

Frank Hertel<sup>a</sup><sup>a</sup> Institut für Biometrie und Medizinische Informatik, Medizinische Fakultät,  
Otto-von-Guericke-Universität Magdeburg

## D1.2: Erweiterter Use Case – Bilddaten<sup>1</sup>

	D1.1: Use Case – Bilddaten
Autor(en)	Frank Hertel
Editor(en)	Frank Dickmann
Datum	07.12.2012

### A: Status des Dokuments

Version 1.0.0

### B: Bezug zum Projektplan

M1.1: Dokumentierter Use Case – Biomedizinische Bilddaten, Version 2

### C: Abstract

*Für die Langzeitarchivierung von Forschungsdaten aus dem UseCase biomedizinische Bilddaten ist es wichtig Informationen über die Herkunft der Ergebnisse, die auf Analysen von Primärdaten basieren, zu erfassen. Diese Informationen über Auswertemethoden und -Werkzeuge sowie ihrer Arbeitsschritte und Parameter werden als Provenienz der Forschungsdaten bezeichnet. Für ein Forschungsdatenmanagementsystem, als Grundlage der Langzeitarchivierung, ist es wichtig die heterogene Informationsvielfalt aus der neurowissenschaftlichen Bildgebung in einem generischen Datenmodell erfassen zu können.*

---

<sup>1</sup>Dieses Dokument wurde im Rahmen des Projekts LABIMI/F erstellt. Das Projekt LABIMI/F wird gefördert von der Deutschen Forschungsgemeinschaft (DFG) unter dem Förderkennzeichen RI1000/2-1.

## D: Änderungen

Version	Datum	Name	Kurzbeschreibung
0.0.1	12.01.2012	Frank Hertel	Entwurf
0.0.2	02.07.2012	Frank Hertel	Inhaltsverzeichnis
0.0.3	23.08.2012	Frank Hertel	Kapitel 1
0.0.4	01.10.2012	Frank Hertel	Kapitel 2.1
0.0.5	30.10.2012	Frank Hertel	Kapitel 2.2
0.0.6	12.11.2012	Frank Hertel	Kapitel 3.1
0.0.7	15.11.2012	Frank Hertel	Kapitel 3.2
0.0.8	19.11.2012	Frank Hertel	Kapitel 4
0.0.9	26.11.2012	F. Dickmann, R. Grütz	Review
1.0.0	07.12.2012	Frank Hertel	Korrekturen laut Review

## E: Inhaltsverzeichnis

1	Einleitung .....	4
2	Material und Methoden .....	5
2.1	Provenienz - von den Primärdaten zur Veröffentlichung.....	5
2.1.1	Rohdaten und Primärdaten .....	6
2.1.2	Aufbereitete Daten.....	7
2.1.3	Ausgewertete Daten .....	7
2.1.4	Veröffentlichte Ergebnisse.....	7
2.2	Auswertewerkzeuge für fMRT und DTI .....	8
2.2.1	BrainVoyager.....	9
2.2.2	Statistical Parametric Mapping .....	10
2.2.3	FMRIB's Software Library .....	10
2.2.4	FreeSurfer .....	12
2.2.5	Caret.....	12
2.2.6	AFNI und SUMA .....	12
2.2.7	3DSlicer.....	12
2.2.8	MedINRIA.....	13
2.2.9	NeuroELF .....	13
2.2.10	MRI Studio .....	13
2.2.11	NeuroDebian .....	13
2.2.12	Fazit .....	14
3	Ergebnisse .....	15
3.1	Metadatenmodelle zur Verwaltung von MR-Projekten .....	15
3.1.1	DICOM .....	15
3.1.2	XCEDE 2 .....	17
3.1.3	CSMDM.....	17
3.1.4	Subject-centric Research Object Model .....	18
3.1.5	EDL .....	19
3.2	Archivierungstoolkits .....	20
3.2.1	eXtensible Neuroimaging Archive Toolkit .....	20
3.2.2	DFBidb Software for Neuroimaging Datamanagement.....	22
3.2.3	Datafinder.....	23

4	Diskussion und Ausblick .....	24
4.1	Wahl des Forschungsdatenmanagementsystems .....	24
4.2	Fallbeispiel Leibniz-Institut für Neurobiologie Magdeburg.....	25
4.3	Daten durch DOIs via REST-API referenzieren .....	26
5	Literaturverzeichnis .....	27

## F: Abbildungsverzeichnis

Abb. 1:	Die Dokumentation von Provenienz steigert die Ebene des Verständnisses am Ende eines wissenschaftlichen Arbeitsablaufs. ....	5
Abb. 2:	Durch Fourier-Transformation werden MRT-Daten vom Frequenzraum in den Bildraum überführt. ....	6
Abb. 3:	Im Vergleich zu der Frequenzraumdarstellung (links) ist die Grauwertdarstellung (rechts) für den Menschen verständlich. ....	6
Abb. 4:	Verschiedene Software-Werkzeuge erzeugen unterschiedliche Analyseergebnisse aus den Primärdaten .....	8
Abb. 5:	Die Versionsgeschichte von BrainVoyager stellt zahlreiche Funktionalitäten der Software dar .....	9
Abb. 6:	Vergleich verschiedener Metadatenformate hinsichtlich ihrer Datenobjekte	15
Abb. 7:	Objektverknüpfungen des Real-World-Modells aus dem DICOM-Standard PS 3.3-2011 .....	16
Abb. 8:	Schematische Darstellung der Objekthierarchie in XCEDE 2 .....	17
Abb. 9:	Die Entitäten des Core Scientific Metadata Models .....	18
Abb. 10:	Entitäten in der Subjekt-bezogenen Forschung .....	19
Abb. 11:	EDL wurde als Metadatenformat speziell für Echtzeit-FMRT-Experimente entwickelt.....	19
Abb. 12:	Die Drei-Schichten-Architektur von XNAT .....	21
Abb. 13:	Die Hierarchie der Entitäten in XNAT .....	22
Abb. 14:	Architektur der DataFinder Speicherinfrastruktur .....	23
Abb. 15:	Metadatenverwaltung mit DataFinder.....	23

## G: Tabellenverzeichnis

Tab. 1:	Merkmale und mehrstufige Ausprägungen zum Thema Forschungsdatenarchivierung nach der GMA-Analyse aus D4.1 .....	24
Tab. 2:	Bewertung der Metadatenformate aus Abschnitt 3.1 .....	25

# 1 Einleitung

Deliverable 1.1 befasste sich vorwiegend mit den Formaten, die bei einem Archivierungsmanagement beachtet werden müssen, sowie mit den Prozessen, bei denen zu archivierende Informationen und Metadaten geschaffen werden. Außerdem wurde bereits ein Ausblick auf vorhandene Frameworks für ein Forschungsdatenmanagement gegeben. Diese Konzepte sollen nun auf Tauglichkeit im Kontext der Langzeitarchivierung von bildbasierten Forschungsdaten untersucht werden. Durch die Darstellung und Analyse der verschiedenen Projektphasen ist es möglich, weiterführende Betrachtungen mit Fokus auf Metadaten für eine Langzeitarchivierung durchzuführen.

Richtlinien hierfür werden z.B. durch die DFG in der Denkschrift *Vorschläge zur Sicherung guter wissenschaftlicher Praxis* gegeben. Darin wird z.B. empfohlen, dass „Primärdaten als Grundlagen für Veröffentlichungen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für zehn Jahre aufbewahrt werden.“ [Deutsche Forschungsgemeinschaft 1998]. Das Referenzmodell des Open Archival Information System (OAIS) wurde durch das Consultative Committee for Space Data Systems (CCSDS) Ende der 90er Jahren ausgearbeitet und wurde 2003 in den ISO-Standard 14721 übernommen. In den letzten Jahren haben weltweit Rechenzentren und Archivbetreiber begonnen, dieses Protokoll anzuwenden. [Consultative Committee for Space Data Systems 2002] [International Organisation for Standardization 2010]. Auch die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) hat in *Principles and Guidelines for Access to Research Data from Public Funding* entsprechende Kriterien festgelegt [OECD 2007].

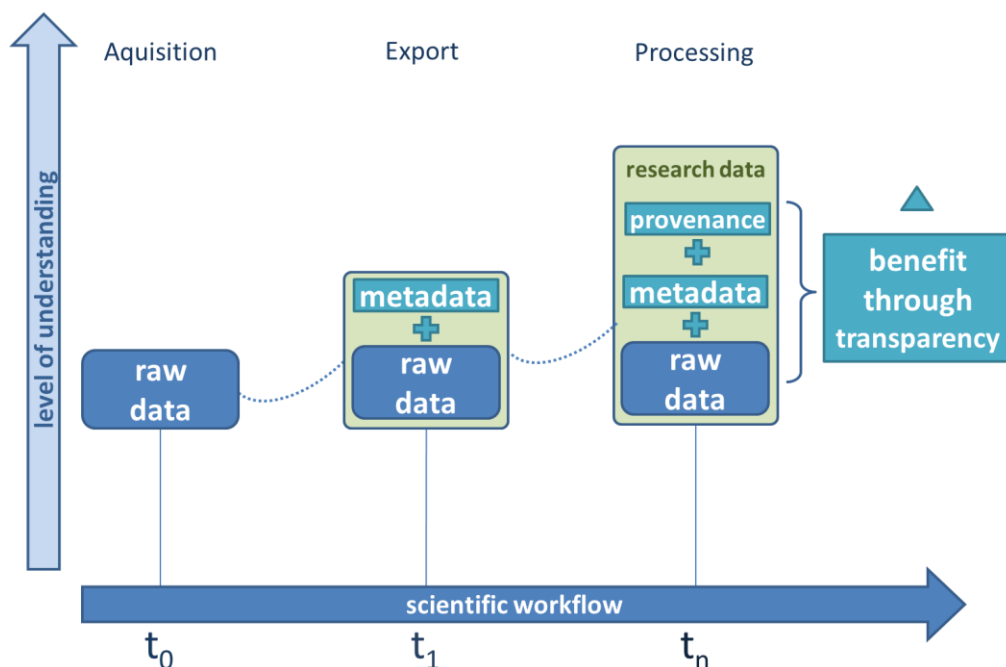
Im nachfolgenden Dokument sollen die Auswerteprozesse, die im Anschluss an die Datenakquise durchgeführt werden, näher beleuchtet werden. Gerade hierbei werden durch unterschiedliche Software-Werkzeuge neue Metainformationen und weiterführende Formate sowie vor allem Ergebnisdaten, die zu einer wissenschaftlich Erkenntnis führen sollen, erzeugt. Diese Prozesse verändern nicht nur die Daten, sondern auch den Kontext, in dem die neuen Ergebnisdaten stehen. Die Erfassung von Weitergabe- und Umwandlungsprozessen bezogen auf Forschungsdaten inklusive ihrer Parameter werden nachfolgend als Metadaten der Provenienz bezeichnet. Im Kontext der wiederauffindbaren und weiterverwertbaren Langzeitarchivierung gilt es, die Provenienz von Ergebnisdaten zu erfassen. Hierbei ist es zunächst wichtig festzulegen, welche Informationen und Parameter zum Verständnis und zur Nachvollziehbarkeit eines Forschungsergebnisses zu erfassen sind. Die gesammelten Informationen müssen dann in einem für das Forschungsfeld einheitlichen Metadatenmodell angeordnet werden. Die das Experiment beschreibenden Metadaten und die Provenienz-Metadaten bilden zusammen mit den Primärdaten und den Ergebnisdaten eine Einheit und unterlegen ein wissenschaftliches Ergebnis bzw. eine Publikation.

Dabei charakterisieren die Primärdaten, wie in Deliverable 1.1 beschrieben den Ausgangspunkt einer wissenschaftlichen Auswertung. Die Methoden und Parameter der Analysen der Primärdaten beschreiben den Pfad einer wissenschaftlichen Auswertung. Die Auswahl dedizierter Ergebnisdaten bilden das Ziel einer wissenschaftlichen Auswertung. Die Einheit aller drei Prozesse können als wissenschaftliche Erkenntnis betrachtet werden.

## 2 Material und Methoden

### 2.1 Provenienz - von den Primärdaten zur Veröffentlichung

Der Begriff der Provenienz wird im Kurationswesen als Herkunft eines Archivierungsobjektes verstanden. Das Prinzip der Provenienz beschreibt ein Ordnungsprinzip, nachdem in Archiven Objekte nach Herkunft und Entstehungszusammenhang erschlossen werden. Die Provenienz digitaler Daten umschreibt somit zum einen die Quelle eines Datums, sowie die Umstände unter welchen es erhoben wurde. Zum anderen gehört zur Provenienz digitaler Daten eine chronologische Auflistung aller weiterverarbeitenden Prozesse. Somit wird zu einer Datenmenge genau aufgeschlüsselt wie und woher die Primärdaten stammen und was mit den Primärdaten geschehen ist, um die zu betrachtende vorliegende Datenlage hervorzubringen. Für die Archivierung von Forschungsdaten aus der biomedizinischen Bildgebung bedeutet dies, neben der Akquiseprotokollierung ebenfalls alle Schritte und Parameter der digitalen Modifikation und Analyse der Daten zu erfassen, so dass eine eindeutige Beziehung bzw. Entstehungsgeschichte zwischen Primärdaten und Ergebnisdaten hergestellt und somit die Qualität des Forschungsprozesses nachgewiesen und gesteigert werden kann. [MacKenzie-Graham 2008]

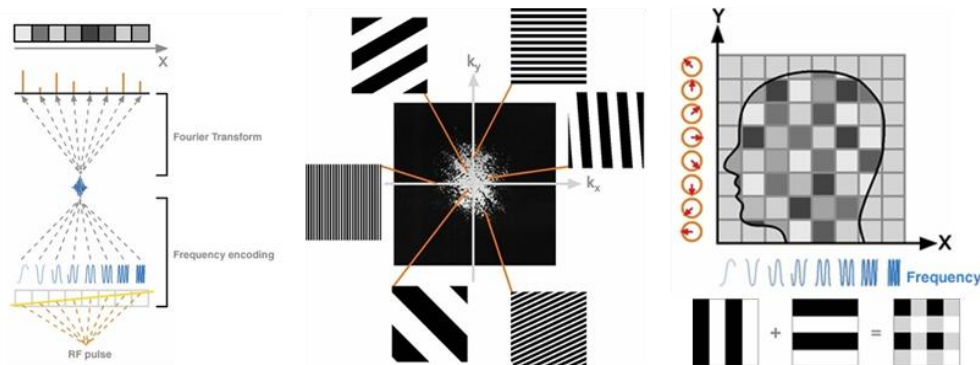


**Abb. 1:** Die Dokumentation von Provenienz steigert die Ebene des Verständnisses am Ende eines wissenschaftlichen Arbeitsablaufs.

Abb. 1 zeigt in Anlehnung an [Dickmann 2011] das Konzept der Steigerung der Transparenz von langfristig gehaltenen Forschungsdaten durch das Hinzufügen von Metadaten und Provenienzinformatoren. Erst die Kombination dieser Daten zusammen mit den Primärdaten können ein Forschungsergebnis beschreiben. Das Erfassen der Provenienz in einem Metadatenmodell dient der Erschließung von Wissenslücken zwischen archivierten Primärdaten und veröffentlichten Ergebnissen, die auf diese Daten basieren und deren Entstehungsweg meist verbal und nicht immer formal vollständig und parametrisiert beschrieben ist.

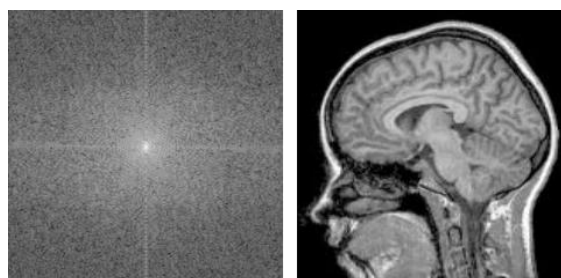
### 2.1.1 Rohdaten und Primärdaten

Der Begriff Rohdaten wird in unterschiedlichen Wissenschaften verschiedenartig definiert und verwendet. In der medizinischen Bildgebung bezeichnen die Rohdaten (raw data), diejenigen digitalen Informationen, die nach einem Scanvorgang, im Steuerrechner des MRT vorliegen. Im Scanner wird zusätzlich zum sehr starken permanenten Magnetfeld durch zusätzliche Gradientenfelder eine Ortskodierung durch kontrollierte Dephasierung der Signale ermöglicht. Durch einen Hochfrequenzpuls werden Veränderungen in den Spins der Wasserstoffkerne des Messobjektes angeregt. Das daraus resultierende Spin-Echo erzeugt in Kombination mit dem Gradientenfeld Signale unterschiedlicher Frequenz für jedes Voxel. Diese Signale werden in Schichten entlang der z-Achse im Frequenzraum (K-Raum) abgebildet. Für die Schichtselektion ist ein weiterer Gradient, der zusammen mit dem HF-Puls geschaltet wird, verantwortlich. Dies erfolgt für jede Schicht an jedem Kanal der Empfangsspule. Die Frequenzdaten verschiedener Kanäle werden zu einem Gesamtsignal rekonstruiert. Die Dekodierung der Daten aus dem Frequenzraum in den Bildraum erfolgt durch Fourier-Transformation. Hierbei werden die unterschiedlichen Frequenzen für jedes Voxel zu einem Grauwert kombiniert, wie Abb. 2 zeigt.<sup>2</sup>



**Abb. 2:** Durch Fourier-Transformation werden MRT-Daten vom Frequenzraum in den Bildraum überführt.

Erst die Grauwertdarstellung im Bildraum ist für den Menschen verständlich, kann interpretiert und zur Analyse weiterverarbeitet werden. Die Rekonstruktion und Umwandlung der Signale vom Frequenz- in den Bildraum, wie Abb. 3 zeigt, erfolgt bereits im Scanner, so dass am Ende eines Scans dem Forscher die Daten in Form von Grauwertbildern vorliegen. Auf den Umwandlungsprozess hat der Forscher in der Regel keinen Einfluss. Mit dem Export der Grauwertbilder vom Scanner beginnt der Vorgang der wissenschaftlichen Weiterverarbeitung der Daten. Deshalb werden im Kontext des Forschungsdatenmanagements und der Langzeitarchivierung die vom Steuerrechner eines MRT rekonstruierten Grauwertbilder als Primärdaten der Forschungsprozesskette bezeichnet.



**Abb. 3:** Im Vergleich zu der Frequenzraumdarstellung (links) ist die Grauwertdarstellung (rechts) für den Menschen verständlich.

<sup>2</sup> <http://www.medical.siemens.com>

### 2.1.2 Aufbereitete Daten

Nachdem die Primärdaten in Form von 16-Bit-Grauwertbildern (65536 mögliche Grauwertabstufungen) im DICOM und/oder Analyze-Format durch den Forscher vom MRT exportiert bzw. übertragen wurden, wird eine Reihe bildoptimierender Maßnahmen ausgeführt. Dabei werden die Bilder nicht nur von Artefakten und Bildrauschen befreit, sondern auch bewegungskorrigiert, geglättet, miteinander coregistriert, segmentiert und gegebenenfalls für vergleichende Analysen normalisiert. In machen Fällen z.B. bei Gruppenanalysen werden die 3-dimensionalen Hirnbilder aufgebläht, so dass die Schluchten der Hirnwindungen verschwinden und die gesamte Fläche des Kortex auf einer glatten Oberfläche dargestellt wird.

### 2.1.3 Ausgewertete Daten

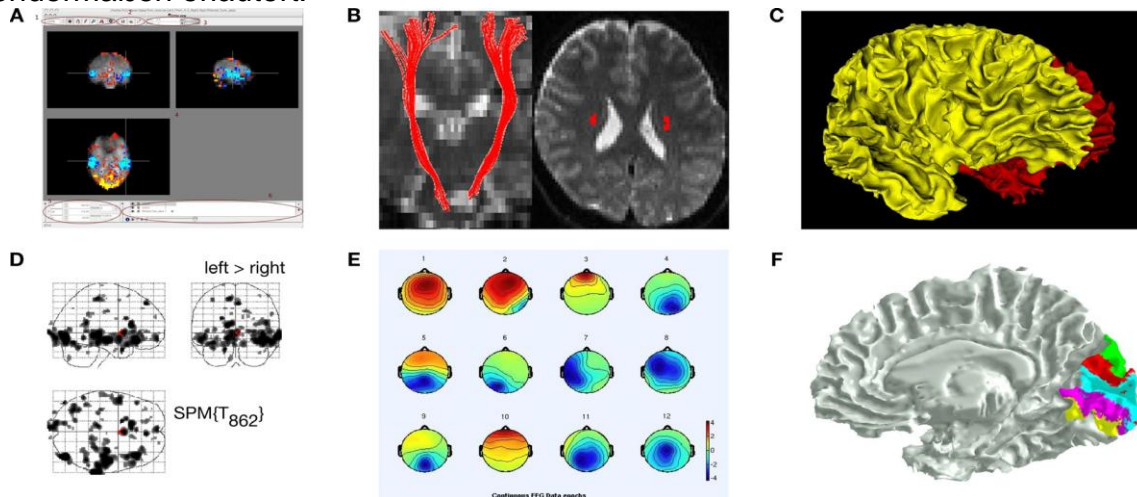
Wenn die Daten in einer bereinigten und optimierten Qualität vorliegen, können Sie durch Verarbeitung mit verschiedenen Programmen ausgewertet werden. Für die funktionelle Magnetresonanztomographie (fMRT) wird durch statistische Schätzungen zur Mustererkennung eine Klassifikation der Daten vorgenommen und somit ermittelt wie hoch die Wahrscheinlichkeit ist, dass in einer Gruppe von Voxeln (ROI) eine Aktivierung ausgelöst durch ein Ereignis (Stimulus/Reaktion) stattgefunden hat. Für Diffusionstensorimaging (DTI) wird durch statistische Verfahren aus mehreren Richtungsaufnahmen ein Ellipsoid ermittelt, der die Wahrscheinlichkeit für eine Diffusion entlang einer Achse im Koordinatensystem beschreibt. Die statistischen Methoden lassen sich an Hand der ihnen zugrundeliegenden statistischen Modelle unterteilen. Die Klassifikation erfolgt auf Zeitserien durch Methoden wie General Linear Modeling (GLM), Independent Component Analysis (ICA), Principal Component Analysis (PCA), Support Vector Machines (SVM), Relevance Vector Machines (RVM) oder Bayes'sche Statistik. Diffusionstensoren können mit Hilfe von Markovketten berechnet werden. Traktographien werden mit Hilfe von Optimierungsstrategien aus der Graphentheorie ermittelt. Das Ergebnis einer Klassifikation ist eine Wahrscheinlichkeitsverteilung in Form von Matrizen entsprechend der Schichten eines fMRT-Datensatzes. Hierbei werden in den Matrizen die Signifikanzwerte der Tests ( $p$ -Werte) dargestellt und können mit den Schnittbildern als farbliche Hervorhebungen überlagert werden. Somit lassen sich Aktivierungen im Schnittbild darstellen. DTI-Auswertungen liefern als Ergebnis die wahrscheinlichste räumliche Auflösung der Diffusionsbewegung von Wassermolekülen entlang der Hirnnervenfasern in Form der berechneten Tensoren. Diese können farbkodiert nach Richtung als Schnittbildreihe dargestellt werden. Die Traktographie liefert Pfade entlang der Diffusionshauptrichtungen in einem Koordinatensystem, welche als 3D-Modell dargestellt werden können.

### 2.1.4 Veröffentlichte Ergebnisse

Für die Veröffentlichung der Forschungsergebnisse kann eine visuelle Aufbereitung stattfinden. Bei anatomischen Datensätzen kann ein Volumerendering bestimmte Hirnstrukturen im 3D-Modell darstellen. Bei fMRT-Resultaten werden meist nur die  $p$ -Werte der Klassifikation tabellarisch ausgewertet und zusammen mit den ROIs der anatomischen Daten eingebracht. Ermittelte Diffusionstensoren werden in farbkodierten Schichtbildern oder Traktographie-Renderings visualisiert. Hierbei soll vor allem die Güte der Auflösung dargestellt werden. Allen Akquisemethoden gemein ist die Auflistung der Messparameter laut Scannerprotokoll zur vollständigen Referenzierung, wie die Daten technisch erhoben wurden.

## 2.2 Auswertewerkzeuge für fMRT und DTI

Abb. 4 zeigt eine Auswahl an Darstellungen der Daten der Neurobildgebung, die in verschiedenen Anwendungen genutzt werden. In [Gouws 2009] werden diese folgendermaßen erläutert.



**Abb. 4:** Verschiedene Software-Werkzeuge erzeugen unterschiedliche Analyseergebnisse aus den Primärdaten

(A) FSLView aus der Oxford FRMIB Toolbox zeigt die Überlagerung von 3D-fMRT-Daten mit anatomischen Bilddaten in drei orthogonalen Ebenen. (B) DTIStudio zeigt Nervenfaserverläufe als farbige Stromlinien die mit orthogonalen Schichtbildern eines anatomischen MRT-Volumens überlagert werden. (C) FreeSurfer wird genutzt, um Oberflächen, die aus MR-Daten extrahiert wurden, darzustellen. Hier im Beispiel wird die Grenzfläche zwischen grauer und weißer Substanz in dreidimensional dargestellt, wobei die linke und rechte Hirn-Hemisphäre unterschiedlich eingefärbt wurde. (D) SPM kann genutzt werden um Projektionen von statistisch signifikanten Regionen auf ein durchsichtiges Hirnvolumen durchzuführen. (E) EEGLab ist in der Lage, zweidimensionale Muster aus dem sich verändernden elektrischen Feld auf der Kopfoberfläche zu generieren. (F) Hier wird mrVista genutzt, um skalare Werte auf eine kortikale Oberfläche abzubilden und so verschiedene visuelle Areale verschiedenfarbig anzuzeigen.

Die nachfolgend vorgestellten Auswertewerkzeuge gelten als die wichtigsten und meist genutzten Softwareprodukte der neurowissenschaftlichen Gemeinschaft. Alle hier gelisteten Programme wurden in der Jubiläumsausgabe von Neuroimage<sup>3</sup> zum Thema „20 YEARS OF fMRI“ in Form von Review-Artikeln vorgestellt. Neben dieser Sammlung gibt es allerdings zahlreiche weitere Analyseprogramme für neurowissenschaftliche Bilddaten. Einige werden ab Abschnitt 2.2.7 vorgestellt. In Magdeburg kommen die drei Programme Brainvoyager, SPM und FSL am häufigsten zum Einsatz, daher werden diese Werkzeuge eingehender dargestellt. Eine komplette Übersicht inklusive Download der über 200 freiverfügbaren neurowissenschaftlichen Programme wird durch NITRC<sup>4</sup> (NeuroImaging Tools and ResSources) bereitgestellt.

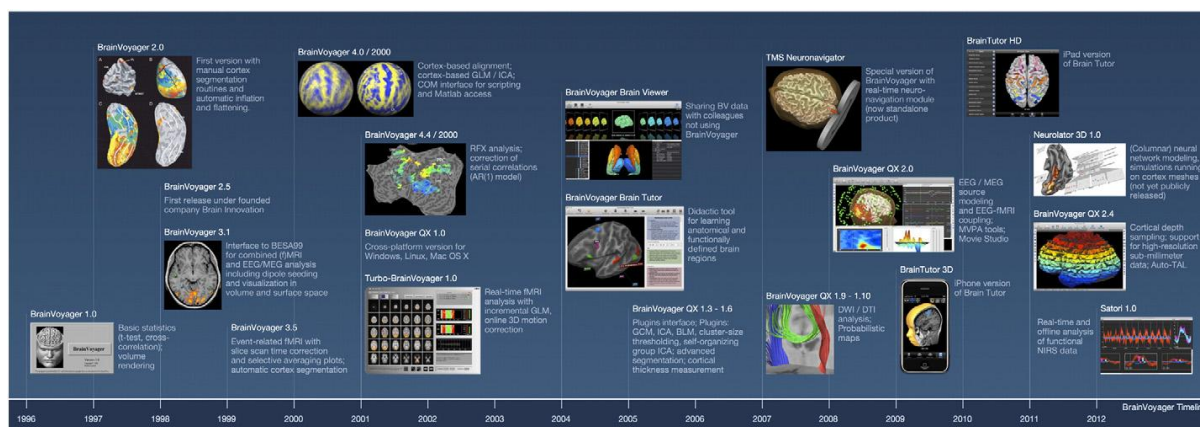
<sup>3</sup> <http://www.sciencedirect.com/science/journal/10538119/62/2>

<sup>4</sup> <http://www.nitrc.org>



## 2.2.1 BrainVoyager

Die kommerzielle Software BrainVoyager<sup>5</sup> gliedert sich in zwei für die Auswertung von MRT-Experimenten spezialisierte Produkte. Der BrainVoyager QX beinhaltet verschiedene Visualisierungs- und Analyse-Funktionalitäten um strukturelle, funktionelle und Diffusions-gewichtete MR-Daten sowie EEG- und MEG-Daten auswerten zu können. Turbo-BrainVoyager wurde speziell dafür entwickelt, fMRT-Experimente in real-time auszuwerten und dadurch Neurofeedbackanwendungen zu ermöglichen. Die Software wird durch die Firma Brain Innovation vertrieben. Die Produktfamilie wurde vom passionierten Programmierer Rainer Goebel seit 1995 entwickelt. Der Psychologe und Informatiker ist Experte auf den Gebieten Neurobiologie, Neuronale Netzwerke, Hirn-Computer-Schnittstellen und Neurofeedback und hält als Professor den Lehrstuhl für Kognitive Neurowissenschaften an der Universität von Maastricht. Er ist ebenfalls Direktor des Maastricht Brain Imaging Center und Leiter der Abteilung Neuroimaging & Neuromodelling im Institute for Neuroscience der Niederlande. Als CEO von Brain Innovations und Chefentwickler von BrainVoyager setzt er durch die Gestaltung der Software-Produktfamilie maßgebliche Impulse in der Gemeinschaft der Neurowissenschaftler.



**Abb. 5:** Die Versionsgeschichte von BrainVoyager stellt zahlreiche Funktionalitäten der Software dar. Seit Version QX ist BrainVoyager eine Multiplattform-Software, die unter Windows, Mac OS und Linux betrieben werden kann. Die Software bietet so [Goebel 2012] folgende Funktionalitäten und Methoden (s. Abb. 5): 2D/3D Analyse und Visualisierung; Volumen- und Kortex-basierte Hypothesen-getriebene statistische Auswertungen durch General-Linear-Modelling; Analyse von Zufallsereignissen durch multifaktorielle Designs und Korrelationen mit externen Variablen; statistische Schwellwertbildung bei multiplen Vergleichen; Volumen-of-Interest- und Path-of-Interest-Analysen in Mehrfach-Probandendaten; Independent Component Analysis; automatische Hirnsegmentierung, Oberflächenrekonstruktion, Kortexaufblähung; kortexbasierte Ausrichtung mehrerer Probandenvolumina; DTI-Analysen in Kombination mit struktureller und funktioneller MRT; Erstellung hochwertiger Volumen- und Oberflächen-Renderings als Bilder oder Filme. Der TurboBrainVoyager liefert zusätzliche Funktionalitäten: Datenexploration während der laufenden Messung; inkrementelle statistische Datenanalyse und Ereignis-gesteuerte Mittlung; automatische Erzeugung von Designmatrizen; Bewegungskorrektur und Gaus'sche Glättung in Echtzeit; online Veränderung statistischer Parameter während der Echtzeitanalyse; Speicherung funktioneller Daten für tiefere Analysen mit BrainVoyager QX. Die Soft-

<sup>5</sup> <http://www.brainvoyager.com>

wareprodukte können unter Verwendung von USB-Dongles lokal oder via Netzwerk freigeschaltet werden. BrainVoyager verarbeitet MRT-Daten, die im DICOM-Format vorliegen. Die Daten, die im Verlauf einer Auswertung erzeugt werden, werden in BrainVoyager-eigenen proprietären Formaten abgespeichert. Renderings lassen sich in Standard-Medien-Formate exportieren.

### 2.2.2 Statistical Parametric Mapping

**Statistical Parametric Mapping**<sup>6</sup> (SPM) ist eine Sammlung von GUI-basierten Modulen zur Schätzung räumlicher statistischer Prozesse und wird genutzt um Hypothesen auf Hirnbild-Datenreihen zu testen. Dies können sowohl zeitlich korrelierte 4D-Datensätze als auch Kohortendatensätze sein. Die Toolbox kommt nicht in Form einer eigenständigen Software daher, sondern bildet eine Erweiterung der Mathworks MATLAB<sup>7</sup>-Umgebung in Form einer Toolbox und wird durch das Trust Center for Neuroimaging des University College London zur Verfügung gestellt und weiterentwickelt. Durch die Toolbox werden innerhalb der MATLAB-Umgebung zusätzliche Eingabeoberflächen, Funktionsaufrufe und Ergebnisanzeigen bereitgestellt. Die erste primitive Version namens SPM91 bzw. SPMclassic wurde laut [Ashburner 2012] bereits 1989 für die Auswertung von PET-Bildern entwickelt und verbreitete sich schnell innerhalb der Forschungsgemeinschaft. Aktuell liegt SPM in Version 8 vor und unterstützt die statistische Auswertung von fMRI-, PET-, SPECT-, EEG- und MEG-Daten mit folgendem Funktionsumfang: Registrierung rigider Körper und nicht-lineare Registrierung; Segmentierung; Normalisierung; Weichzeichnung; Voxel-basierte Morphometrie; Generalisierte Lineare Modelle; Hierarchische Modelle für EEG und MEG; Zufallseffektanalyse; Varianzanalyse; Konvolutionsmodelle für fMRT; Methoden der klassischen Ableitung wie Random Field Theory oder False Discovery Rate; Bayes'sche Ableitungen und vieles mehr. Der komplette Funktionsumfang wird in [Friston 2007] beschrieben. Die Toolbox ist in der Lage MR-Bilddaten im DICOM und Analyse-Format zu verarbeiten. Im Rahmen der MR-basierten Neurowissenschaft wird SPM für die Vorverarbeitung und statistische Auswertung sowie Visualisierung der Ergebnisse in der fMRT eingesetzt. Da SPM auf grundlegende Funktionen und Methodenaufrufe von MATLAB zurückgreift, ist bei der Dokumentation der verwendeten Version ebenfalls die Angabe der Version der MATLAB-Umgebung von Bedeutung.

### 2.2.3 FMRIB's Software Library

FSL ist eine weitverbreitete Softwaresammlung zur Auswertung von (f)MRT und DTI-Daten. FSL steht für **FMRIB's Software Library**<sup>8</sup>, wobei FMRIB das Kürzel des Oxford Centre for Functional MRI of the Brain repräsentiert, dessen Mitarbeiter die Programme und Algorithmen entwickelt haben. Die Sammlung ist quell-offen und frei erhältlich zur nicht-kommerzielle Anwendung. FSL ist unabhängig von anderen Anwendungen wie MATLAB und ist für die Betriebssysteme Linux und Mac OS X erhältlich. Die Sammlung besteht aus mehreren Werkzeugen, die teilweise über grafische Oberflächen verfügen hinter denen sich Shell-Skripte verbergen, die die Daten schließlich an die in C programmierten Binaries übergeben. FSL-Auswertungen können per GUI oder via Kommandozeile gestartet werden. Die Möglichkeit der Kommandozeilenausführung macht FSL nutzbar für Anwendung auf Rechenclustern und

---

<sup>6</sup> <http://www.fil.ion.ucl.ac.uk/spm>

<sup>7</sup> <http://www.mathworks.de/products/matlab>

<sup>8</sup> <http://www.fmrib.ox.ac.uk/fsl>

Serverfarmen. Von Haus aus, lassen sich einige FSL-Programme mit den Cluster-Schedulern Oracle Grid Engine und Condor parallelisieren. Die FSL-Algorithmen implementieren alle klassischen Vorverarbeitungsverfahren und die Anwendung Bayes'scher Statistik zur Auswertung von Neurobilddaten. [Crum 2011]

Folgende Programme und Funktionalitäten werden in der Sammlung für strukturelle MRT-Daten bereitgestellt: *BET (Brain Extraction Tool)* segmentiert MR-Bilder des ganzen Kopfes, so dass nur noch das Hirngewebe dargestellt wird; *FAST (FMRIB's Automated Segmentation Tool)* segmentiert Hirnbilder, so dass verschiedene Gewebetypen verschiedenfarbig dargestellt werden mit Hilfe des Hidden Markov Random Field Model; *FIRST* ist ein Programm zur Registrierung und Normalisierung von Hirnbilddaten auf Grundlage von Vergleichsdatensätzen; *FLIRT (FMRIB's Linear Image Registration Tool)* ist ein Programm zur automatischen Registrierung und Bewegungskorrektur; *FSL-VBM* dient der Voxel-basierten Morphometrie; *SIENA* dient der Ermittlung von Volumenänderungen zwischen zwei 3D-Aufnahmen; *fs\_l\_anat* ist ein Pipeline-Skript, das die zuvor genannten Programme hintereinander auf einen Datensatz anwendet, um die Vorverarbeitung zu vereinfachen.

Folgende Programme und Funktionalitäten werden in der Sammlung für funktionelle MRT-Daten bereitgestellt: *FEAT* stellt einen GLM-basierten Ansatz für die Auswertung von Zeitserien bei maximaler Vorgabe von Standard-Parametern zur Verfügung; *MELODIC* verfolgt hingegen den Ansatz der ICA, um Aktivierungen über die Zerlegung verschiedener 4D-Datensätze in räumliche und zeitliche Komponenten zu ermitteln; *FABBER* implementiert Nonlinear Model Fitting auf Perfusionsdaten; *BASIL* implementiert Bayes'sche Verfahren für Arterial Spin Labeling auf Perfusionsdaten.

Folgende Programme und Funktionalitäten werden in der Sammlung für Diffusions-Daten bereitgestellt: *FDT (FMRIB's Diffusion Toolbox)* ist eine Software für die Analyse Diffusions-gewichteter MR-Daten und enthält Werkzeuge zur Datenvorverarbeitung/-bereinigung, Ermittlung von Diffusionstensoren und zur Traktographie sowie Konnektivitäts-basierten Segmentierung; *TBSS (Tract-Based Spatial Statistics)* ermöglicht Voxel-weise Analyse von Diffusionsdaten aus Studien mit mehreren Probanden; *EDDY* und *TOPUP* sind Vorverarbeitungsprogramme speziell für Wirbelstrom-, Bewegungs- und Feldverzerrungskorrekturen in Diffusionsdaten.

Neben weiteren speziellen Werkzeugen, auf die nicht weiter eingegangen werden soll, ist *FSLView* zu nennen. Das Programm ermöglicht es alle mit FSL erzeugten Daten und Auswertungen in eine grafische Oberfläche zu laden und zu visualisieren. So können z.B. Aktivierungen und Segmentierungen als Einfärbungen über anatomische Datensätze gelegt werden oder Diffusionstensoren durch Farbwerte und Traktographien als 3D-Objekte dargestellt werden. FSL liegt momentan in Version 5.0 vor und wird kontinuierlich weiterentwickelt. Hervorzuheben ist, dass FSL nur mit den Bildformaten NIFTI und Analyze kompatibel ist. DICOM-Bilder lassen sich mit FSL leider nicht verarbeiten [Jenkins 2012].

### 2.2.4 FreeSurfer

FreeSurfer<sup>9</sup> ist eine Sammlung von Werkzeugen zur Analyse von Daten der Neurobildgebung, wie (f)MRT-, DTI- und EEG/MEG-Daten. Es stellt eine Reihe Algorithmen zur Verfügung, mit denen funktionale, strukturelle und konnektive Eigenschaften des menschlichen Gehirns ausgewertet werden können. Die freiverfügbare und quelloffene Software ist für die Betriebssysteme CentOS- sowie Debian-Linux, Mac OS X und für Windows in einer VirtualBox (Ubuntu) erhältlich. Unterstützte Formate sind DICOM, Analyze und MINC (Medical Imaging NetCDF, proprietäres Format des McConnell Brain Imaging Centre am Montreal Neurological Institute, Kanada) [Fischl 2012].

### 2.2.5 Caret

Das freie und quelloffene Softwarepaket Caret<sup>10</sup> des VanEssen Lab (Department of Anatomy & Neurobiology, Washington University in St. Louis, Missouri, USA) unterstützt die Visualisierung und Analyse der Kortexoberfläche struktureller und funktionaler MRT-Daten. Die Software, die seit den 90er Jahren kontinuierlich weiterentwickelt wurde, besteht aus einer grafischen Oberfläche (Caret5) zum Anzeigen der Datensätze und einem Kommandozeilenprogramm (caret\_command) mit dessen Hilfe Pipelines mit 200 unterschiedlichen Operationen zur Datenauswertung abgearbeitet werden können. Die Software ist für Linux, Mac OS X und Windows erhältlich und unterstützt die Bildformate Analyze und NIfTI sowie Datenformate aus AFNI und Freesurfer. [Van Essen 2012] Im Rahmen des Human Connectome Project<sup>11</sup> wurde eine Datenbank, namens SumsDB (Surface Management System Database), für öffentliche und private Sammlungen an kortikalen Hirnbilddaten erstellt. Die Komponente WebCaret dient hierbei zur Visualisierung der Daten im Webbrowser und ist aus der Caret GUI hervorgegangen.

### 2.2.6 AFNI und SUMA

AFNI<sup>12</sup> (Analysis of Funktional NeuroImages) ist eine freie quelloffene Sammlung von C-Programmen um funktionale MRT-Daten zu verarbeiten, zu analysieren und anzuzeigen. Die Programme werden seit 1994 weiterentwickelt, sind unter allen Unix-ähnlichen Betriebssystemen mit X11- und Motiv-Unterstützung lauffähig, und somit auch unter Linux und Mac OS X [Cox 1996]. SUMA<sup>13</sup> ist ein Programm, welches die Funktionalität von AFNI um Analysen der kortikalen Oberfläche von fMRT-Daten erweitert [Saad 2012]. Da AFNI und SUMA vom National Institute of Health entwickelt wurden, unterstützen sie nur das Format NIfTI, welches ebenfalls durch das NIH entworfen und bereitgestellt wurde (siehe Deliverable 1.1).

### 2.2.7 3DSlicer

3DSlicer<sup>14</sup> ist ein freies quelloffenes Programm, welches sich in Bedienung und Funktionalität eng an PACS-Workstations anlehnt und klinische Protokolle unterstützt. Daher wird es oft in klinischen Studien eingesetzt [Fedorov 2012]. Mit der Software lassen sich neben MRT auch Modalitäten wie CT, Ultraschall und nuklear-

---

<sup>9</sup> <http://surfer.nmr.mgh.harvard.edu>

<sup>10</sup> <http://brainvis.wustl.edu/wiki/index.php/Caret:About>

<sup>11</sup> <http://www.humanconnectome.org>

<sup>12</sup> <http://afni.nimh.nih.gov>

<sup>13</sup> <http://afni.nimh.nih.gov/afni/suma>

<sup>14</sup> <http://www.slicer.org>

medizinische Daten sowie Mikroskopierdaten direkt aus dem PACS oder anderen DICOM-Bilddatenbanken laden, darstellen und auswerten. Außerdem ist die Software in der Lage medizinische Bilder vom Kopf bis zum Fuß eines Patienten auszuwerten und ist somit nicht auf die neurowissenschaftliche Forschung beschränkt. Das Programm ist für Linux, Mac OS X und Windows erhältlich und verarbeitet das Datenformat DICOM.

### 2.2.8 MedINRIA

MedINRIA<sup>15</sup> ist ein freiverfügbares Programm zur Auswertung von fMRT und vor allem DTI-Daten. Es ist aus verschiedenen Projekten des französischen Forschungsinstitutes Inria hervorgegangen [Toussaint 2007]. Es ist für die Plattformen Mac OS X, Debian- und Fedora-Linux sowie Windows erhältlich. Ähnlich 3DSlicer setzt MedINRIA auf die Funktionalität von PAC-System auf. So lassen sich die DICOM-Bilder aus einem PACS in eine lokale Datenbank laden um diese anschließend zu bearbeiten und auszuwerten. Dabei gehören neben DTI- und fMRI-Analysen ebenfalls Registrierung, Segmentierung, Filterung, Traktographie, Volumerendering und 3D/4D-Darstellung zum Funktionsumfang.

### 2.2.9 NeuroELF

NeuroELF<sup>16</sup> ist ebenfalls eine Toolbox, die auf MATLAB basiert und für das Arbeiten mit Daten aus der Neurobildgebung entwickelt wurde. Auf der Webseite wird darauf hingewiesen, dass diese Toolbox nicht als Ersatz sondern vielmehr als Ergänzung zu den traditionellen Auswerte-Werkzeugen, die in diesem Kapitel vorgestellt wurden, anzusehen ist. Die Toolbox bietet eine grafische Oberfläche und verschiedene Methoden, um die Formate der Programme Brainvoyager QX, SPM, AFNI, FSL und Freesurfer weiterzuverarbeiten, in den Schichtbildern zu navigieren und Konfigurationen und Einstellungen in einem allgemein lesbaren Format abzuspeichern.

### 2.2.10 MRI Studio

MRI Studio<sup>17</sup> ist ein Programm, welches speziell zur Auswertung von DTI-Daten entwickelt wurde. Es fasst einen Bildbetrachter und beherrscht Tensorkalkulation, Traktographie, 3D Visualisierungen, Registrierung und ROI-Statistik. Die Software ist ausschließlich für das Betriebssystem Windows verfügbar und ist am Functional Brain Imaging Center (Johns Hopkins University in Baltimore, Maryland, USA) entstanden. Die Software ist weder ausführlich dokumentiert noch wird sie in Publikationen erwähnt. Allerdings wird sie trotz dessen von vielen DW-MRT-Forschern ergänzt zu oben genannten Werkzeugen genutzt und wird im Rahmen des BIRN-Projektes (Biomedical Informatics Research Network) gefördert und verbreitet.

### 2.2.11 NeuroDebian

Neben Slackware, aus dem SuSE-Linux hervorgegangen ist, und Red Hat, aus dem CentOS und Scientific-Linux entstanden sind, ist Debian eines der drei ältesten bis heute weiterentwickelten Linux-Distributionen. Da Debian ausschließlich freie Software enthält, die zumeist aus dem GNU-Projekt stammt, und weltweit die am meisten genutzte Linux-Distribution ist, entschlossen sich Halchenko und Hanke [Halchenko 2012] daraus eine Linux-Distribution zusammenzustellen, die bereits alle

---

<sup>15</sup> <http://med.inria.fr>

<sup>16</sup> <http://neuroelf.net>

<sup>17</sup> <https://www.mristudio.org>

existierenden freiverfügbaren Werkzeuge für die Neurobildgebung für Debian kompiliert durch ein entsprechendes Paket-Repository enthält und zugleich binärkompatibel mit Debian und seinen Derivaten wie Ubuntu bleibt. Diese Distribution nannten sie NeuroDebian<sup>18</sup>. Sie wird unter anderem als VirtualBox bereitgestellt und kann somit schnell heruntergeladen und in Betrieb genommen werden. Dies ist die einfachste Methode die oben genannten freien Linux-Programme für einen Wissenschaftler zur Verfügung zu stellen und markiert gleichzeitig einen Meilenstein in der neurowissenschaftlichen Zusammenarbeit und Communitybildung. [Poline 2012] Die vollständige Auflistung der aktuell 322 Programme/Pakete in NeuroDebian kann unter <http://neuro.debian.net/pkg.html> eingesehen werden.

### 2.2.12 Fazit

Aus zahlreichen neurowissenschaftliche Großprojekten, Zentren, Kooperationen und persönlichem Engagement einzelner Forscher ist eine Vielzahl an Software, Programmen, Paketen und Toolboxen hervorgegangen. Dabei sind die Auswertewerkzeuge nicht von einer zentralen Instanz verifiziert und validiert wurden. Es ist also nicht immer ganz klar, ob Analyseergebnisse eines Programmes korrekt sind und im Kontext der jeweiligen wissenschaftlichen Fragestellung die Realität abbilden. So konnte z.B. in [Eklund 2012] gezeigt werden, dass bei bis zu 70% von 1484 fMRT-Datensätze aus Resting-State-Experimenten, also Messungen ohne Stimulus, Aufgabenstellung oder Paradigma, die mit SPM ausgewertet wurden, ein statistischer Fehler größer 5% zu verzeichnen war. D.h. Datensätze in den keine Paradigmen-gesteuerte Aktivierung, sondern zufällige Aktivierungen im Ruhezustand des Probanden, berechnet hätte werden dürfen, zeigten nach der Auswertung mit SPM Aktivierungen gemäß verschiedener bekannter Paradigmen. Es hat auch schon Nachweise gegeben, dass bei Auswertung eines fMRT-Datensatzes eines tiefgefrorenen Atlantik-Lachs, einige Programme trotzdem eine Aktivierung von Hirnarealen berechnen und anzeigen [Bennett 2009].

Somit ist es unerlässlich bei der Publikation von Forschungserkenntnissen anzugeben, welches Programm inklusive Parameter, Versionsnummer und Betriebssystem für die Auswertungen herangezogen wurde. Es zeigt sich auch, dass die unterschiedlichen Programme verschiedene Kombinationen von Formaten unterstützen. Es ist nicht immer gegeben, dass der internationale Standard DICOM genutzt werden kann. Oftmals sind Neurowissenschaftler dazu angehalten für bestimmte Auswertungen ihre Daten in NIfTI oder Analyze zu konvertieren und basierend auf diesen Formaten weiter zu arbeiten. Außerdem werden die Ergebnisse der Auswertungen von der Vielzahl an Programmen in einer Vielzahl an Ergebnisdaten-Formaten gespeichert. Dadurch zeigt sich, dass ein Forschungsdatenmanagementsystem hinsichtlich der Provenienzinformationen in Form von Metadaten und hinsichtlich der Formate eingepflegter Daten höchst flexibel sein muss. Aufgrund der Heterogenität der Vorgehensweisen und Ergebnisse muss ein Metadatenformat für die neurowissenschaftliche Bildgebung prinzipiell generisch, offen und erweiterbar sein.

---

<sup>18</sup> <http://neuro.debian.net>

## 3 Ergebnisse

### 3.1 Metadatenmodelle zur Verwaltung von MR-Projekten

Nachfolgend sollen einige Metadatenmodelle, die im Rahmen der biomedizinischen Forschung entwickelt wurden, vorgestellt werden. Es handelt sich hierbei teilweise um theoretische Ansätze, die es zu implementieren gilt und teilweise um Metadatenmodelle, die im Rahmen der Entwicklung von Forschungsdatenmanagementsystemen entstanden sind und implementiert wurden. Eine vergleichende Übersicht nach [Lohrey 2009] ist in Abb. 6 zu sehen.

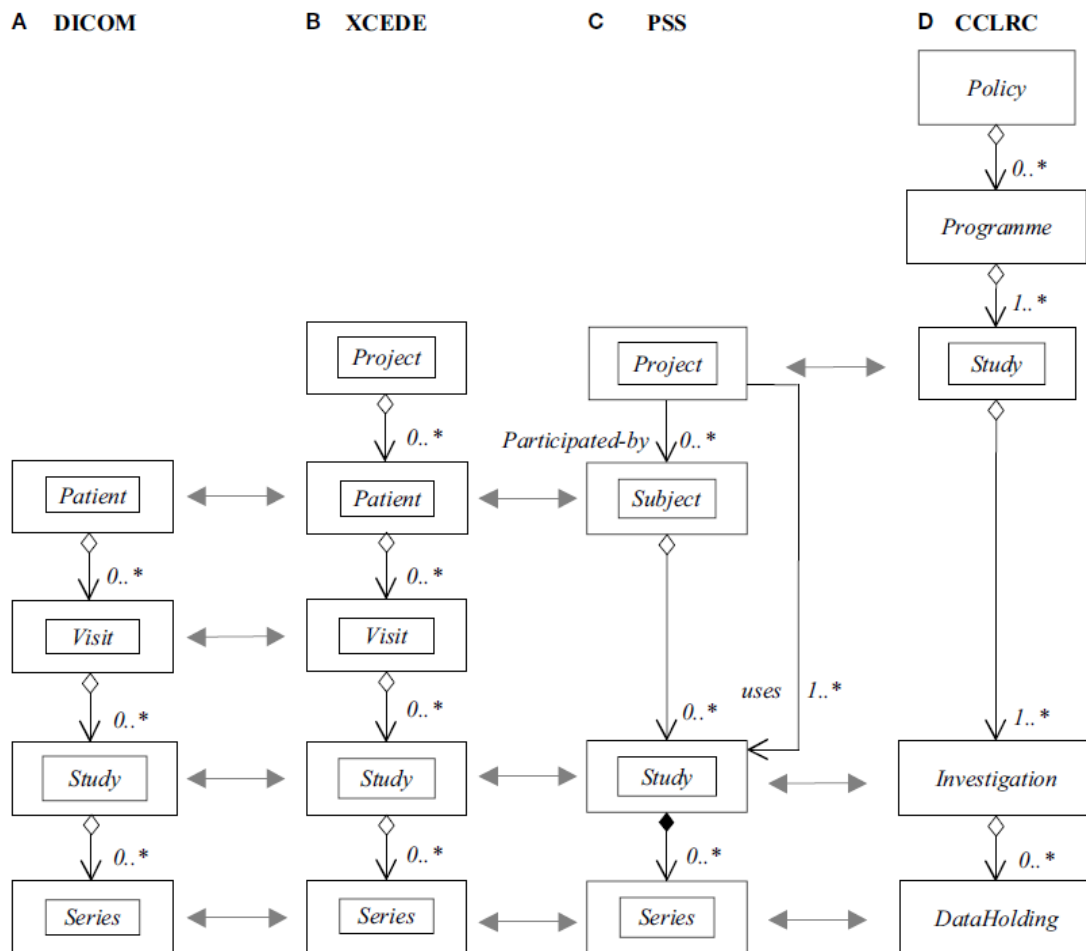


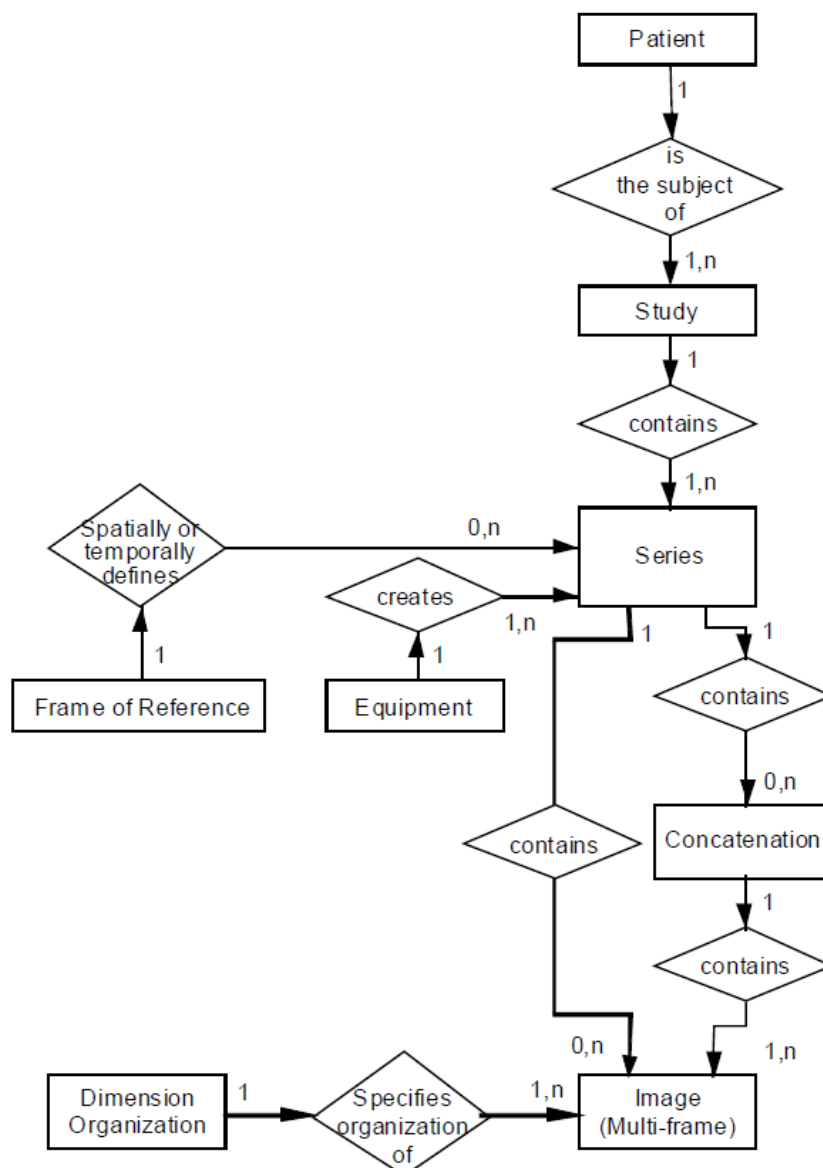
Abb. 6: Vergleich verschiedener Metadatenformate hinsichtlich ihrer Datenobjekte

#### 3.1.1 DICOM

DICOM<sup>19</sup> (Digital Imaging and Communications in Medicine) wurde als Format bereits im Deliverable 1.1 behandelt. Es handelt sich hierbei neben einem Bilddatenformat und einem Kommunikationsprotokoll ebenfalls um ein Metadatenformat, welches aus einer Hierarchie von Entitäten, die in Relation zu einander stehen, besteht [Mildenberger 2002]. Die Metainformationen werden im Header der DICOM-Dateien gehalten. So wird in diesem Metadatenmodell beschrieben das ein Patient im Rahmen einer Untersuchung (*Visit*) ein Subjekt einer Studie (*Study*) ist. Diese Studie enthält eine Serie von Bildern (*Series*). Die Serie wird durch verschiedene Gerä-

<sup>19</sup> <http://medical.nema.org/standard.html>

te/Modalitäten (Equipment) erzeugt. Ein Bild (*Image*) einer Serie besteht aus mehreren Einzelbildern (*Frames*), welche wiederum durch Dimension, Auflösung usw. charakterisiert sind, wie Abb. 7 zeigt. Somit kann in diesen Objekten der gesamte klinische Ablauf (Patienteninformationen, Informationen zur Untersuchung wie Klinik und Arzt, Informationen zur Modalität wie Hersteller und technische Daten usw.) der Bilderhebung dokumentiert werden. DICOM ist international standardisiert und sehr komplex. Es ist darauf ausgelegt verschiedene bildgebende Geräte und Bildspeicher interoperabel zu machen. Die Grenzen für die Forschung liegen darin, dass DICOM den klinischen Ablauf der Bilderhebung beschreibt. Es fehlen zum einen Objekte, die Forschungsprojekte und Experimente abbilden und zum anderen Objekte, die die Weiterverarbeitung und Auswertung der Bilddaten beschreiben. Somit ist DICOM für ein Forschungsdatenmanagement weniger geeignet, obwohl freie PACS-Lösungen, wie dcm4chee<sup>20</sup> in Kombination mit OsiriX<sup>21</sup>, durchaus zur Speicherung von Forschungsdaten genutzt werden.



**Abb. 7:** Objektverknüpfungen des Real-World-Modells aus dem DICOM-Standard PS 3.3-2011

<sup>20</sup> <http://www.dcm4chee.org>

<sup>21</sup> <http://www.osirix-viewer.com>



### 3.1.2 XCEDE 2

XCEDE 2 (**X**ML-Based **C**linical **E**xperiment **D**ata **E**xchange Schema Version 2) wurde für die Human Imaging Database (HID) im Rahmen des BIRN-Projektes an der Washington University (St. Louis, Missouri, USA) entwickelt. Es steht in Form einer XML Schema Definition (XSD) frei zur Verfügung<sup>22</sup>. Das Schema beschreibt eine weitreichende Metadatenhierarchie um Forschungsdaten und Daten aus klinischen Studien zu beschreiben und zu dokumentieren. Es erweitert die Objekthierarchie von DICOM um die Entität *Project*, wie Abb. 8 zeigt, und stellt gegenüber DICOM eine Vielzahl weiterer Attribute für die Entitäten zur Verfügung, die z.B. auch Information zur Provenienz der Daten abbilden können. Das Schema sollte laut [Gadde 2011] als Export- und Austauschformat zwischen Datenbanken, Auswertewerkzeugen und Webanwendungen genutzt werden, da es für den Menschen verständliche Inhalte organisiert in einem Maschinen-lesbaren Format abspeichert. In [Keator 2006] wurde das Schema für fMRT in SPM integriert und eine Toolbox dazu angeboten, um Auswertedaten aus SPM in XML zu exportieren. Da das Format allerdings kein internationaler Standard ist, wird es auch von keinem der in 2.3 genannten Werkzeuge unterstützt. Das Metadatenschema findet momentan Anwendung in der oben genannten HID und SPM-Toolbox.

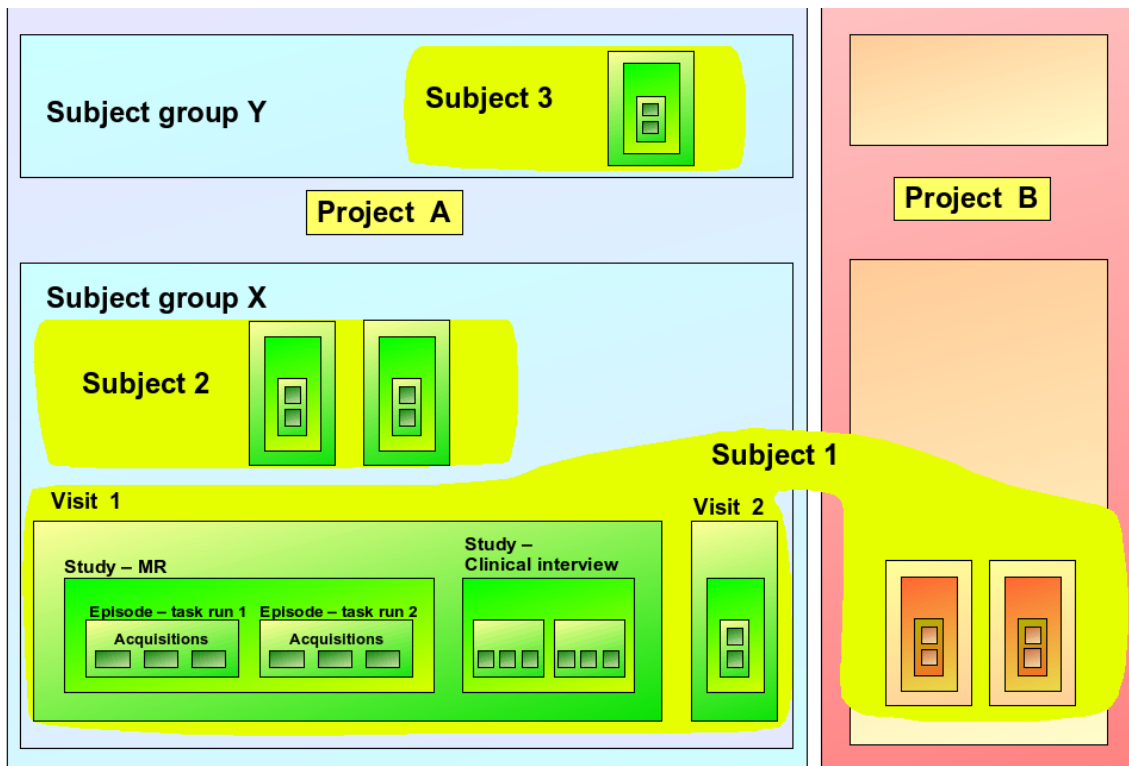


Abb. 8: Schematische Darstellung der Objekthierarchie<sup>23</sup> in XCEDE 2

### 3.1.3 CSMDM

CSMDM (**C**ore **S**cientific **M**eta**D**ata **M**odel) des CCLRC<sup>24</sup> (Council for the Central Laboratory of the Research Councils, Großbritannien) wird nicht öffentlich präsentiert und ist nur über die Veröffentlichung [Matthews 2010] und im technischen Report des CCLRC [Sufi 2004] zu erschließen. Es wurde als formaler Lösungsansatz zur Hand-

<sup>22</sup> <http://www.xcede.org/xcede-2/xcede-2.0-core.xsd>

<sup>23</sup> <http://www.birncommunity.org/tools-catalog/xcede>

<sup>24</sup> <http://www.cclrc.ac.uk>

habung großer Datenbestände in Forschungseinrichtungen entwickelt und bietet einen allgemeinen Ansatz für die Organisation von Forschungsdaten. Es ist somit nicht auf den Einsatz für bildgebende Neurowissenschaften begrenzt. Im Kern bietet CSMD folgende Entitäten: *Policy* ist das Förderprogramm für das Forschungsvorhaben; *Programme* ist die Zusammenfassung von Studien zu einem Forschungsthema; *Study* besteht aus mehreren *Investigation* als fundamentale Einheiten einer Studie; *Investigation* gliedert sich in *Measurement*, *Simulation* oder *Experiment*; *Investigator* ist das forschende Personal; *Topic* und *Keyword* beschreiben das Thema des Forschungsvorhabens; *Publication* beinhaltet alle motivierenden und entstandenen Referenzen zum Projekt; *Sample* beschreibt das Material bezüglich wessen geforscht wird; *Dataset* beschreibt eine Sammlung von Daten mit *Datafile* als Elemente in denen Daten und Ergebnisse abgelegt sind; *Parameter* beschreiben die messbaren Qualitäten und Quantitäten von *Investigation*; *Authorization* definiert die Zugriffsrechte für die Elementen von *Investigation*.

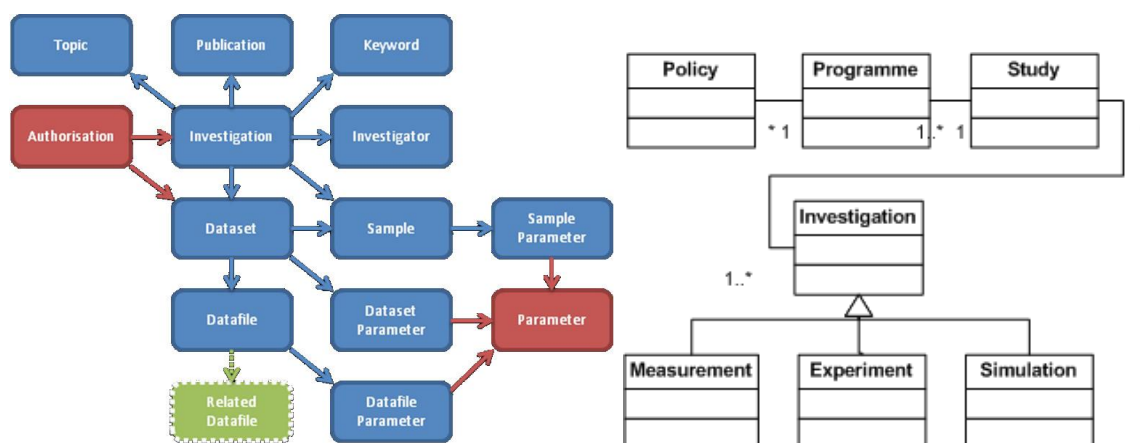


Abb. 9: Die Entitäten des Core Scientific Metadata Modells

### 3.1.4 Subject-centric Research Object Model

Das Metadatenmodell Subject-centric Research Object Model<sup>25</sup> ist aus seinem Vorgänger PSS (**P**roject **S**ubject **S**tudy object model) hervorgegangen. Es stellt eine Kombination aus DICOM, XCEDE2 und CSMDM dar. In [Lohrey 2009] wird das Objekt-Modell zusammen mit Workflows für Forschungsmethoden, einer Datenablage für Experimente und dem Konzept der Provenienz als Framework vorgestellt. Das Framework wurde basierend auf dem Objektmodell mit der kommerziellen Datenmanagement-Plattform Mediflux<sup>TM26</sup>, welche ebenfalls von Lohrey als CTO der australischen Firma Arcitectra<sup>TM</sup> entwickelt wurde, in Kooperation mit dem Centre for Neuroscience (University of Melbourne, Victoria, Australien) umgesetzt. Das Modell wurde entwickelt um neurowissenschaftliche MRT- und Mikroskopie-Experimente begleitend zu dokumentieren, wobei das Subjekt der Forschung, also Patient, Proband, Versuchstier oder Phantom, im Mittelpunkt der Betrachtung steht. Hervorzuheben ist, dass ein Subjekt, welches innerhalb eines Forschungsprojektes untersucht wird, mit verschiedenen Untersuchungsmethoden, die wiederum aus unterschiedlichen Schritten bestehen, assoziiert wird, wie Abb. 10 zeigt. Das Metadatenmodell wurde in XML implementiert, ist jedoch nicht freiverfügbar.

<sup>25</sup> <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2715266/>

<sup>26</sup> <http://www.arcitectra.com/Products/Mediaflux>

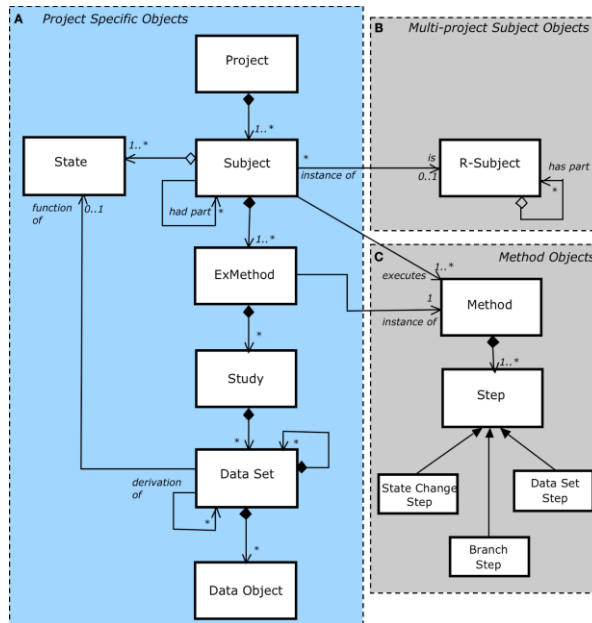


Abb. 10: Entitäten in der Subjekt-bezogenen Forschung

### 3.1.5 EDL

EDL (**Experiment Description Language**) wurde in [Hollmann 2008] vorgestellt und in Magdeburg am Institut für Biometrie und Informatik entwickelt. Das Metadatenmodell, wie es in Abb. 11 gezeigt wird, ist in XML implementiert und dient der einfachen und schnellen Verteilung von Konfigurationen für Echtzeit-fMRT-Experimente. Es wurde entwickelt, damit ein Experimentalleiter alle Einstellungen und Parameter der an einem rt-fMRT-Experiment beteiligten Geräte im Vorfeld bestimmen und in XML speichern kann. Diese Einstellungen können so durch die XML-Beschreibung in die Geräte geladen werden und das Experiment kann zügig durchgeführt werden.

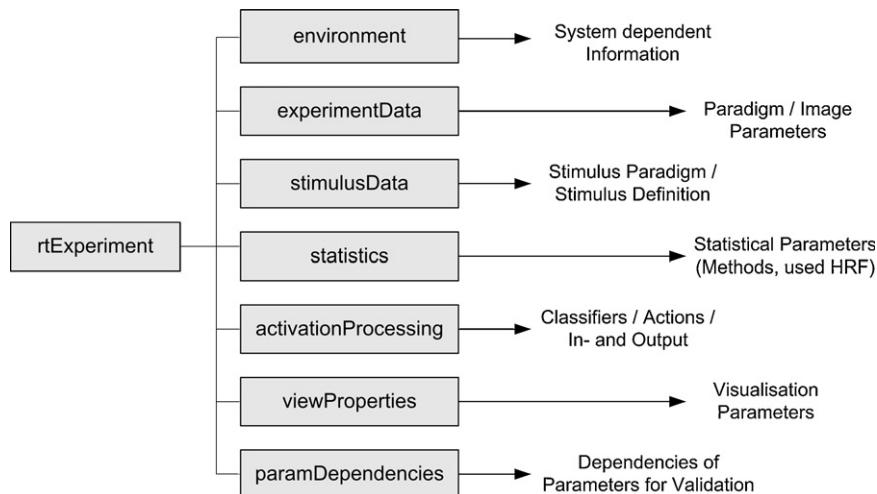


Abb. 11: EDL wurde als Metadatenformat speziell für Echtzeit-FMRT-Experimente entwickelt

Neben Experimentalplanung und Gerätekonfiguration dienen die Metainformationen nach dem Experiment der Zuordnung der Daten und Umgebungsbedingungen zu den durchgeführten Auswertungen, da die statistischen Parameter, Informationen zum Stimulus und anderen Attribute des Paradigmas in EDL beschrieben werden. Das Metadatenformat ist allerdings nicht für die Archivierung von Forschungsdaten vorgesehen. Da es sich hierbei um eine Eigenentwicklung handelt, gibt es kein Managementsystem, das dieses Metadatenformat unterstützt.

## 3.2 Archivierungstoolkits

In einschlägigen Publikationen werden verschiedene Forschungsdatenmanagementsysteme vorgestellt. Einige davon wie COINS (**C**ollaborative **I**nformatics and **N**euroimaging **S**uite) [Scott 2011], RepSIS (**R**epository for **S**tandardized Medical Image and **S**ignal case data references) [Deserno 2011], SciPort (Platform for Scientific Data Management and Integration) [Wang 2009], LORIS (**L**ongitudinal **O**nline **R**esearch and **I**maging **S**ystem) [Das 2012] und Mediaflux<sup>TM</sup> [Lohrey 2009] sind Eigenentwicklungen von Forschungseinrichtungen oder kommerziellen Firmen und stehen nicht frei zur Verfügung. HID (**H**uman **I**maging **D**atabase) ist eine Entwicklung aus dem durch die NIH finanzierte Projekt BIRN (**B**iomedical **I**nformatics **R**esearch **N**etwork), die nur durch teilnehmende Institute in den USA genutzt werden kann. Neben diesen gibt es auch im Internet veröffentlichte frei nutzbare Plattformen. Hierzu gehören DataFinder [Schlauch 2007] und DFBidb [Adamson 2010]. Diese stellen allerdings keine zentrale Server-basierte Architektur, die per Web-Schnittstelle erreichbar ist, zur Verfügung, sondern stellen vielmehr digitale Laborbücher dar, in denen sich zusätzlich die Primärdaten aus den lokalen Experimenten verwalten lassen. Das einzige frei nutzbare, quelloffene und erweiterbare Framework, um zentral Bild- und Forschungsdaten verschiedener Einrichtungen zu verwalten und im Internet frei zugeben, ist XNAT.

### 3.2.1 eXtensible Neuroimaging Archive Toolkit

XNAT<sup>27</sup> (**eX**tensible **N**euroimaging **A**rchive **T**oolkit) ist eine Management-Plattform für Projekte aus der medizinischen Bildgebung. Es werden von Haus aus die Modalitäten MRT, PET, CT, und Röntgen für Forschungsprojekte als auch für klinische Studien unterstützt. Das quelloffene Framework ist erweiterbar und individuell anpassbar und wird durch die Entwickler sowie einer starken Community unterstützt. Das Copyright für das Framework liegt bei der Harvard University (Cambridge, Massachusetts, USA) und dem Howard Hughes Medical Institute der Washington University (St. Louis, Missouri, USA). Entwickelt wurde es von Randy Buckner und Daniel Marcus zusammen mit der Neuroinformatics Research Group<sup>28</sup> [Marcus 2007].

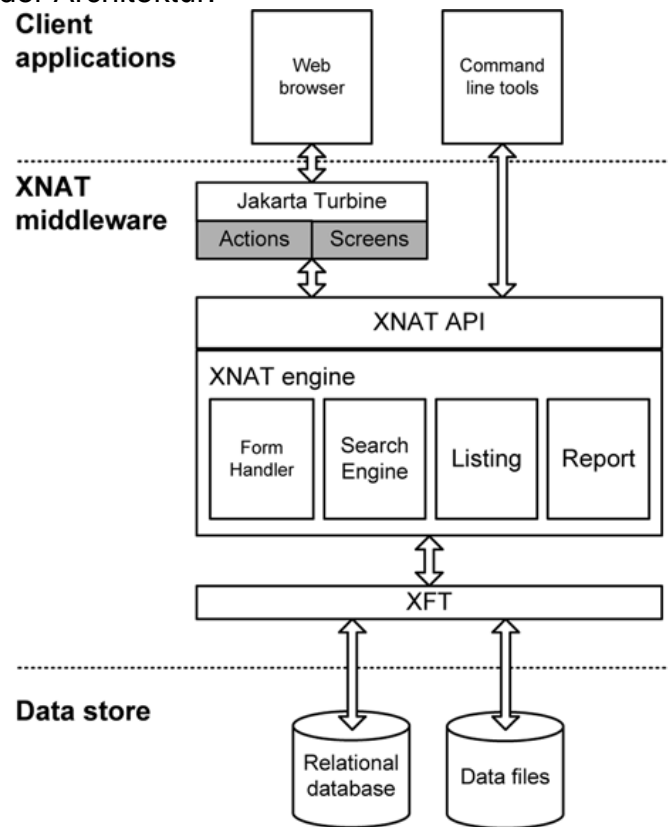
Die Architektur des Frameworks ist in drei Schichten aufgebaut, wie Abb. 12 zeigt. Die untere Schicht bildet die serverseitige Datenhaltung. Bilddaten, Dokumente und sonstige Dateien werden hier in einem Archiv-Verzeichnis auf dem Server gespeichert. Alle Metainformationen zu Projekten, Experimenten Messungen, Probanden und Bilddaten sowie der Speicherort der Dateien im Verzeichnis und alle Nutzer, Rollen und Zugriffsberechtigungen werden in einer relationalen SQL-Datenbank verwaltet. Die zweite Schicht, die Middleware, besteht aus dem eXtensible Formatting Tool (XFT) welches per SQL-Query auf die Datenbank zugreift und die Informationen im XML-Format an die XNAT engine weiterreicht. Die XNAT engine setzt sich aus Formhandler, Search engine sowie Listing- und Report-Funktionalität zusammen, verarbeitet die XML-Informationen und generiert entsprechende Formulare, Masken und Reportansichten in HTML für die Dartsellung der XNAT-Dateninhalte in einer Webseite. Auf diese Komponenten setzt das Application Programming Interface auf (API). Es dient zum einen als Schnittstelle für Entwickler, die zusätzliche funktionelle Erweiterungen für XNAT schaffen wollen. Zum anderen können die Nutzer von

---

<sup>27</sup> <http://xnat.org/about>

<sup>28</sup> <http://nrg.wustl.edu>

XNAT durch die vom Webapplication Framework Apache Turbine™<sup>29</sup> generierten Webseiten oder ein Kommandozeilentool die XNAT engine über die API ansteuern. Die Nutzerschnittstelle in Form von Web-Frontend und Kommandozeilentool bilden die obere Schicht der Architektur.



**Abb. 12:** Die Drei-Schichten-Architektur von XNAT

Die Kernfunktionalitäten von XNAT bilden der Datenupload, die Datenorganisation inklusive Datenfreigabe, die Online-Betrachtung und der Download der Bilddaten, die Sicherheits- und Zugriffsmechanismen, die Suchfunktion und Weiterleitung der Bilddaten an einer Verarbeitung-Pipeline. Der Datei-Upload wird durch DICOM-Protokoll (C-STORE), Upload-Tool für ZIP-Archive oder formularbasierte Eingabemaske im Browser bewerkstelligt. Die Datenorganisation erfolgt analog zu den Metadatenmodellen aus Abschnitt 3.2 in einer Hierarchie der Entitäten *Project*, *Subject* (Proband, Patient, Tier, Phantom) und *Experiment*, *Subject Assessment* (Untersuchung eines Subjekts), *Imaging Session*, *Scan*, *Reconstruction* und *Image Assessment* (Weiterverarbeitung der Bilddaten), wie Abb. 13 zeigt. Diese Entitäten besitzen zahlreiche generalisierte Attribute mit qualitativen und quantitativen Merkmalen für die Neurowissenschaftliche Bildgebung, wie z.B. Projekteigenschaften, Probandendaten und Scannerparameter. Die Entitäten und Ihre Merkmale werden intern als XML Schema Definitionen (XSDs) verwaltet. Die konkreten die Metainformationen zu einem Projekt, Experiment usw. werden entsprechenden in XML repräsentiert und kommuniziert und analog zu den XSDs in SQL an die Datenbank übergeben. Bemerkenswert ist, dass in XNAT kein starres Metadatenschema wie in den Beispielen aus Abschnitt 3.1 verwendet wird. Jede der vorgegebenen Entitäten lässt sich erweitern. Dies erfolgt durch Definition neuer Entitäten in XSD unter Vererbung der Attribute vorhandener Entitäten. Somit werden allgemeine Entitäten erweitert und spezialisiert.

<sup>29</sup> <http://turbine.apache.org>

Die Datenfreigabe erfolgt über eine mehrstufige Autorisierung. So können einzelne Entitäten für Auflistung, zur Einsicht oder zum Download für verschiedene Nutzergruppen freigegeben werden. Die Betrachtung der Daten erfolgt im Webbrowser durch eine Java-Applet-Implementation von ImageJ<sup>30</sup>, welches Zahlreiche medizinische Datenformate darstellen kann. Die Exploration großer Datenmengen wird durch spezielle Suchmasken, die ebenfalls aus den XSDs generiert werden, ermöglicht. Ergebnisse einer Attributbasierten Suche lassen sich übersichtlich in Listen oder detailliert in Reports im Webbrowser anzeigen. Angebundene HPC-Ressourcen wie Rechencluster können Daten aus dem System heraus verarbeiten und auswerten, sofern entsprechende Verarbeitungspipelines definiert wurden und im Webbrowser auf bestimmte Datensätze angewendet werden. Der klassische Workflow gliedert sich in Datenupload ins Prearchiv (Daten-Cache), Validierung der Daten und anschließende Übertragung ins Zentralarchiv sowie Bearbeitung und Freigabe der Daten auf. Die Vorteile des XNAT-Systems liegen in der Erweiterbarkeit, den Sicherheitsmechanismen, dem webbasierten Zugriff und der Bildvorschau. Nachteile sind der hohe administrative Aufwand sowie die Ressourcenbereitstellung und Metadatenerstellung und -pflege. Die Dokumentation des Frameworks kann auf den Projektseiten im Wiki oder in der Veröffentlichung [Marcus 2007] eingehend nachgeschlagen werden oder aus der [Präsentation vom Workshop](#) in Magdeburg rekapituliert werden.

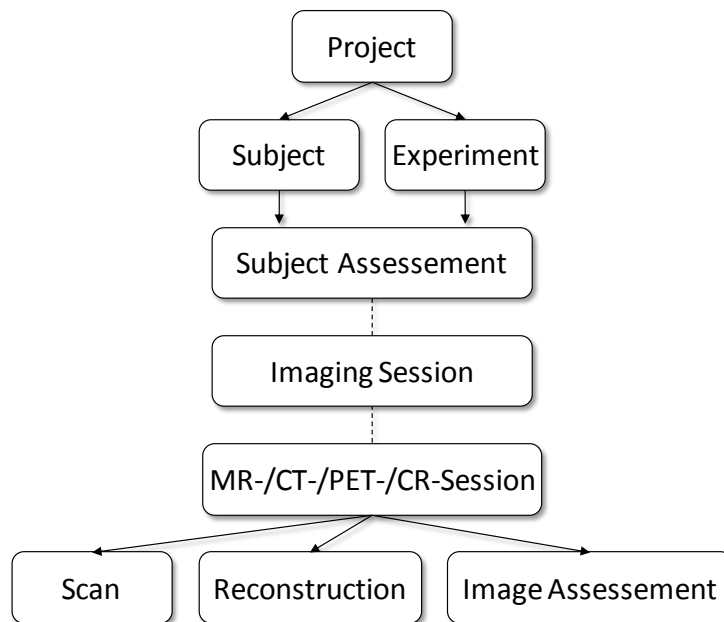


Abb. 13: Die Hierarchie der Entitäten in XNAT

### 3.2.2 DFBidb Software for Neuroimaging Datamanagement

DFBidb<sup>31</sup> ist in Python programmiert und läuft nur auf Linux-Plattformen. Es stellt keine Client-Server-Lösung dar und stellt keine Netzwerkanbindung bereit. Die Datenbasis wird durch in XML beschriebene Verzeichnisbäume mit TAR-Archiven realisiert, wobei die Zugriffsrechte und Sicherheitsmechanismen ausschließlich über Unix-basierte Access Control Lists (ACLs) umgesetzt wird. DFBidb stellt somit eher eine Laborsoftware bzw. eine digitales Laborbuch mit Primärdatenspeicherung dar, als ein zentrales Archivsystem nach OAIS-Modell. Es wurde dazu entwickelt im lokalen Netzwerk eines MR-Scanners alle Experimente und Studien der dort arbeitenden

<sup>30</sup> <http://rsb.info.nih.gov/ij/applet>

<sup>31</sup> <http://www.nitrc.org/projects/dfbidb>

Forschergruppen zu verwalten. Positiv ist, dass das Tool mit einer Erweiterung des Metadaten-Formats XCEDE2 arbeitet und somit Datensätze detailliert mit Attributen und Eigenschaften dokumentiert werden können und somit auch komplexe Suchanfragen an die Datenbank ausgeführt werden können. [Adamson 2010]

### 3.2.3 Datafinder

DataFinder<sup>32</sup> ist ein im Auftrag des Deutschen Luft- und Raumfahrtzentrums (DLR) entwickeltes Client/Server Framework zur Dokumentation und Verwaltung von großen wissenschaftlichen und technischen Datenmengen, die während der experimentellen Arbeit anfallen. Die Metainformationen werden strukturiert in XML auf einem WebDAV-Server abgelegt, die Daten werden in heterogenen Stageservern gehalten, wie Abb. 14 zeigt. Der auf Python und Qt basierende Client ist für Linux, Windows und SGI Irix verfügbar und kann die Daten über das Netzwerk abrufen [Schlauch 2007], [Ney 2011].

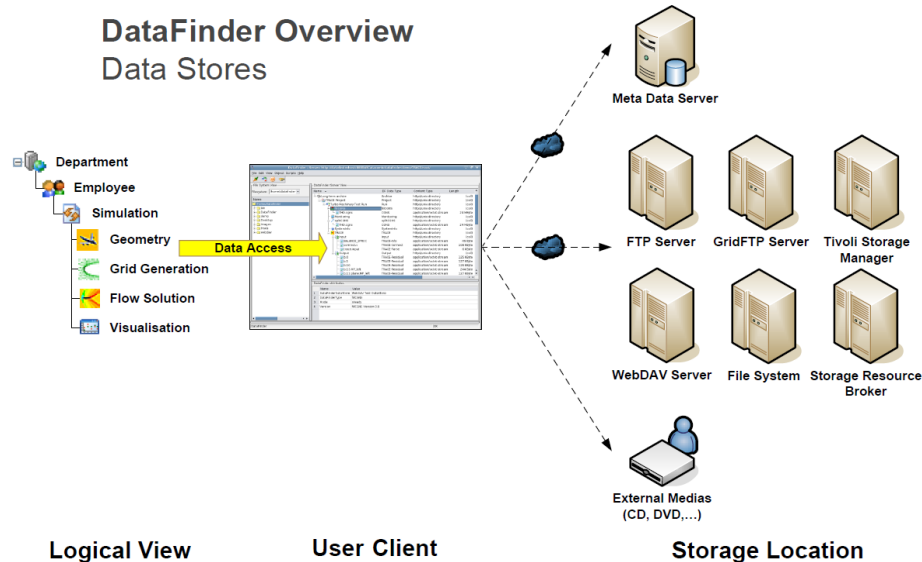


Abb. 14: Architektur der DataFinder Speicherinfrastruktur

Die Client-Anwendung dient der Administration dem Nutzerzugang zum Datenarchiv. Der WebDAV-Server muss mit Hilfe eines Apache Webservers inklusive der DAV-Erweiterung (mod\_dav) und dem Modul Catacomb sowie einer relationalen Datenbank durch die Forschungseinrichtung bereitgestellt werden. Die Metadaten zum Anwendungsfall müssen durch die jeweiligen Forscher selbst definiert und modelliert werden, wie Abb. 15 zeigt.

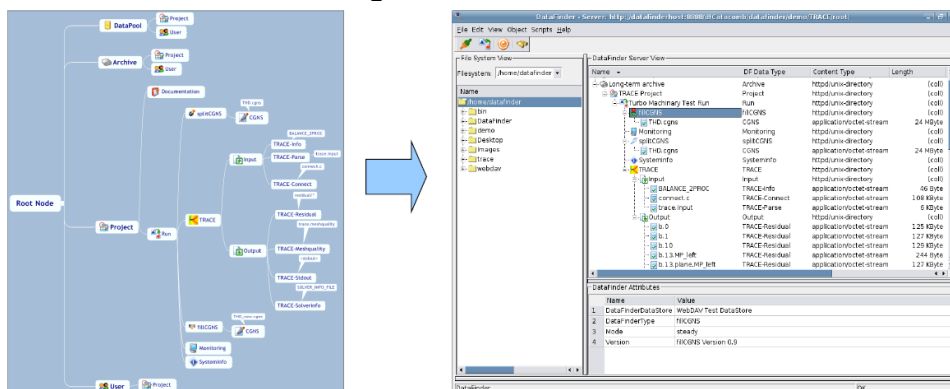


Abb. 15: Metadatenverwaltung mit DataFinder

<sup>32</sup> <http://www.dlr.de/datafinder>

## 4 Diskussion und Ausblick

### 4.1 Wahl des Forschungsdatenmanagementsystems

HID bietet leider nur die Clienttools, um Daten im BIRN-Netzwerk zu speichern, fällt also somit für das Projekt raus. fMRI DataCenter<sup>33</sup> [Horn 2002] ist ein reiner Webdienst, der Primärdaten zu Publikationen bereitstellt und kann hier als Referenzbeispiel für den Zugriff und die Veröffentlichung von Primärdaten betrachtet werden. SciPort von Siemens Research existiert nur in der Literatur und kann nur über direkte Kooperation mit Siemens genutzt werden. Autor des Frameworks ist Fusheng Wang<sup>34</sup> und Ansprechpartner in Deutschland ist Fabian Moerchen<sup>35</sup>.

Anhand einschlägiger Literatur sind weitere Systeme relevant: COINS, da es DICOM-Funktionalität umsetzt, sowie RepSIS, da es auf Basis des OAIS-Modells erstellt wurde und von der DFG gefördert wurde. Leider sind diese beiden Systeme nicht freiverfügbar, sondern wurden aus Fördermittel in dedizierten Einrichtungen entwickelt und außerhalb publizierter Journalartikel nicht öffentlich dokumentiert. Die lokale Datendokumentation durch DFBidb kann bereits vor der eigentlichen Langzeitarchivierung im Forschungsprozess helfen, Rohdaten und Metadaten kontinuierlich und vollständig zu erfassen. DFBidb ist allerdings für ein Forschungsdatenmanagement kooperierender Einrichtungen nicht geeignet, da das Programm als Lokalinstallation auf einem einzelnen Laborrechner nicht an ein zentrales Archiv angebunden werden kann und das Programm nicht via Standard-Netzwerkprotokolle von verschiedenen lokal getrennten Nutzern angesteuert werden kann.

**Tab. 1:** Merkmale und mehrstufige Ausprägungen zum Thema Forschungsdatenarchivierung nach der GMA-Analyse aus D4.1

Merkmal	Ausprägung 1	Ausprägung 2	Ausprägung 3	Ausprägung 4	Ausprägung 5	Ausprägung 6
Infrastrukturarchitektur	lokal	zentral	hierarchisch	verteilt	hybrid	
Schnittstellen	PACS	MR-Datenbank	Labordatenbank			
Software	Open Source	Frei	Kommerziell	Eigenentwicklung	Auftragsentwicklung	Hybrid
Speicherperformance	Langsam: Nearline	Schnell: Online	Hochverfügbar	HSM		
Datenspeicherstruktur	Dateisystem	XML-DB	SQL-DB	Hybrid	Proprietär	
Datenabnehmer	Intern	Extern	Verbund	Wissenschaft	öffentlich	
Datenschutz	kein Schutz	Pseudonymisierung	Pseudonymisierung 2.Stufe	Anonymisierung		
Nutzerrollen	Besitzer	Mitarbeiter	Kollege	Partner	Öffentlich	
Authentifizierung	keine	Beantragung	User/PWD	PKI	Biometrisch	
Domäne	Projekt	Abteilung	Einrichtung	Verbund	National	International
Umfang der Archivierung	nur Rohdaten	inkl. Metadaten	inkl. Methodik	inkl. Provinienz	inkl. Ergebnis	inkl. Publikation
Speicherbedarf	bis 10 Terabyte	bis 100 Terabyte	bis 1 Petabyte	bis 10 Petabyte	bis 100 Petabyte	bis 1 Exabyte
Nachnutzungsszenarien	Wiederaufnahme	Wiederverwendung	Weiterverwendung	Bewertung	Vergleichsstudie	Verifizierung (GSP)
Archivierungsfrist	bis 5 Jahre	bis 10 Jahre	bis 15 Jahre	bis 30 Jahre	bis 100 Jahre	unlimitiert
Archivierungsbedarf	unlimitiert	bis 100 Jahre	bis 30 Jahre	bis 15 Jahre	bis 10 Jahre	bis 5 Jahre
Löschauslöser	Projektende	Publikation	Führungswechsel	Umstrukturierung	Auflösung	Gesetzliche Frist
Metadatengenerierung	manuell	importiert	halbautomatisch	vollautomatisch		
Datenformat	nur DICOM	nur Archive	Definierte Formate	Diverse Formate		
Metadatenmodell	DICOM	CSMD	XCEDE 2	XNAT	COINS	

XNAT, die Client-Server-Lösung mit erweiterbarem Metadatenmodell, hat sich in den letzten Jahren rasant in der neurowissenschaftlichen Gemeinschaft verbreitet und etabliert. Es ist hoch skalierbar und anpassbar und wird vom Entwicklerteam kontinuierlich weiterentwickelt. Die Komponenten des Frameworks sind hervorragend dokumentiert<sup>36</sup> zum stehen frei zum Download bereit<sup>37</sup>. Für den schnellen Einstieg kann

<sup>33</sup> <http://www.fmridc.org/fmridc/aboutus/index.html>

<sup>34</sup> [http://cci.emory.edu/cms/people/pages/wang\\_fusheng.html](http://cci.emory.edu/cms/people/pages/wang_fusheng.html)

<sup>35</sup> <http://www.linkedin.com/in/fabianmoerchen>

<sup>36</sup> <https://wiki.xnat.org>



XNAT sogar als Produktivsystem in einer virtuellen Maschine vorkonfiguriert heruntergeladen werden. XNAT unterstützt zwar bisher noch keine kombinierten Studien wie z.B. EEG/fMRT, ist allerdings so flexibel erweiterbar und anpassbar, dass dies, wenn notwendig, in Eigenentwicklung umgesetzt und dann der Community zugänglich gemacht werden kann. Vor allem aber erfüllt XNAT die in der GMA-Analyse aus Deliverable D4.1 gestellten Anforderungen an ein Forschungsdatenmanagement in der medizinischen Bildgebung (siehe Tab. 1).

Auswertewerkzeuge besitzen eine weitreichende Versionsgeschichte und werden fortlaufend in ihrem Funktionsumfang erweitert. Die Auswertungen werden zum Teil GUI-gesteuert durch zahlreiche Parameter konfiguriert. Um lückenlos dokumentieren zu können, wie mit einem Auswerte-Werkzeug gearbeitet wurde, um ein wissenschaftliches Ergebnis zu erzeugen, ist es notwendig Softwareversion und Ausprägung aller Auswerteparameter zu erfassen. Hierfür gibt es weder Exportfunktionen in den Werkzeugen noch wird die Erfassung dieser Provenienzinformationen durch Forschungsdatenmanagementsysteme unterstützt. XNAT bietet jedoch mit seinem generischen Metadatenmodell die Möglichkeit ein vorhandenes Metadatentupel um entsprechende Attribute zu erweitern. Diese generische Erweiterbarkeit gepaart mit dem Webbasierten Zugriffs- und Veröffentlichungsmöglichkeiten machen XNAT gegenüber anderer Metadatenmodelle, wie Tab. 2 zeigt, zu einer idealen Basis für Forschungsdatenmanagement und Langzeitarchivierung.

**Tab. 2:** Bewertung der Metadatenformate aus Abschnitt 3.1

Format	Hierarchie	Projektebene	Provenienz	Erweiterbarkeit
DICOM	+	-	0	-
XCEDE 2	+	+	+	-
CSMDM	+	+	+	0
ScROM	+	+	+	0
EDL	0	-	0	-
XNAT	+	+	+	+

Anforderung erfüllt (+), teilweise erfüllt (0), nicht erfüllt (-)

## 4.2 Fallbeispiel Leibniz-Institut für Neurobiologie Magdeburg

XNAT wird im Leibniz-Institut für Neurobiologie (LIN) in Magdeburg bereits zur Grundlagenforschung eingesetzt. Nutzungszahlen zeigen, dass bis zum 20.06.2012 in XNAT 81 Projekte mit insgesamt 2708 Messungen von 705 Probanden gespeichert wurden. Pro Jahr kommen ca. 10 Projekte hinzu. Bislang gibt es 14 aktive Nutzer und einen Administrator. Pro Tag ist ein Nutzer aktiv. Daneben erfolgt ein wesentlicher Teil der Zugriffe automatisiert über Skripte zur internen Abrechnung durchgeführter Messungen und Studien. Nach einer Messung arbeiten die Forscher auf lokalen Kopien und verwenden XNAT in erster Linie, um für aktuelle Studien bereits im System vorhandene Probanden nach demographischer Eignung zu rekrutieren.

Hier werden mit XNAT allerdings nur Projekte, Experimente, Messungen und Probanden als Metainformationen verwaltet. Die tatsächlichen Forschungsdaten liegen auf einem gespiegelten Storage-Area-Network und werden als Samba-Netzwerk-Partition jedem Mitarbeiter auf der jeweiligen Workstation bereitgestellt. Die Daten

<sup>37</sup> <http://xnat.org/download>

der unterschiedlichen Projekte werden mittels Unix-basierten Access-Control-Lists unterschiedliche je nach Nutzern und Gruppen eines Projektes zum Lesen, Schreiben und Ausführen freigegeben. Dabei folgt die Struktur und Benennung des Verzeichnisbaums eines Projektes und der darin enthaltenen Daten im Dateisystem einer strengen Konvention, die von jedem Mitarbeiter und Projekt eingehalten werden muss. So finden sich in jedem Projektverzeichnis die Projektdokumente, Primärdaten, vorverarbeiteten Daten und Auswertergebnisse am gleichen Ort und sind von jedem Mitarbeiter nach einem vorgegebenen Schema auffindbar. Die Projekt- und Experimentnamen entsprechen denen die in XNAT verwendet wurden. So kann ein Mitarbeiter in XNAT über verschiedene Metainformationen zu einem Messvorgang, ein bestimmtes Experiment identifizieren und die Daten dazu im Samba-Share unter dem entsprechenden Namen finden.

XNAT wird hier also als Metadaten-Managementsystem zur Erfassung und Verwaltung von Projekten, Experimenten und Messungen genutzt, wohingegen die eigentlichen Forschungsdaten in einem Dateisystem auf einem Unix-basierten Stageserver lagern. Die ist zum Teil deshalb möglich, weil alle Mitarbeiter unter den gleichen Bedingungen an die IT-Infrastruktur des LIN angebunden sind und im LIN ausschließlich das DICOM-Format verarbeitet wird und alle Auswertungen mit BrainVoyager durchgeführt werden. Dadurch wird innerhalb des Instituts eine gewisse Homogenität des Forschungsvorgangs gewährleistet und somit Transparenz der Daten und Vorgänge garantiert. Zusätzlich werden alle am Scanner erzeugten Rohdaten gekennzeichnet nach Aufnahmedatum und Messverantwortlichen gesondert gesichert. Dazu ist ein weiteres Stagesystem als Backupereinheit mit dem Scanner verbunden.

### **4.3 Daten durch DOIs via REST-API referenzieren**

In den Geowissenschaften werden seit einiger Zeit nicht nur Publikationen durch Digital Object Identifier (DOIs) referenziert und im Internet veröffentlicht sondern ebenfalls die mit der Publikation verknüpften Forschungsprimär- und -ergebnisdaten [Klump 2008]. Würde sich dieses System ebenfalls in der medizinischen Bildgebung etablieren so könnten in XNAT freigegebene Datensätze einfach mit DOIs verknüpft werden. Dies wird durch die REST-API<sup>38</sup> von XNAT ermöglicht. Mit Hilfe der REST-API lässt sich ein Datenbankquery, durch Kombination der Attribute, die bei der Suche nach einem Datensatz diesen Datensatz als eindeutigen Treffer zurückgeben würden, an XNAT ausführen. Dabei wird die Syntax der REST-API an die URL der jeweiligen XNAT-Instanz angehängen. Es kann also durch die Angabe einer URL ein öffentlich freigegebener XNAT-Datensatz eindeutig referenziert und im Web-Browser aufgerufen werden. Diese URL kann dann mit einer DOI verknüpft werden und bei Publikationen als zusätzliche Quelle für ergänzende Informationen angegeben werden. Hierin zeigt sich erneut, dass XNAT eine hervorragende Basis für ein langfristiges und nachhaltiges Forschungsdatenmanagement für die medizinische Bildgebung bereitstellt.

---

<sup>38</sup> <http://xnat15.wikispaces.com/XNAT+REST+API+Usage>

## 5 Literaturverzeichnis

- Adamson, C. L.; Wood, A. G.: *DFBdb: a software package for neuroimaging data management*. In: Neuroinformatics 8(4), S. 273–284, **2010**.
- Ashburner, J.: *SPM: A history*. In: NeuroImage 62(2), S. 791–800, **2012**.
- Bennett, C. M.; Baird, A. A.; Miller, M. B.; Wolford, G. L.: *Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction*. In: Neuroimage 47 Supplement 1(15th Annual Meeting of the Organization for Human Brain Mapping), 456 SA-PM, **2009**.
- Büttner, S.; Hobohm, H.-C.; Müller, L.: *Handbuch Forschungsdatenmanagement*. Bock + Herchen, Bad Honnef, **2011**.
- Consultative Committee for Space Data Systems (Hrsg.): *Reference Model for an Open Archival Information System (OAIS). Recommendation for Space Data System Standards. Blue Book, Issue 1*. Program Integration Division. National Aeronautics and Space Administration, Washington DC, **2002**.
- Cox, R. W.: *AFNI: software for analysis and visualization of functional magnetic resonance neuroimages*. In: Comput. Biomed. Res. 29(3), S. 162–173, **1996**.
- Crum, W. R.: *Magnetic Resonance Brain Image Processing and Arithmetic with FSL*. In: Methods Mol Biol 711(2), S. 109–126, **2011**.
- Das, S.; Zijdenbos, A. P.; Harlap, J.; Vins, D.; Evans, A. C.: *LORIS: a web-based data management system for multi-center studies*. In: Front. Neuroinform. 5, **2012**.
- Deserno, T. M.; Welter, P.; Horsch, A.: *Towards a Repository for Standardized Medical Image and Signal Case Data Annotated with Ground Truth*. In: J Digit Imaging 24(11), **2011**.
- Deutsche Forschungsgemeinschaft (Hrsg.): *Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Empfehlungen der Kommission Selbstkontrolle in der Wissenschaft. Denkschrift*. Wiley-VCH, Weinheim, **1998**.
- Dickmann, F.; Grütz, R.: *LABIMI/F - Digital Preservation of Biomedical Research Data*. In: Knowledge Exchange Workshop Research Data Management - Activities and Challenges(Wissenschaftszentrum Bonn), **2011**.
- Eklund, A.; Andersson, M.; Josephson, C.; Johannesson, M.; Knutsson, H.: *Does parametric fMRI analysis with SPM yield valid results? An empirical study of 1484 rest datasets*. In: NeuroImage 61(3), S. 565–578, **2012**.
- Fedorov, A.; Beichel, R.; Kalpathy-Cramer, J.; Finet, J.; Fillion-Robin, J.-C.; Pujol, S. et al.: *3D Slicer as an image computing platform for the Quantitative Imaging Network*. In: Magn Reson Imaging 30(9), S. 1323–1341, **2012**.
- Friston, K. J.: *Statistical parametric mapping. The analysis of functional brain images*. 1. Aufl. Elsevier/Academic Press, Amsterdam ;Boston, **2007**.
- Fischl, B.: *FreeSurfer*. In: NeuroImage 62(2), S. 774–781, **2012**.
- Gadde, S.; Aucoin, N.; Grethe, J. S.; Keator, D. B.; Marcus, D. S.; Pieper, S.: *XCEDE: An Extensible Schema for Biomedical Data*. In: Neuroinformatics, **2011**.
- Goebel, R.: *Brain Voyager - Past, present, future*. In: NeuroImage 62(2), S. 748–756, **2012**.

- Gouws, A.; Woods, W.; Millman, R.; Morland, A.; Green, G.: *DataViewer3D: An Open-Source, Cross-Platform Multi-Modal Neuroimaging Data Visualization Tool*. In: *Front Neuroinform* 3, S. 9, **2009**.
- Halchenko, Y. O.; Hanke, M.: *Open is Not Enough. Let's Take the Next Step: An Integrated, Community-Driven Computing Platform for Neuroscience*. In: *Front. Neuroinform.* 6, **2012**.
- Hodge, G. M.: *Best Practices for Digital Archiving. An Information Life Cycle Approach*. In: *D-Lib Magazine* 6(1), **2000**.
- Hollmann, M.; Mönch, T.; Mulla-Osman, S.; Tempelmann, C.; Stadler, J.; Bernarding, J.: *A new concept of a unified parameter management, experiment control, and data analysis in fMRI: application to real-time fMRI at 3T and 7T*. In: *J Neurosci Methods* 175(1), S. 154–162, **2008**.
- Horn, J. D. van; Gazzaniga, M. S.: *Databasing fMRI studies — towards a 'discovery science' of brain function*. In: *Nat Rev Neurosci* 3(4), S. 314–318, **2002**.
- Horn, J. D. van; Toga, A. W.: *Is it time to re-prioritize neuroimaging databases and digital repositories?* In: *Neuroimage* 47(4), S. 1720–1734, **2009**.
- International Organisation for Standardization (Hrsg.): *Space data and information transfer systems. Open archival information system. Reference Model*. ISO14721:2003 (49.140), **2010**.
- Jenkinson, M.; Beckmann, C. F.; Behrens, T. E. J.; Woolrich, M. W.; Smith, S. M.: *Fsl*. In: *NeuroImage* 62(2), S. 782–790, **2012**.
- Keator, D. B.; Gadde, S.; Grethe, J. S.; Taylor, D. V.; Potkin, S. G.: *A general XML schema and SPM toolbox for storage of neuro-imaging results and anatomical labels*. In: *Neuroinformatics* 4(2), S. 199–212, **2006**.
- Keator, D. B.; Wei, D.; Gadde, S.; Bockholt, J.; Grethe, J. S.; Marcus, D. et al.: *Derived Data Storage and Exchange Workflow for Large-Scale Neuroimaging Analyses on the BIRN Grid*. In: *Front Neuroinform* 3, S. 30, **2009**.
- Klump, J.; Brase, J.; Diepenbroek, M.; Grobe, H.; Hildenbrandt, B.; Höck, H. et al.: *Use of persistent identifiers in the publication and citation of scientific data*. In: *AGU fall meeting, San Francisco, CA, USA(519)*, **2008**.
- Lohrey, J. M.; Killeen, N. E. B.; Egan, G. F.: *An Integrated Object Model and Method Framework for Subject-Centric e-Research Applications*. In: *Front Neuroinform* 3(19), **2009**.
- MacKenzie-Graham, A. J.; Horn, J. D. van; Woods, R. P.; Crawford, K. L.; Toga, A. W.: *Provenance in neuroimaging*. In: *Neuroimage* 42(1), S. 178–195, **2008**.
- Marcus, D. S.; Olsen, T. R.; Ramaratnam, M.; Buckner, R. L.: *The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data*. In: *Neuroinformatics* 5(1), S. 11–34, **2007**.
- Matthews, B.; Sufi, S.; Flannery, D.; Lerusse, L.; Griffin, T.; Gleaves, M.; Kleese, K.: *Using a Core Scientific Metadata Model in Large-Scale Facilities*. In: *Int J Digit Curation* 1(5), S. 106–118, **2010**.
- Menne, K. M. L.; Hofmann, U. G.: *Neurowissenschaftliche Datenbanken - eine Herausforderung*. In: *Datenbank-Spektrum* 10(9), S. 5–13, **2004**.

- Mildenberger, P.; Eichelberg, M.; Martin, E.: *Introduction to the DICOM standard*. In: Eur Radiol 12(4), S. 920–927, **2002**.
- Mullinger, K.; Bowtell, R.: *Combining EEG and fMRI*. In: Methods Mol Biol 711(2), S. 303–326, **2011**.
- Ney, M.: *Enabling a Data Management System to Support the Good Laboratory Practice. DataFinder*. Freie Universität, Berlin, **2011**.
- OECD (Hrsg.): *Principles and guidelines for access to research data from public funding*. OECD Publishing, Paris, **2007**.
- Pinowar, H. A.: *Who Shares? Who Doesn't? Factors Associated with Openly Archiving Raw Research Data*. In: PLoS ONE 6(7), S. e18657, **2011**.
- Poline, J.-B.; Breeze, J. L.; Ghosh, S.; Gorgolewski, K.; Halchenko, Y. O.; Hanke, M. et al.: *Data sharing in neuroimaging research*. In: Front Neuroinform 6(9), **2012**.
- Saad, Z. S.; Reynolds, R. C.: *SUMA*. In: NeuroImage 62(2), S. 768–773, **2012**.
- Schlauch, T.; Schreiber, A.: *DataFinder. A Scientific Data Management Solution*. In: PV 2007 Conference, **2007**.
- Scott, A.; Courtney, W.; Wood, D.; La Garza, R. de; Lane, S.; King, M. et al.: *COINS: An Innovative Informatics and Neuroimaging Tool Suite Built for Large Heterogeneous Datasets*. In: Front Neuroinform 5, S. 33, **2011**.
- Sufi, S.; Mathews, B.: *CCLRC Scientific Metadata Model Version 2*. In: CCLRC Tech Report(DL-TR-2004-001), **2004**.
- Toga, A. W.: *Neuroimage databases: the good, the bad and the ugly*. In: Nat Rev Neurosci 3(4), S. 302–309, **2002**.
- Toussaint, N.; Souplet, J.-C.; Fillard, P.: *Medinria: medical image navigation and research tool by INRIA*. In: Proceedings of MICCAI Workshop on Interaction in Medical Image Analysis and Visualization 10, **2007**.
- Van Essen, D. C.: *Cortical cartography and Caret software*. In: NeuroImage 62(2), S. 757–764, **2012**.
- Wang, F.; Hussels, P.; Liu, P.: *Securely and Flexibly Sharing a Biomedical Data Management System*. In: Proc. SPIE Med. Imag. 37(7264), **2009**.