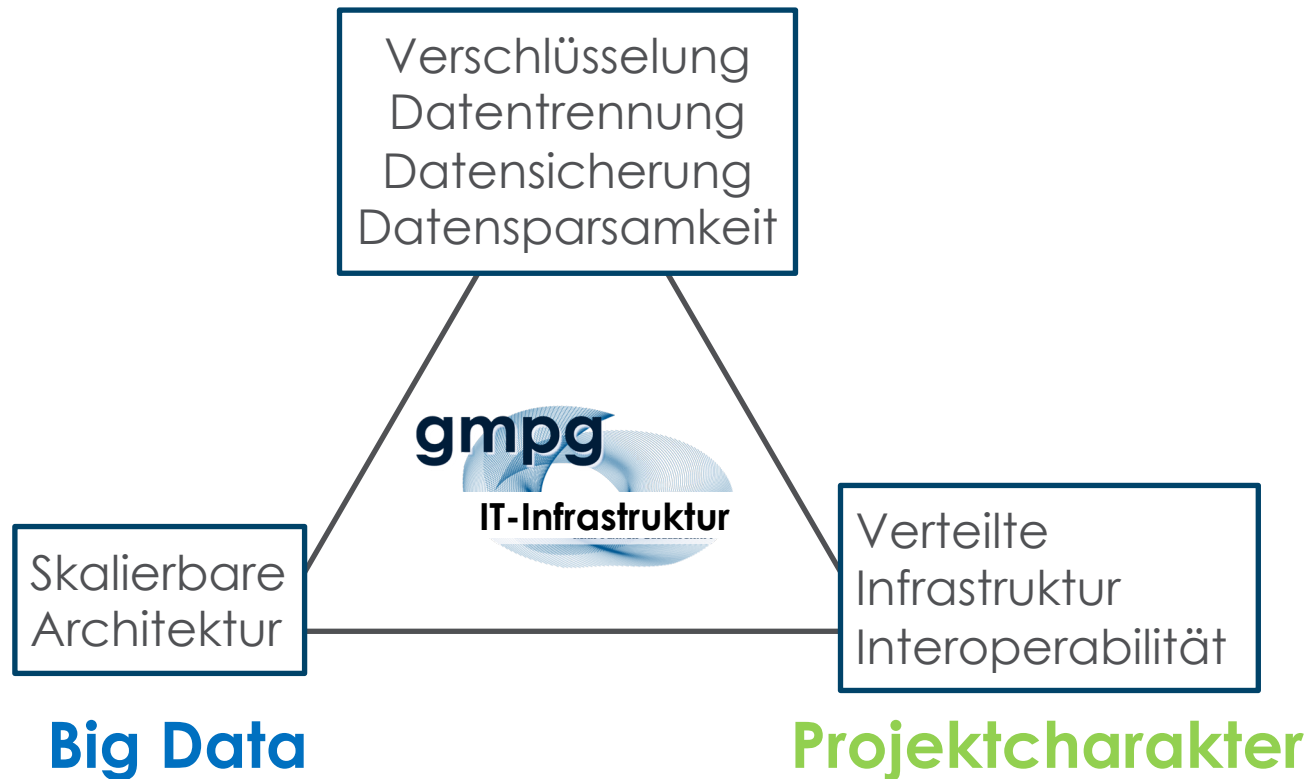
Make: Datenmanagement



# Technische Umsetzung

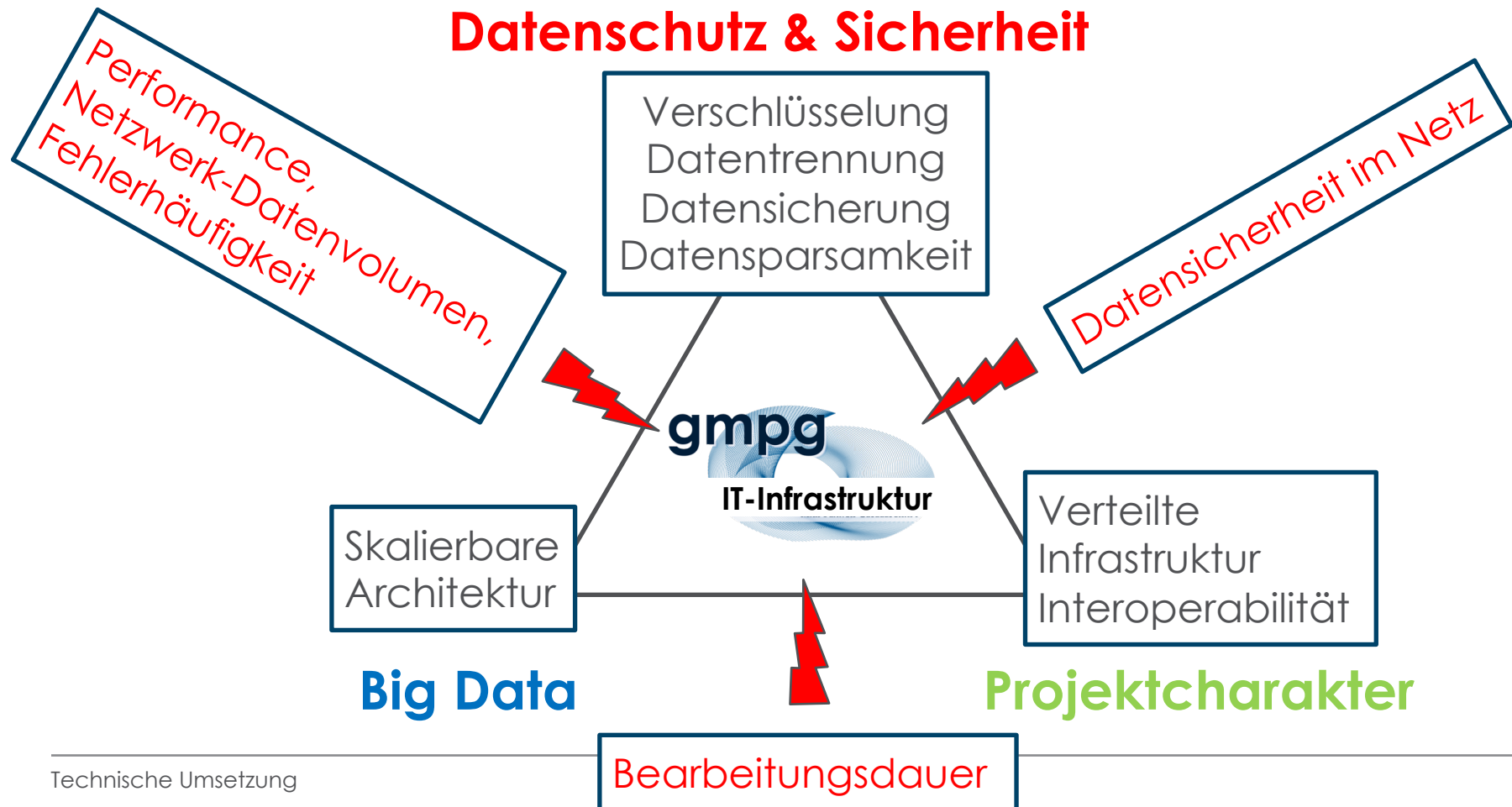
Optimierung hinsichtlich gegebener Randbedingungen

## Datenschutz & Sicherheit



# Technische Umsetzung

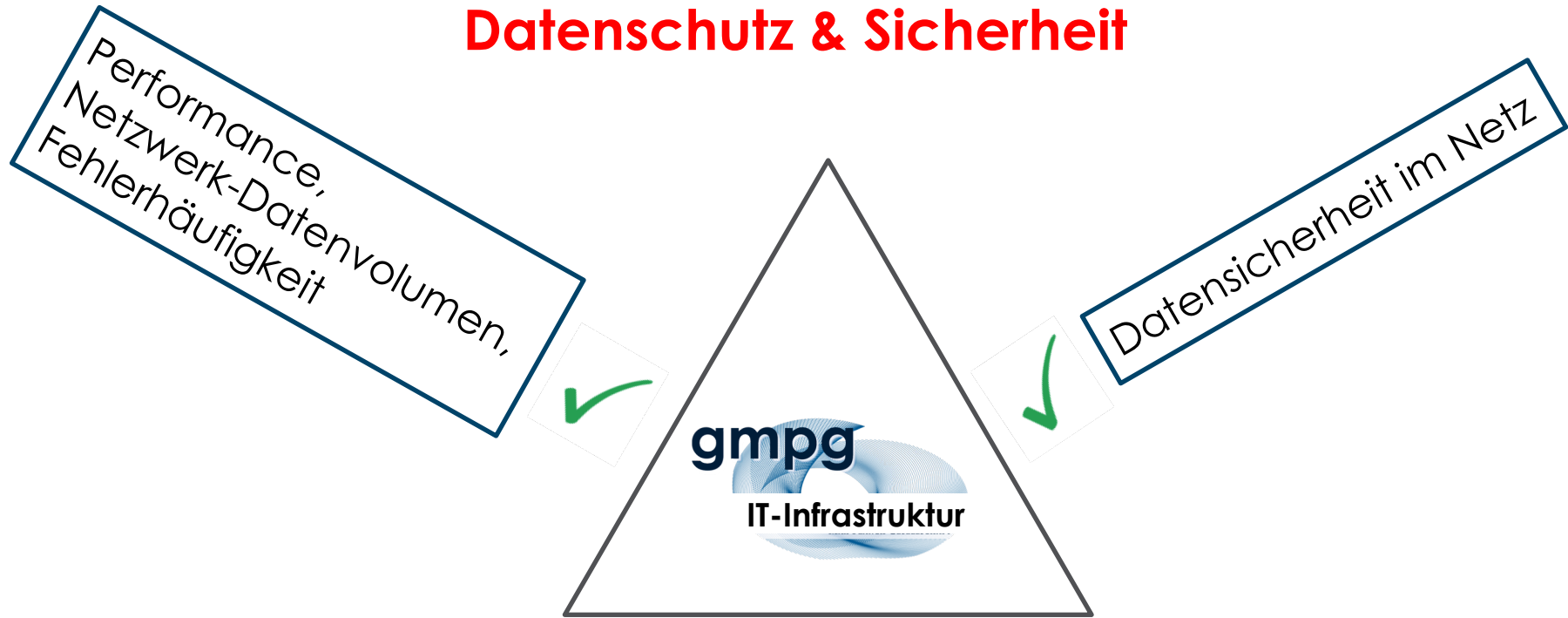
Potentielle Zielkonflikte



# Technische Umsetzung

Lösungen

## Datenschutz & Sicherheit



Big Data



Projektcharakter

Bearbeitungsdauer

# Lessons Learned

## Datentransfer/-prozessierung

- Skalierungsprobleme (Performance/Stabilität) liegen eher bei Hardware/OS/Netzwerk als bei der (Workflow-)Software.  
  
→ In Projektkontext schwer plan- und testbar
- Mobile Datenträger mit UPS unter Umständen schneller als (S)FTP



Rocstor DE5 Case Waterproof Military Standard, Quelle: Rocstor

---

# Lessons Learned

## Softwarearchitektur

- Im Ergebnis hat sich Mischung aus proprietärer Software (mit Beratung) und Eigenentwicklung bewährt
- Reagieren auf Erfahrungen aus der laufenden Produktion



# Lessons Learned

## Sicherheit

- Unmöglich, Datensicherheit in einem performance intensiven Workflow „umsonst“ zu bekommen.
- Zusätzliche Kosten
  - Personell: IT-Monitoring-Aufwand ist höher
  - Zeitlich: Prozesse dauern länger
  - Technisch: Workflows sind aufwändiger
  - Finanziell: Spezifischer Hardware notwendig
- Zuweilen Inkompatibilität mit präferierten IT-Standardlösungen

---

Vielen Dank für Ihre Aufmerksamkeit.



Kontakt:

**Felix Lange & Dr. Felix Schäfer**

Forschungsprogramm

Geschichte der Max-Planck-Gesellschaft

Max-Planck-Institut für Wissenschaftsgeschichte

Boltzmannstraße 22

14195 Berlin

[flange@mpiwg-berlin.mpg.de](mailto:flange@mpiwg-berlin.mpg.de)

[ffschaefer@mpiwg-berlin.mpg.de](mailto:ffschaefer@mpiwg-berlin.mpg.de)

[gmpg.mpiwg-berlin.mpg.de](http://gmpg.mpiwg-berlin.mpg.de)