



# Design von Dateninfrastrukturen

das FENIX-Konzept

Daniel Mallmann  
Forschungszentrum Jülich GmbH  
Jülich Supercomputing Centre

# Das Human Brain Project

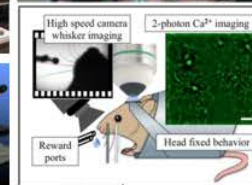
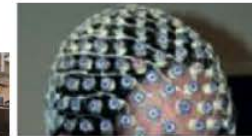
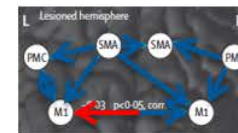
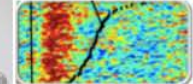
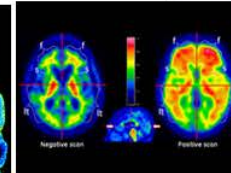
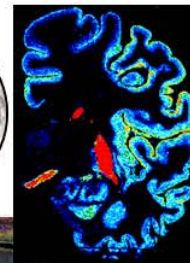
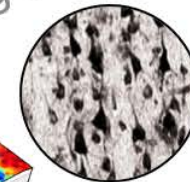
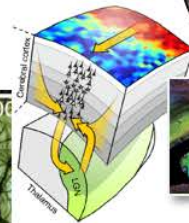
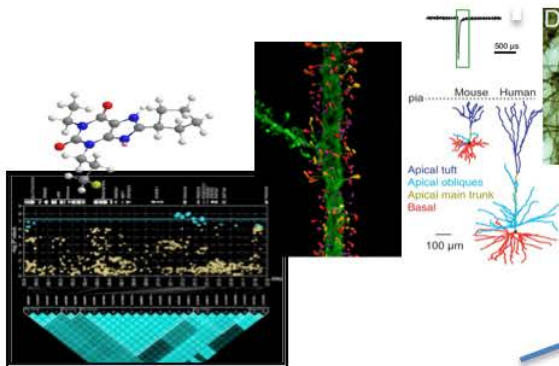
## ETHIK & GESELLSCHAFT

### FORSCHUNGSINFRASTRUKTUR DATENANALYSE & SIMULATION

Decoding the Human Brain

Multiskalensystem in Raum und Zeit, multimodal

### NEUROWISSENSCHAFT EXPERIMENT & THEORIE



Amunts et al., The Human Brain Project: Towards a European Infrastructure to Decode the Human Brain; Neuron, 2016

# Human Brain Project

## Herausforderung

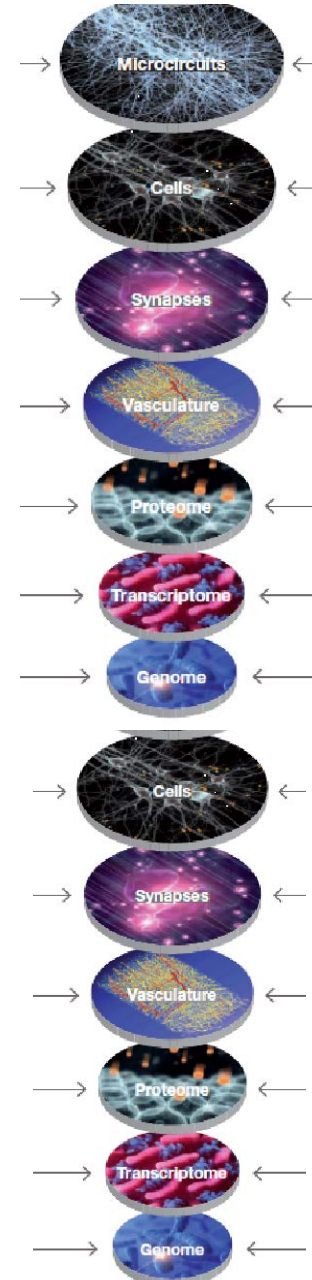
### Komplexität des menschlichen Gehirns

- 100 Milliarden Neuronen
- 100 Billionen Synapsen
- Leistungsaufnahme: 30 Watt
- Über 100.000 neuro-wissenschaftliche Veröffentlichung pro Jahr

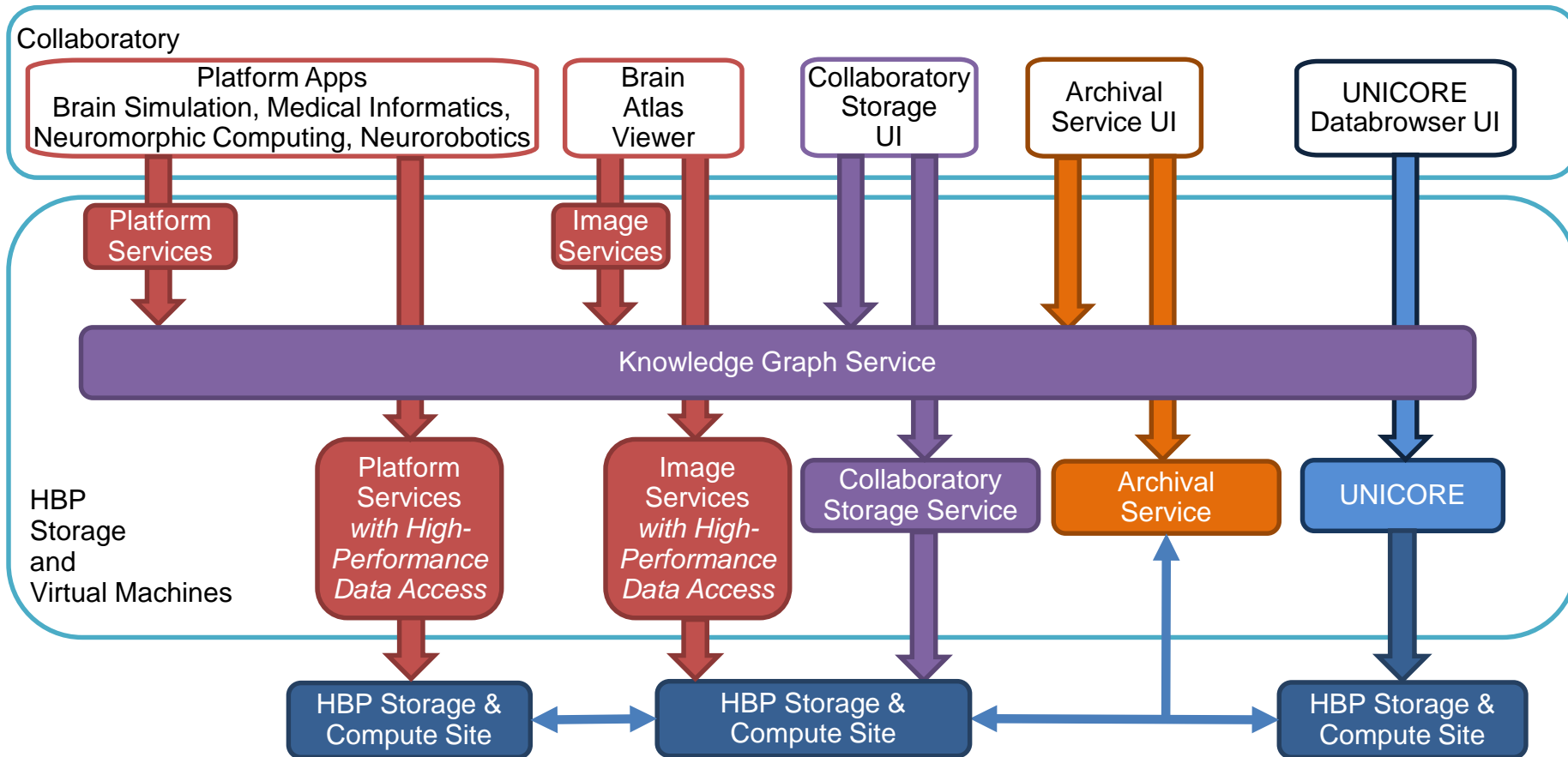
## Herausforderung

### Verständnis des menschlichen Gehirns

- Molecular
- Cellular
- Brain regions
- Whole brain



# Dateninfrastruktur des Human Brain Project



# HBP Collaboratory

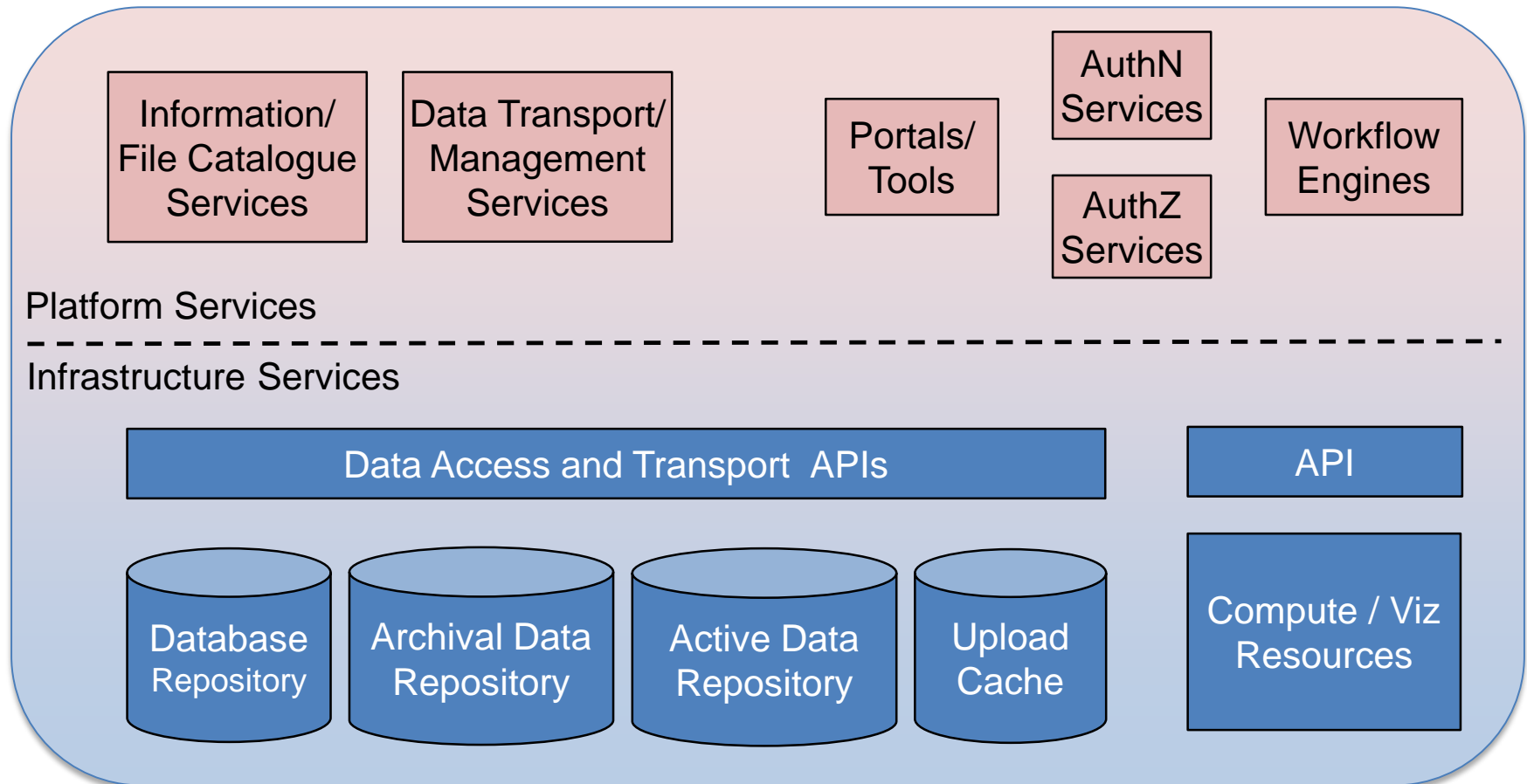
- Community-spezifisches Software Deployment (SaaS)
- Entwicklungsplattform (PaaS)
- Datenmanagement
- Data Curation
- Provenance Tracking
- Interaktive Datenanalyse (lokal und remote)
- Visualisierung von Batch-prozessierten Daten, interaktiv über Web
- Direkter und indirekter Zugriff auf HBP Storage Sites (IaaS)

# FENIX

## Grundlegende Eigenschaften und Funktionen

- Enge Kopplung von Daten- und Computerressourcen für die Analyse
- Replikation von wesentlichen Daten für Langlebigkeit, Verfügbarkeit, und Zugriffsperformanz
- Communities mit hohem Grad an Selbstorganisation sollen FENIX als „IaaS“ nutzen können
- Sicherheit, Flexibilität, Erweiterbarkeit und Skalierbarkeit
- Nutzung von offenen Standards für Infrastruktur Dienste
- Geprägt durch die Anwendungsfälle der verschiedenen Communities

# FENIX – Federated EuropeaN Information eXchange





# FENIX Storage Services

Archival Data Repository	Active Data Repository
<ul style="list-style-type: none"> <li>• Optimiert für Kapazität, Langlebigkeit, Verfügbarkeit</li> <li>• Großen Datenmengen, die nicht neu erstellt werden können und die dauerhaft zugreifbar sein müssen</li> </ul>	<ul style="list-style-type: none"> <li>• Platzierung nahe an Compute- und Visualisierungsressourcen</li> <li>• Temporäre (Read-only) Kopien für höhere Performanz</li> <li>• Datenverlust führt zu neuer Replikation</li> </ul>
Database Repository	Upload Cache
<ul style="list-style-type: none"> <li>• Optimiert für Suche und Abruf von kleinen Datenmengen</li> <li>• Permanente Datenablage (z.B. für Metadaten), remote-Zugriff ausreichend schnell</li> <li>• Datenbankreplikation zur Erhöhung des Lesezugriffs</li> </ul>	<ul style="list-style-type: none"> <li>• Platzierung nahe an Datenquellen (nicht-HPC)</li> <li>• Hält temporäre Kopien von Großen Datenmengen, die nicht neu erstellt werden können, bevor sie ins „Archival Data Repository“ transferiert werden</li> </ul>



# FENIX

## Erfahrungen aus HBP

- Infrastruktur (IaaS)
  - Ebene universeller, föderierter Forschungsdateninfrastrukturen
  - Interoperabilität
  - Vertrauen
  - Langfristige Perspektive
  - Vernetzung
- Plattform (PaaS), Software (SaaS)
  - Ebene der projekt-spezifischen Infrastrukturelemente
  - Akzeptanz der Benutzer
  - Metadatengenerierung durch die Forscher
  - Provenance Tracking