

# MAX PLANCK INSTITUTE

FOR  
HUMAN  
COGNITIVE AND BRAIN SCIENCES  
LEIPZIG



# Data Infrastructure at the MPI for Human Cognitive and Brain Sciences (MPI CBS)

---

**Dr Roberto Cózatl**

Coordinator, Database Management Group, MPI CBS

MPG eScience Seminar 2009

Repository Systems

Munich/Garching 25 & 26 June 2009

- Introduction (Background)
- Nature of our data
- Infrastructure and current strategy
- Our needs
- Roadmap and the future

# MPI for Human Cognitive and Brain Sciences – Background

1994: Max Planck Institute of Cognitive Neuroscience,  
Leipzig

Prof. Dr. A. D. Friederici, Prof. Dr. D. Y. von Cramon



1998: New building Stephanstrasse



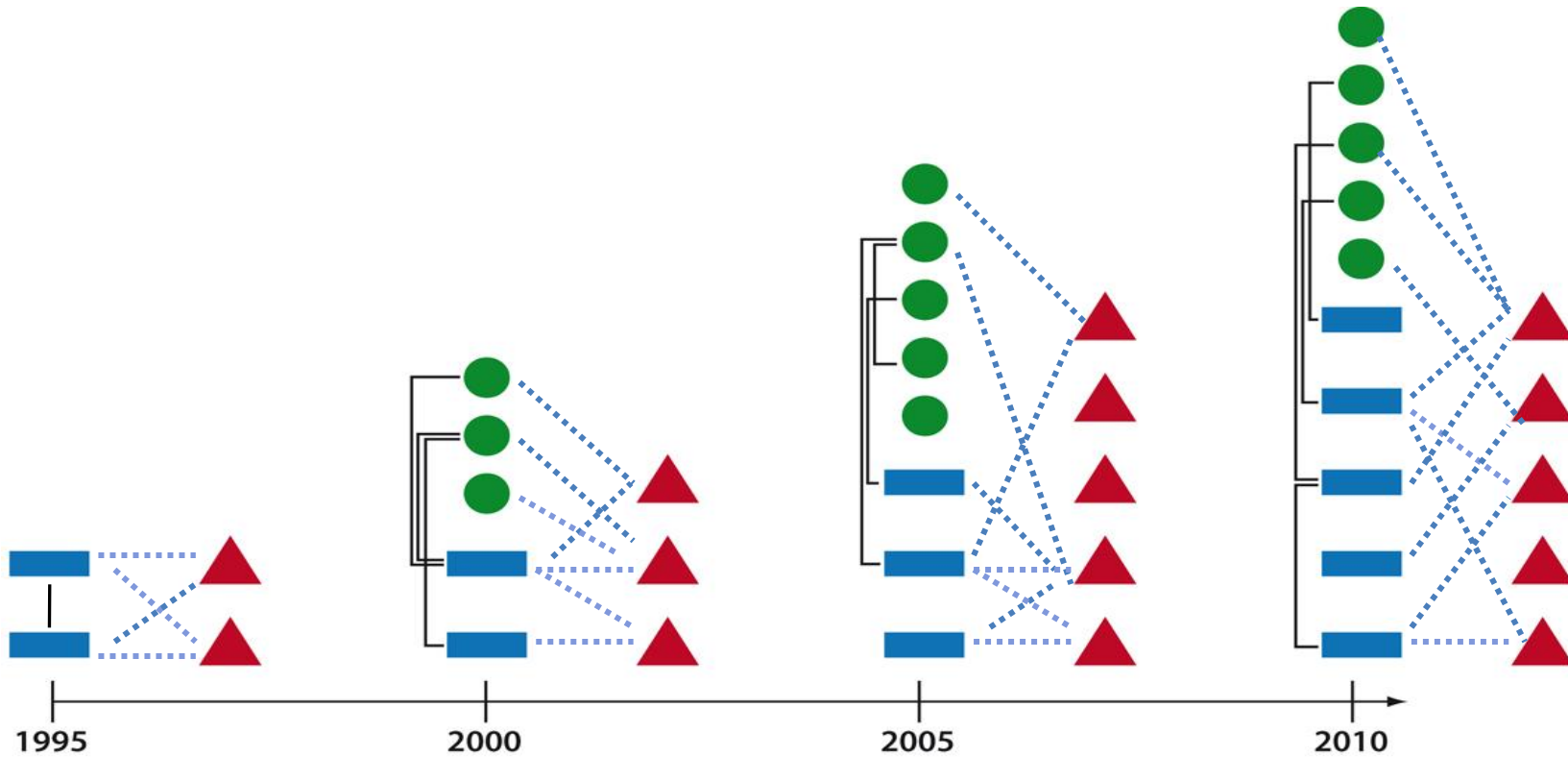
2004: Fusion with Max Planck Institute for Psychological  
Research, Munich (est. in 1981)

Prof. Dr. W. Prinz

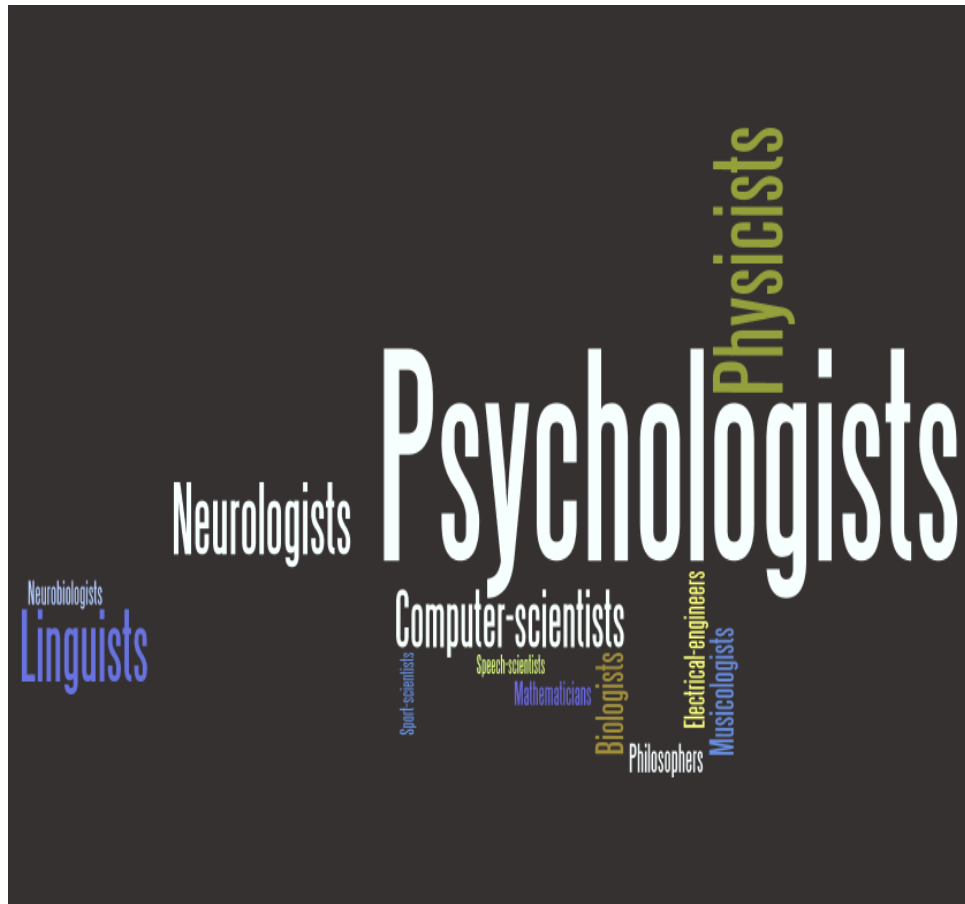


2004: Max Planck Institute for Human Cognitive and Brain  
Sciences, Leipzig

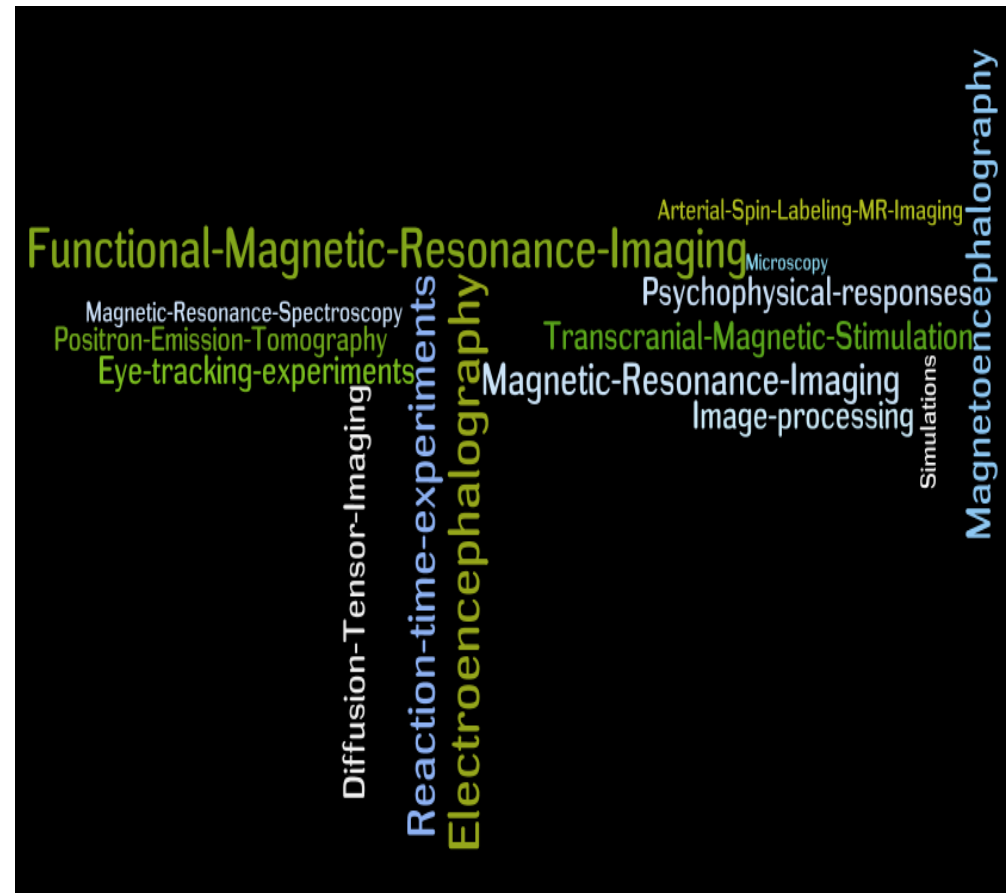
Prof. Dr. A. D. Friederici, Prof. Dr. W. Prinz, Prof. Dr. R. Turner, Prof. Dr. A. Villringer



- Departments
- ▲ Methods
- Research groups

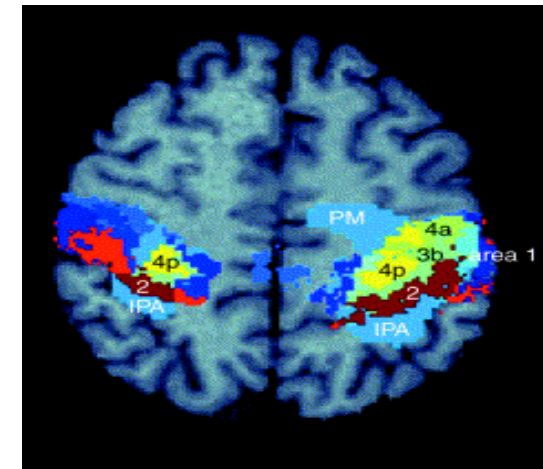
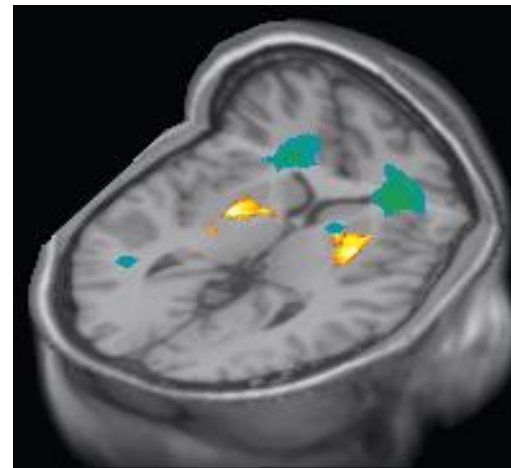


+



=

Data Diversity & Complexity



**Technical metadata:** image acquisition parameters, resolution, modality, data format

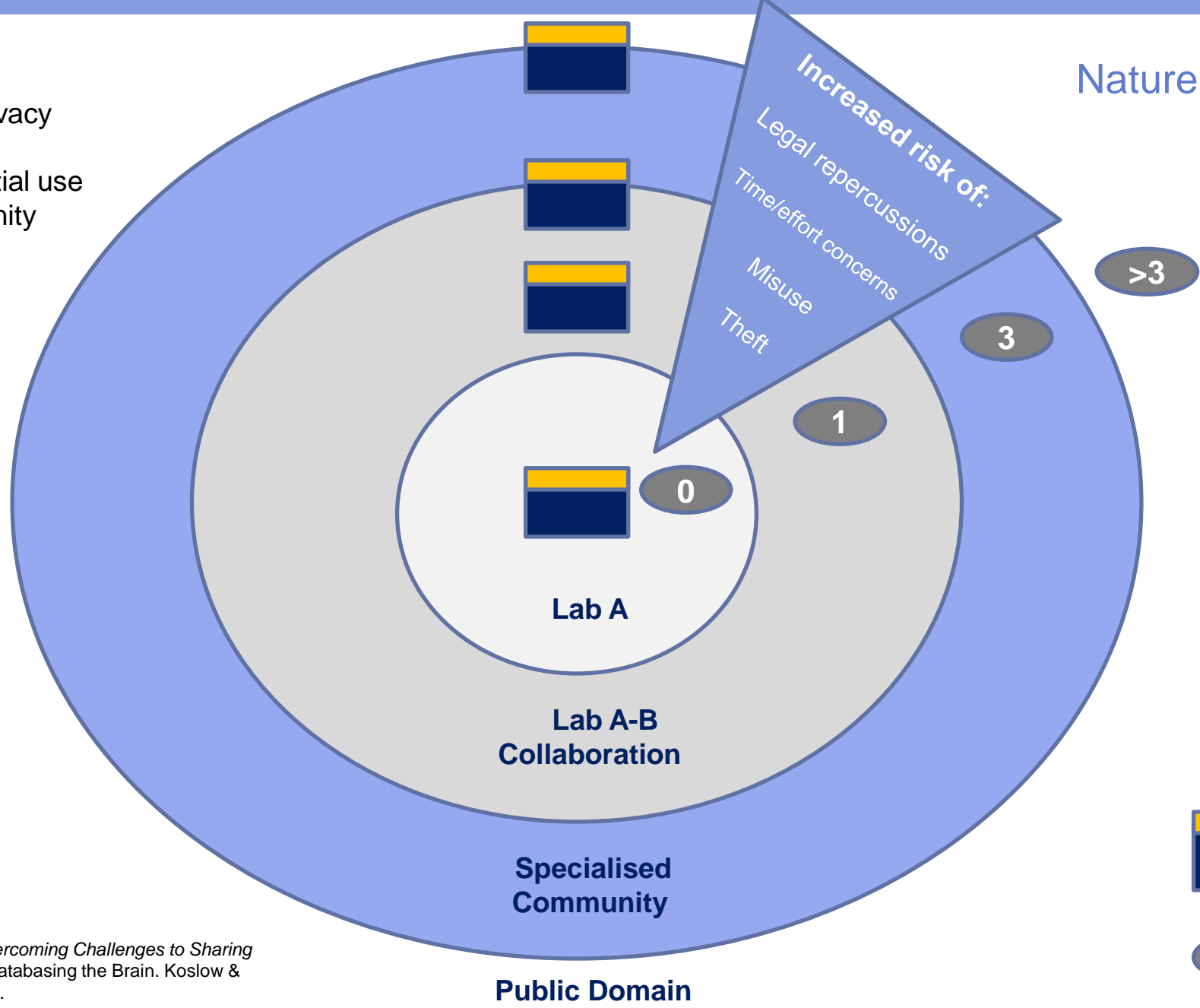
**Subject metadata:** age, gender, handedness, medical or clinical history

**Experimental metadata:** experimental setting, experimental design, hypothesis

**Data access/risk management:** data privacy, rights management, controlled data sharing

# Nature of our Data

- \* Data privacy
- \* Differential use of community standards



Adapted from:  
 Smith, K. 2005. *Overcoming Challenges to Sharing Neuroimaging*. In: *Databasing the Brain*. Koslow & Subramaniam (eds.).



General principles: data security, infrastructure flexibility and scalability, good computational power, and data accessibility and compatibility

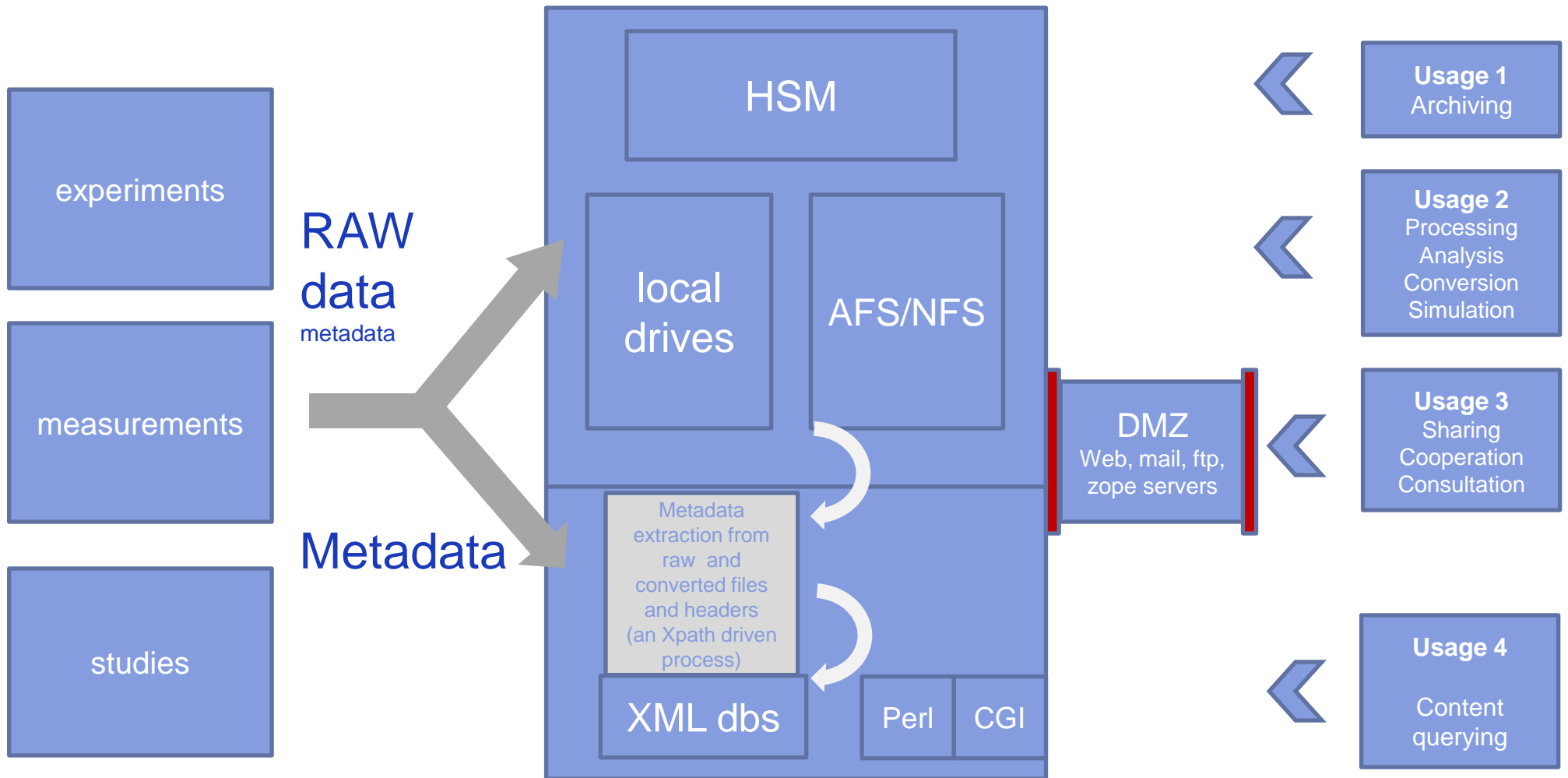
A distributed environment with ca. 750 work stations; 70 servers; Linux, Win, Mac OSs supported, RAID systems directly connected (NAS Server) ca. 55 TB

NFS – standard client/server architecture, efficient file sharing services under a Linux network, facilitates backups and updates through a centralised administration. Used for data exchange, analysis, mails, documents, scripts, etc.

AFS – transparent access to files, adequate data security, good control over data accessibility, basic authentication. Used for raw data from projects and studies

HSM - (SUN) Server ca.104 TB (easily expandable) – suits our high ratio of nearline to online data

# Infrastructure – Strategy

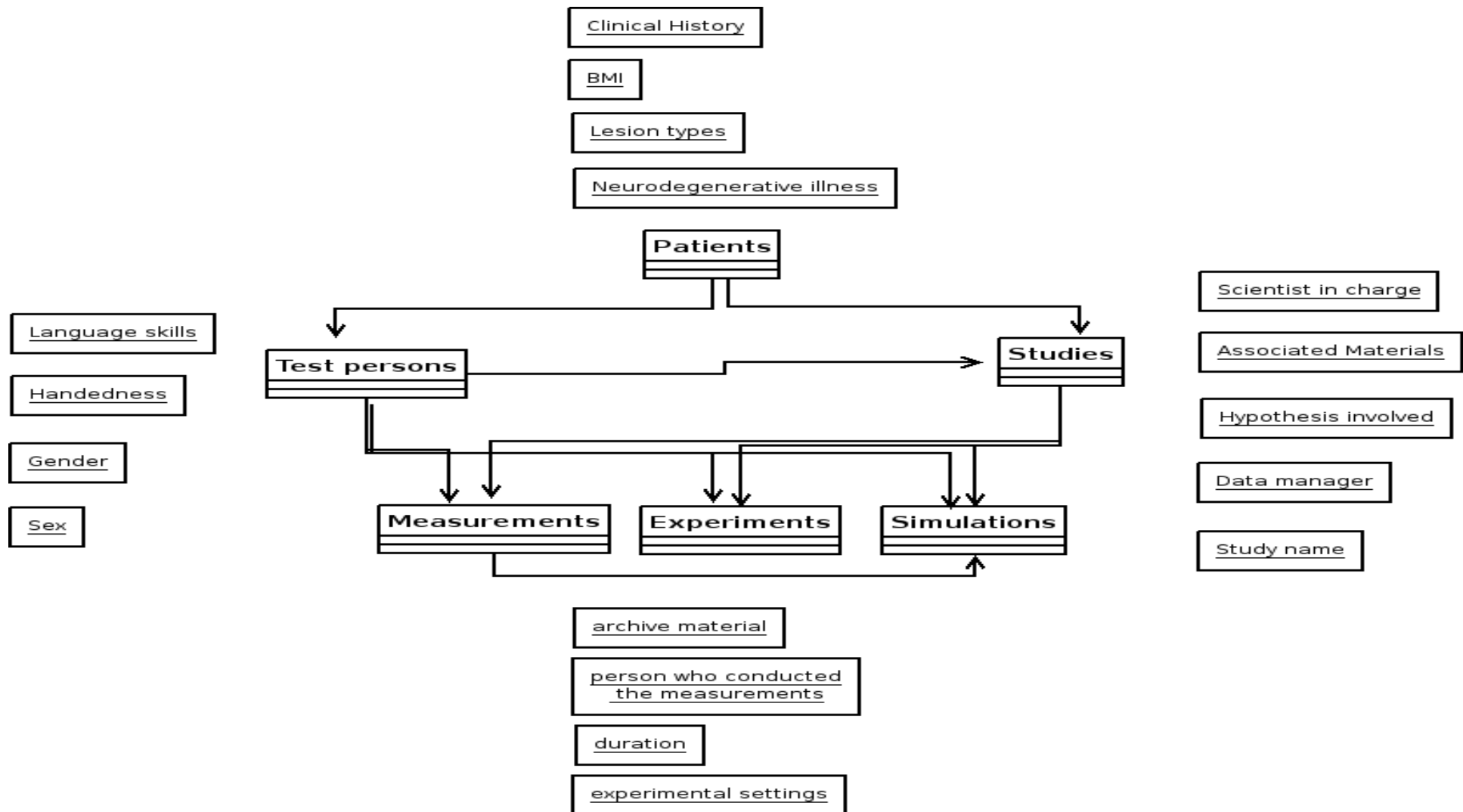


The screenshot shows the website <https://data.cbs.mpg.de/Startseite> in a Mozilla Firefox browser. The page features the Max Planck Institute logo and the text "MAX PLANCK INSTITUT FÜR KOGNITIONS- UND NEUROWISSENSCHAFTEN LEIPZIG". Below the logo is a navigation menu with the following items: DB Berlin, DB GLaD, DB Infant, DB Patient, DB Proband, DB Studien, DB Schulen, DB Research, Internetanmeldungen, and Bestellliste. The main heading reads "Willkommen auf der Startseite unserer Datenbanken". Three blue callout boxes provide details:

- db Patients:** patients from the University of Leipzig Day Clinic who get scanned at our MPI (restricted access) - points to DB Patient.
- db Proband:** data from test persons who participate in various experiments at our MPI (restricted access to personal data) - points to DB Proband.
- db Studies:** data from studies and experiments - points to DB Studien.

On the right side, the text "metadata xml url-based" is displayed in red.

# Infrastructure – Strategy



Tools for the automatic extraction and better management of metadata from our various research data types

Right-management tools that can help us to implement significant custodial responsibilities by data owners which are understood by both humans and software and that increase the trust of data producers in our repository

Enhancements to our repository to be able to ingest and adequately preserve adjacent/complementary material derived from experiments and studies that are important for interpretation, replication, and contextualisation of results

Infrastructural developments that contemplate work with – if necessary – proprietary solutions (i.e. software, middleware) and that can help us to correctly express and define our sharing policies and intents with collaborators and or other communities at the academic, industry, and public level

Writing of more reference and consultation documents as well as the creation and enhancement and updating of our current data policies

Continue to improve our capacity to deal with data aggregation and linkage of datasets as well as our capacity to map data and metadata flows at our Institute

Increase the collaboration with other professional organisations (data archives, test beds, validation services) to seek technical solutions that suit our needs. These might include solutions for data protection, long-term preservation, authentication, risk management, and data repository assessment methods



MAX  
PLANCK  
INSTITUTE

FOR  
**HUMAN  
COGNITIVE AND BRAIN SCIENCES**  
LEIPZIG

Thank you!

- Organisers of the eScience Seminar
- Colleagues at CBS-MPI Leipzig, esp.

**Elke Maess**

db Administrator, Database Management Group

**Dr. Helmut Hayd**

Cordinator, IT Group

**Dr. Stefan Kiebel**

Group Leader, Department of Neurology

**Dr. Burkhard Maess**

Group Leader, MEG/EEG: Signal Analysis and Modelling

**Enrico Reimer**

Programmer, Department of Neurophysics

- Collaborators UK Data Archive,  
University of Essex

**Prof. Kevin Schürer**

Director

**Dr. Matthew Woollard**

Head of Digital Preservation and Systems