




Datenmanagement in Forschung und Entwicklung mit Open-Source-Werkzeugen

Tobias Schlauch, DLR Simulations- und Softwaretechnik
03. März 2009

Deutsches Zentrum für Luft- und Raumfahrt (DLR)

- 5.700 Mitarbeiter arbeiten in 29 Forschungsinstituten und Einrichtungen in
 - 13 Standorten
- Büros in Brüssel, Paris und Washington





Datenhaltung in Forschung und Entwicklung

Position der Deutschen Forschungsgemeinschaft (DFG)

„Primärdaten, die Grundlage einer wissenschaftlichen Veröffentlichung sind, werden auf haltbaren und gesicherten Trägern in der Einrichtung, in der sie entstanden sind, mindestens zehn Jahre aufbewahrt.“

Aus Regel 3: DFG, Sicherung guter wissenschaftlicher Praxis und Verfahren bei wissenschaftlichem Fehlverhalten





Datenhaltung in Forschung und Entwicklung

Herausforderungen

- Projektorientiertes Arbeiten
 - Mitarbeiter an verschiedenen Instituten
 - Heterogene IT-Infrastrukturen
- Hohe Mitarbeiterfluktuation bedingt durch
 - Zeitverträge (Projektstellen)
 - Promotionen, Gastwissenschaftler

Entwicklung des Werkzeuges DataFinder zur Unterstützung der Mitarbeiter beim Datenmanagement.



DataFinder

DataFinder ist ein Werkzeug zur

- **Organisation gemeinsamer Daten** in einem Repository
- **Archivierung von Daten**

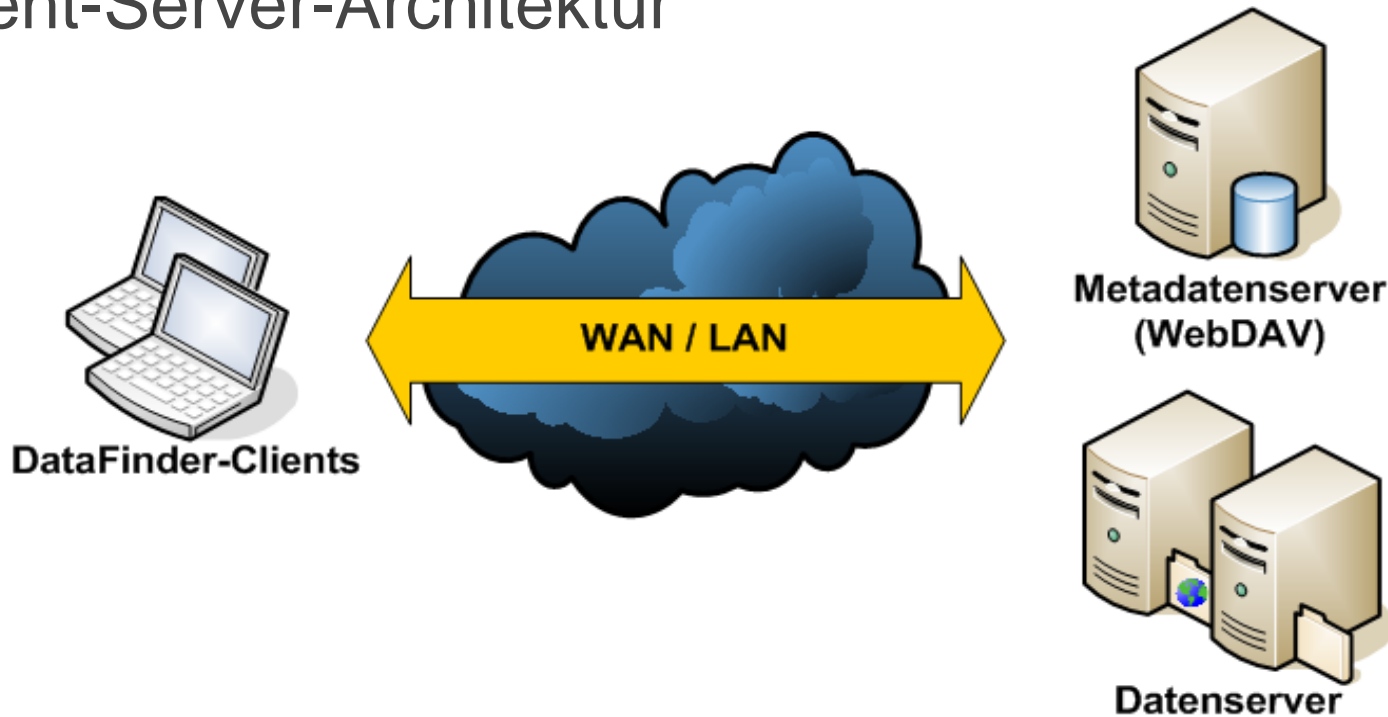
Verfügbar als Open-Source-Software (BSD-Lizenz)

- Weiterentwicklung durch das DLR
- Projektseite: <http://sourceforge.net/projects/datafinder/>



DataFinder

Client-Server-Architektur



Clients auf **unterschiedlichsten Plattformen verfügbar** (Linux, Windows, ...)

WebDAV-Server zur Speicherung der Konfiguration, Struktur und Metadaten erforderlich

Zusätzliche Speicherressourcen zur verteilten Datenspeicherung nutzbar

DataFinder

Datenmanagement-Client



DataFinder - Server: http://192.168.125.130/repos/data/sample/

File Edit View Object Scripts Help

File System View

Filesystem: /

Name	Size
lib	(directory)
lost+fo...	(directory)
media	(directory)
mnt	(directory)
opt	(directory)
proc	(directory)
root	(directory)
sbin	(directory)
srv	(directory)
sys	(directory)
tmp	(directory)
usr	(directory)
var	(directory)

DataFinder Server View

Name	Data Type	Content Type
sample		
Project	Directory	http://unix-direc
Documentation	Directory	http://unix-direc
project description.pdf	File	application/pdf
SimulationRun	Directory	http://unix-direc
Input Files	Directory	http://unix-direc
grid.dat	File	
params.txt	File	text/plain
Output Files	Directory	http://unix-direc

DataFinder Attributes

Name	Value
1	Data Type Directory

Log | Search Results | Script Output

Collection...

- Upload File...
- Download...
- Create
- Modify Attributes...
- Access Privileges...
- Find
- Cut (Ctrl+X)
- Copy (Ctrl+C)
- Copy Meta Data (Ctrl+M)
- Paste (Ctrl+V)
- Delete
- Rename
- Archive
- Collection...
- Offline Resource...
- Link
- Archive...

WebDAV-Server Catacomb

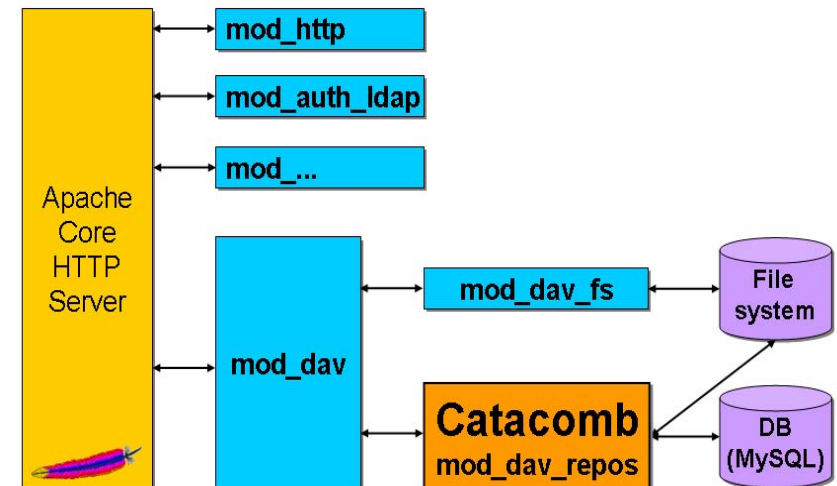


Erweiterte Alternative zum Apache-WebDAV-Modul

- Implementierung zusätzlicher WebDAV-Standards (Metadatenuche, Zugriffsrechte, Versionierung)
- Speicherung der Informationen im Dateisystem + Datenbank (Oracle, PostgreSQL, MySQL, SQLite 2/3, ...)

Verfügbar als Open-Source-Software (Apache-2.0-Lizenz)

- Weiterentwicklung durch das DLR
- Projektseite: <http://catacomb.tigris.org/>

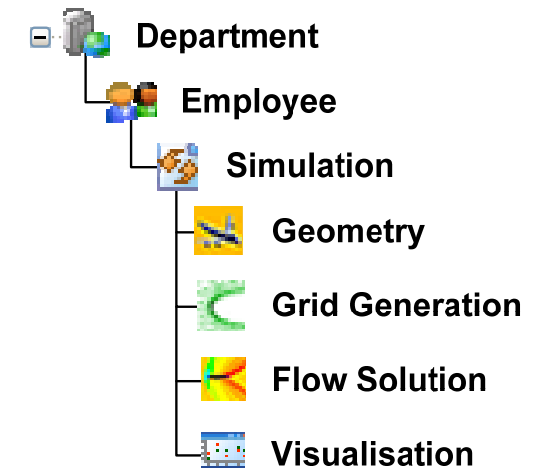


DataFinder

Vorteile im Datenmanagement

Strukturierung und Kennzeichnung der Daten

- Einführung einer **logischen Sicht** auf die Daten
- Modellierung der Metadaten & Struktur durch ein Datenmodell



Integrierbar in bestehende IT-Infrastrukturen

- Verwendung von offenen Standards (WebDAV, XML)
- Abstraktion der Speicherorts
- Flexible Einbindung von Speicherressourcen zur Datenablage durch die Unterstützung verschiedener Schnittstellen
 - z.B. WebDAV, FTP, GridFTP, Tivoli Storage Manager (TSM), Storage Resource Broker (SRB), ...

DataFinder

Anpassung an die Anwendung

Anforderungsanalyse

- Auswahl der Daten und Bestimmung der Zusammenhänge
- Festlegung der darauf aufbauenden Funktionalitäten

Datenmodellierung

- Abbildung der Datenzusammenhänge
- Definition von Eigenschaften der Datenobjekte

Funktionale Erweiterungen mit Hilfe von Python-Skripten

- Implementierung als Erweiterungen des Datenmanagementclients
- Python-API

DataFinder

Anwendungsbeispiele

Testdatenverwaltung im Automobilbau

- Organisation von Simulationsdaten zur Crash-Testsimulation eines Automobilzulieferers



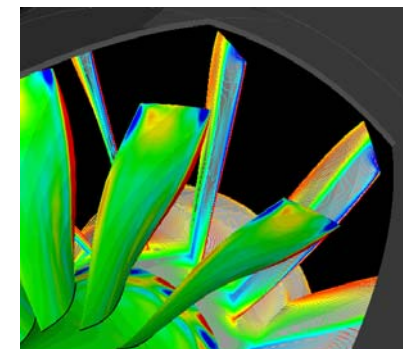
Dokumentenverwaltung im Luftverkehr

- Aufbau einer Datenbank zur Unterstützung des Luftverkehrs-Monitorings



Datenverwaltung in der Luftfahrtforschung

- Simulationsautomatisierung von CFD-Rechnungen im Rahmen des Projekts AeroGrid



Anwendungsbeispiel – AeroGrid

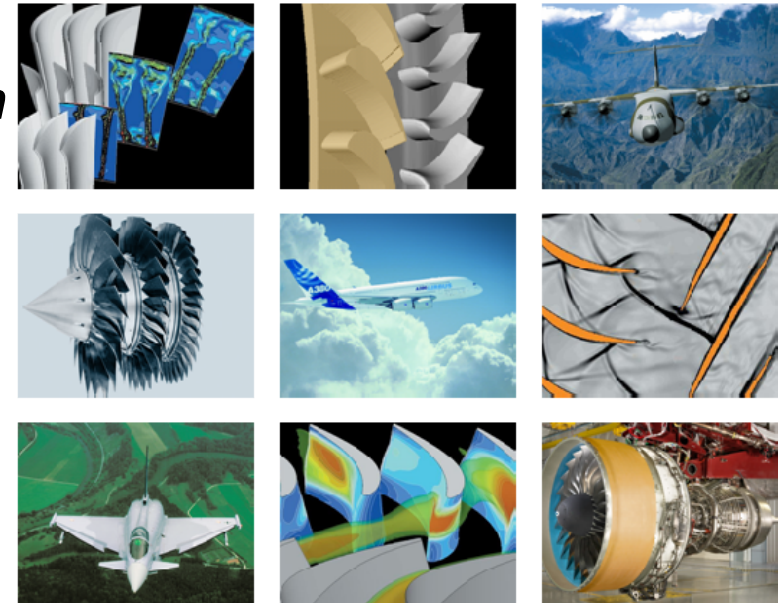
Projektübersicht



Grid-basierte Zusammenarbeit zwischen Industrie, Großforschung und Universitäten in der Luftfahrtforschung

Zielstellung

- Nutzung von Rechen-Ressourcen
- Zusammenarbeit bei der Auslegung von Triebwerkskomponenten



➤ Projektseite: <http://www.aero-grid.de/>



GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Anwendungsbeispiel – AeroGrid

Simulationsautomatisierung von CFD-Rechnungen



Projektorientierte Verwaltung
von Simulationsergebnissen

Automatisierung der
Simulationsdurchführung

Visualisierung von
Ein- und Ausgabedaten

Integration mit
Grid-Schnittstellen

The screenshot shows the DataFinder web interface. On the left is a file system tree view showing a directory structure for simulation results. On the right is a table listing data types and their attributes. Below the table is a 'DataFinder Attributes' section with a small table. At the bottom, there is a log window showing search results for 'DataFinder Type == Run'.

Name	DF Data Type	Content Type	Length	Modified
trace	Project	http/unix-directory	(Collection)	12. Feb. 16:52
MTU-12	User	http/unix-directory	(Collection)	17. Feb. 11:37
Müller	Project	http/unix-directory	(Collection)	17. Feb. 11:41
Verdichter	Project	http/unix-directory	(Collection)	17. Feb. 11:41
BC_Fourier	Run	http/unix-directory	(Collection)	17. Feb. 11:51
Monitoring	Monitoring	http/unix-directory	(Collection)	17. Feb. 11:52
SystemInfo	SystemInfo	http/unix-directory	(Collection)	17. Feb. 11:52
TRACE	TRACE	http/unix-directory	(Collection)	17. Feb. 11:51
Input	Input	http/unix-directory	(Collection)	17. Feb. 11:51
BALANCE_1PROC	TRACE-Info	application/octet-stream	15 Byte	17. Feb. 11:51
stcf10_1.cgns	CGNS	application/octet-stream	135.289 MByte	17. Feb. 11:51
TRACE_control.input	TRACE-Parse	application/octet-stream	3.334 KByte	17. Feb. 11:51
TRACE_entry.input	TRACE-Entry	application/octet-stream	898 Byte	17. Feb. 11:51
TRACE_exit.input	TRACE-Exit	application/octet-stream	25 Byte	17. Feb. 11:52
TRACE_S2.input	TRACE-S2	application/octet-stream	174 Byte	17. Feb. 11:51
Output	Output			
BC_Giles1	Run			17. Feb. 11:55
BC_Giles2	Run			17. Feb. 11:44
BC_Riemann	Run			17. Feb. 11:58

Name	Value
1 CPUs	5
2 DataFinderType	TRACE
3 Version	6.3.72

Log Search Results Script Output

11:44:45: INFO: Search results for [DataFinder Type == Run]:

- /datafinder/data/trace/Müller/Verdichter/BC_Fourier
- /datafinder/data/trace/Müller/Verdichter/BC_Riemann
- /datafinder/data/trace/Müller/Verdichter/BC_Giles1
- /datafinder/data/trace/Müller/Verdichter/BC_Giles2

4 item(s) found.



Vielen Dank für ihre Aufmerksamkeit!

Weiterführende Links

DataFinder: <http://sourceforge.net/projects/datafinder/>

Catacomb: <http://catacomb.tigris.org/>

AeroGrid: <http://www.aero-grid.de/>

Kontakt

Email: Tobias.Schlauch@dlr.de

