



Architektur eines Managementsystems für wissenschaftliche Daten

Roland Betz, Rolf Hempel, Volker Padur
Simulation and Software Technology (SISTEC)
<http://www.sistec.dlr.de>

Workshop „Wege aus dem Daten-Chaos“,
DLR Köln-Porz, 16. Mai 2002



Übersicht

DLR-Anforderungen an ein Datenmanagementsystem

Verfügbare Lösungsansätze

SISTEC-Konzept und seine Realisierung

Zusammenfassung



Allgemeine Lage (im DLR):

Jeder organisiert die Verwaltung seiner Daten selbst, zumeist in einem (mehr oder weniger geschickt) strukturierten Directory-Baum.

Probleme:

- Struktur für andere Benutzer nicht offensichtlich
- Gemeinsames Arbeiten an Projektdaten schwierig (Kopien)
- Wiederfinden älterer Ergebnisse dem Zufall überlassen
- Ausscheiden von Mitarbeitern \Rightarrow Datenchaos

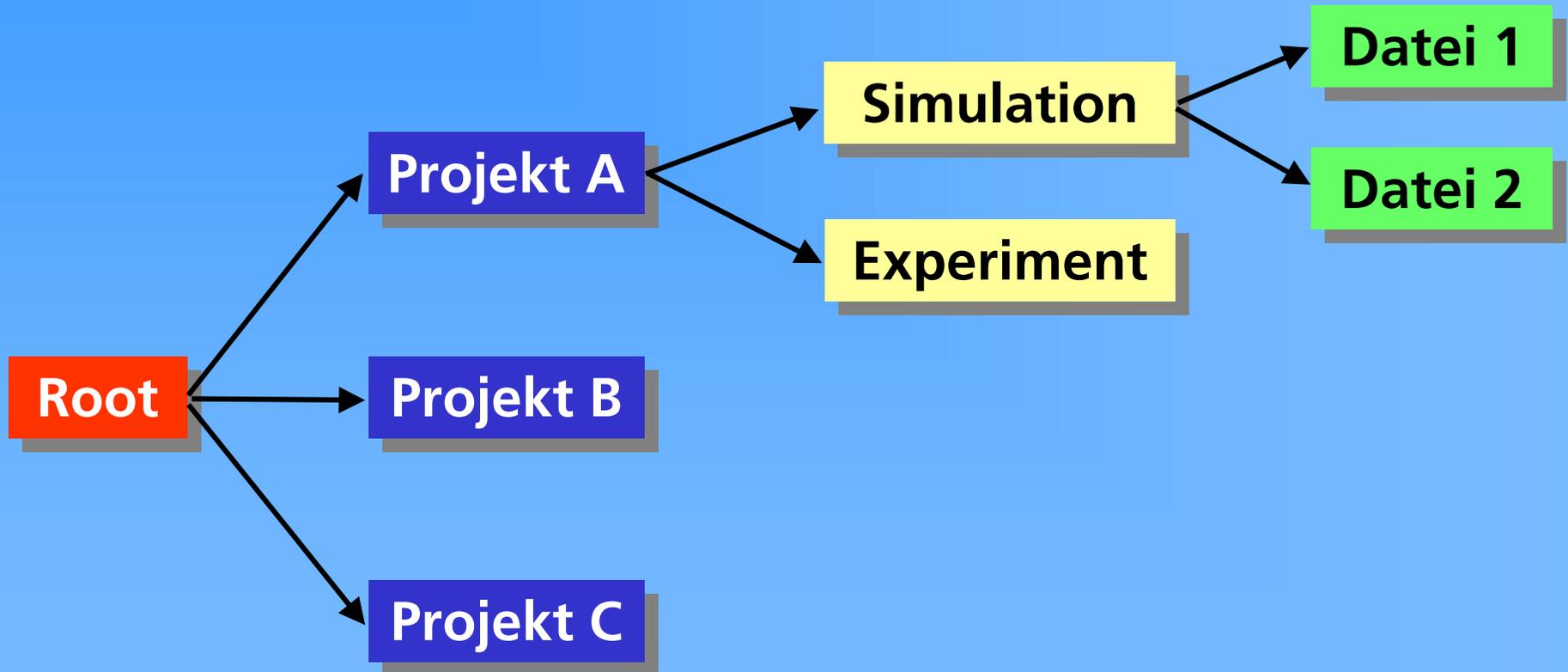
Datenmanagementsystem: Anforderungen

Geforderte Grundfunktionalität: (z.B. im Institut AS)

- Freie hierarchische Strukturierung der Projektdaten
- Versionsverwaltung
- Vermeidung von mehrfacher Datenhaltung
- Kollaboratives Arbeiten
- Unterstützung von Meta-Daten

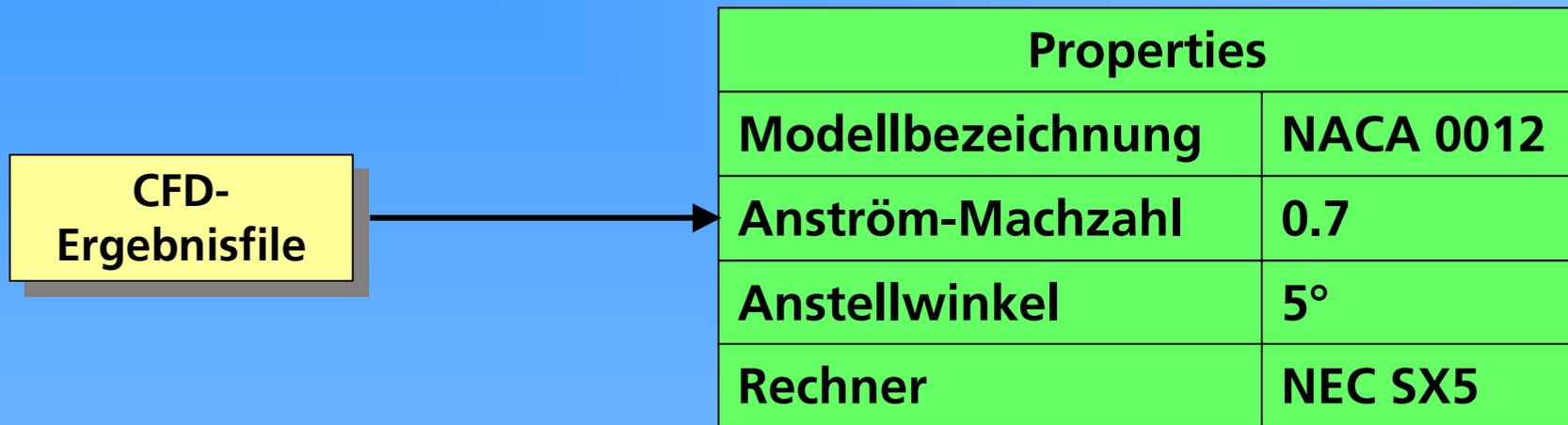
Hierarchische Strukturierung:

Frei wählbare Verzeichnisstruktur (vgl.: lokales Dateisystem)



Meta-Daten:

Zuordnung von beliebigen Attributen (Properties) zu einer Datei:

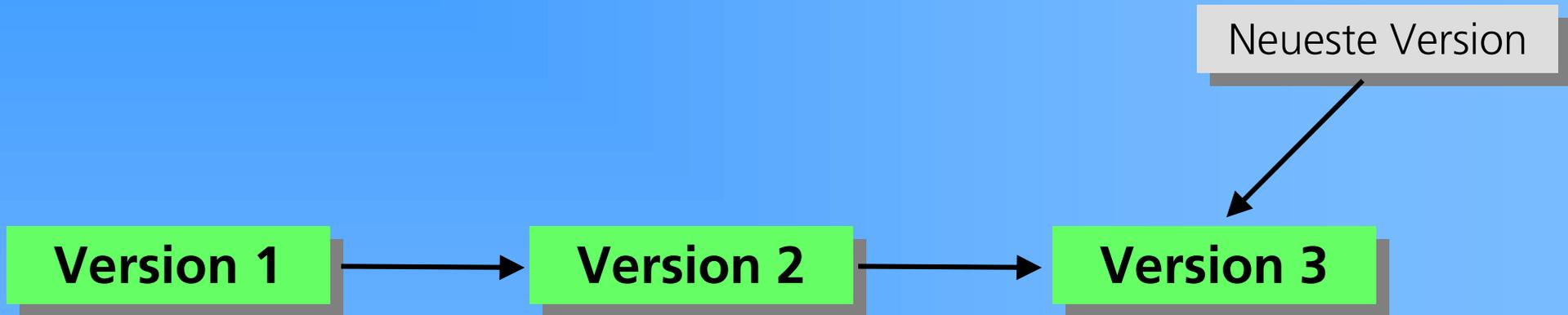


Wichtig:

- Schnelle Suche nach Files mit bestimmten Attributen
- Liste der Attribute jederzeit erweiterbar

Versionsverwaltung:

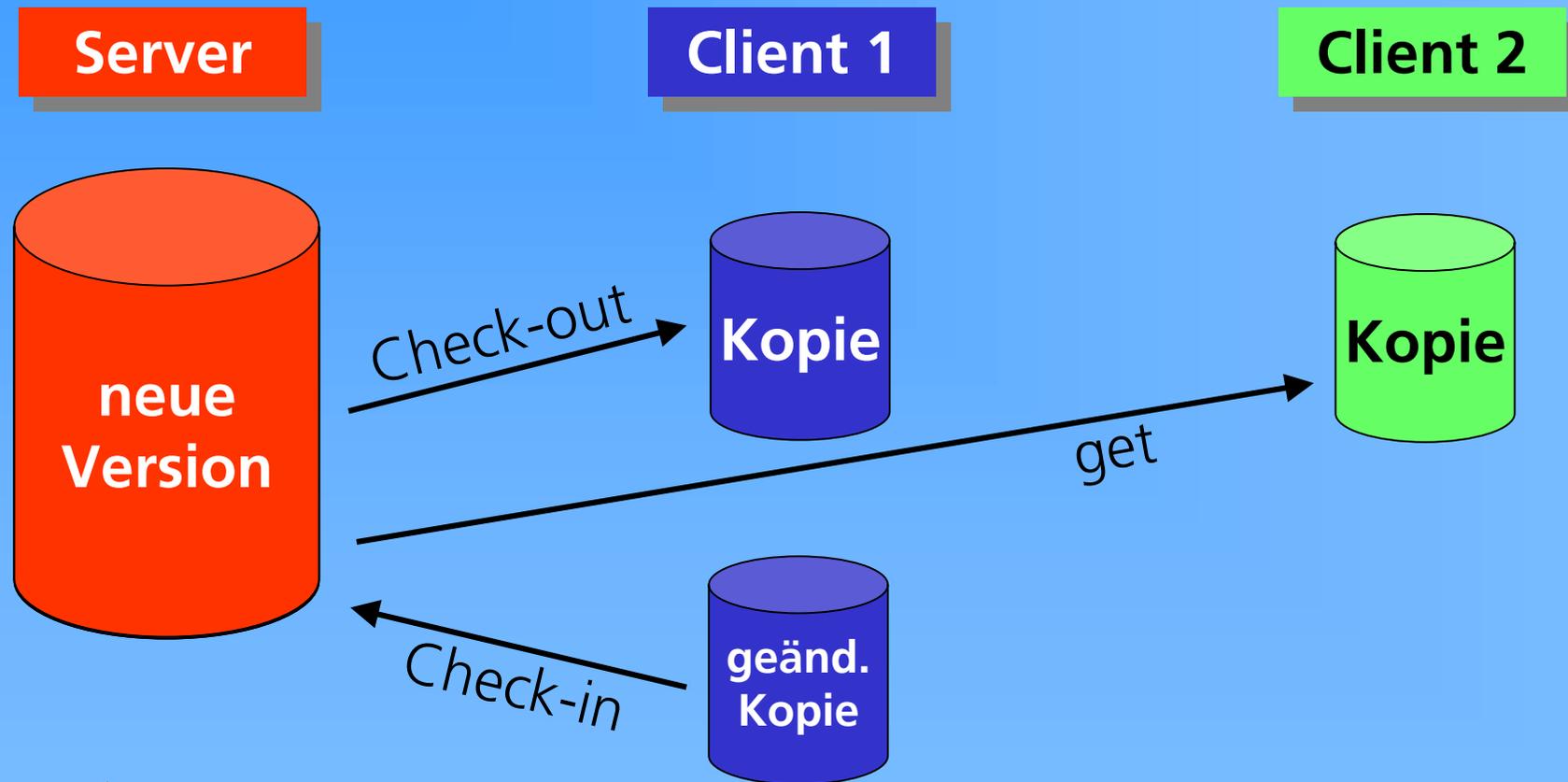
Der Datenserver speichert eine Historie der Projektdaten.



Daten werden erst auf Anfrage gelöscht.

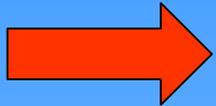
Kollaboratives Arbeiten I:

Synchronisation von Änderungen verschiedener Benutzer: (vgl. cvs)



Kollaboratives Arbeiten II:

Benutzer kündigt mit „check-out“ eine bevorstehende Änderung an.



Datei ist für andere Benutzer schreibgeschützt.

Zugriffsberechtigung über Access Control List (ACL):

ACLs enthalten eine Liste von Benutzern mit Schreib-/Leserechten.

Übersicht

DLR-Anforderungen an ein Datenmanagementsystem

Verfügbare Lösungsansätze

SISTEC-Konzept und seine Realisierung

Zusammenfassung

Gibt es eine fertige Software-Lösung?

Naheliegende Annahme:

Wir sind nicht die einzigen mit diesem Problem

Im Rahmen der TENT-Entwicklung bei SISTEC:

Suche nach verfügbaren Produkten

Vorhandene Optionen (Auswahl):

- PDM-Systeme (z.B. eMatrix, Windchill)
- Co-operative Work Systems (z.B. BSCW)

Beobachtungen:

- Problemstellungen sind in rascher Entwicklung (z.B. kollaboratives Arbeiten)
- Lösungen für „unser“ Problem nicht vorhanden
- Aber: wichtige Bausteine verwendbar
- Offene Plattform wichtiger als fertige Lösung

Konzept bei SISTEC:

Prinzipien:

- Nur neu entwickeln, was es nicht fertig gibt.
Ansonsten: Integration vorhandener Tools.
- Beachtung von Standards, soweit möglich:
 - XML für Metadaten / Konfigurationsdaten
 - WebDAV für Web File Service
 - Etablierte SW für File / Datenbank - Services
- System stellt eine offene Plattform dar:
Andere Software kann einfach (standardisiert) auf die Daten zugreifen.

Kurzeinführung des WebDAV-Protokolls:

- Protokollerweiterung zu http für „**Distributed Authoring & Versioning**“, basierend auf XML
- Realisiert ein **Dateisystem über das Internet**
- WebDAV unterstützt:
 - Verzeichnisse (Collections)
 - Metadaten (Properties)
 - Synchronisation von Änderungen (Locking)
 - Versionen (DeltaV)
 - Zugriffskontrolle (ACL)
 - Suchen (DASL)
 - Erweiterte Verzeichnisse (Links, Sortierungen)

Vorteile des WebDAV-Protokolls:

Standard mit großem Entwicklungspotential:

- Aufbauend auf bewährter **http-Technologie**
- **Sicherheits**mechanismen von http sind nutzbar
- Bereits integriert in **Microsoft Office 2000**, sowie einige Adobe-Produkte

Flexible Kombination von Client- und Server-Software:

- Standardisierte Schnittstelle zwischen Client und Server
- Client und Server sind strikt voneinander getrennt
- Client- und Serverapplikationen können beliebig kombiniert werden

Übersicht

DLR-Anforderungen an ein Datenmanagementsystem

Verfügbare Lösungsansätze

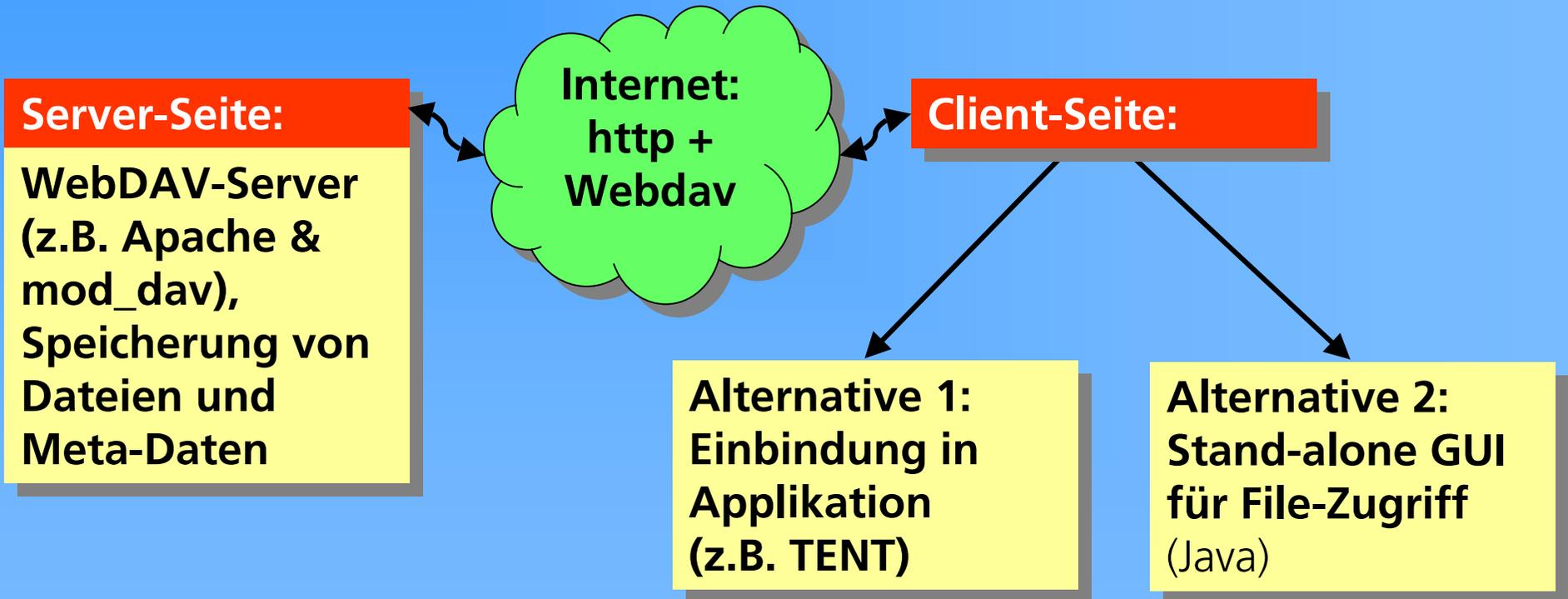
SISTEC-Konzept und seine Realisierung

Zusammenfassung



Architekturübersicht:

Konzeption als Client/Server - Anwendung:



Lösungsansatz (Server-Seite) I:

Web-Server:

- **Apache** (open source)
- WebDAV-Unterstützung durch **mod_dav** - Modul (erlaubt Speichern von Files und Properties)

Erster Schritt: Organisation der Daten in Filesystem

Sicherheit: Secure Socket Layer (**HTTPS**)

Skalierbarkeit:

- Apache kann mehrere File-Server bedienen
- Mehrere Apache-Server können parallel eingesetzt werden

Erweiterungen:

- Zugriffskontrolle ist noch nicht in mod_dav implementiert (bald zu erwarten)
- Bis dahin auf Client-Seite oder selbst in mod_dav implementieren

Lösungsansatz (Server-Seite) II:

Zukunft: - **XML-Datenbank** statt Filesystem

- Kommerzielle Lösungen: Infonyte (FhG-Ausgründung), Tamino (Software AG)
- Andere mögliche Server-Lösung: Sharepoint (Microsoft)

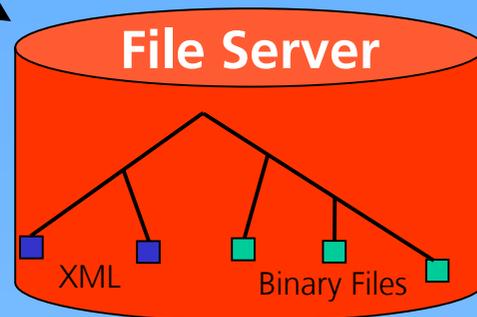
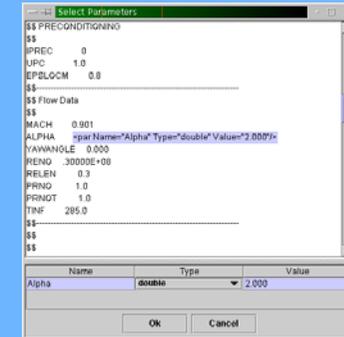
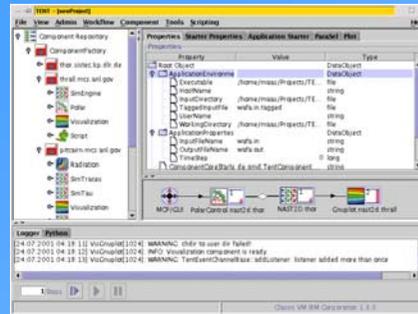
Vorteil: Datenfiles werden nicht nach Properties durchsucht
⇒ **bessere Skalierbarkeit**

Datenzugriff (Client-Seite):

Andere Anwendung

Stand-alone GUI

TENT-GUI



Normaler File-Zugriff
auch weiterhin über
Standard-Betriebssystem

Option 1: Stand-alone GUI

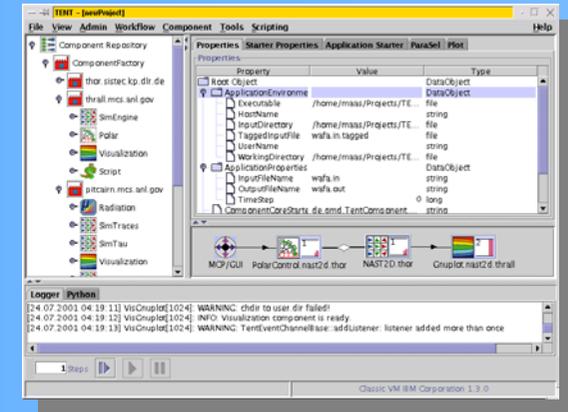


Eine Java-Applikation erlaubt den Zugriff auf den Datenbestand von jedem Arbeitsplatzrechner aus.

Damit steht folgende Grundfunktionalität zur Verfügung:

- Einsehen und Ändern von Dateiattributen (Properties)
- Definition von Verweisen zwischen Dateien
- Durchsuchen des Datenbestandes
- Laden von Dateien vom Server
- Versionsverwaltung
- Synchronisation von Änderungen
- Zugriffsbeschränkungen

Option 2: Zugriff über TENT-GUI

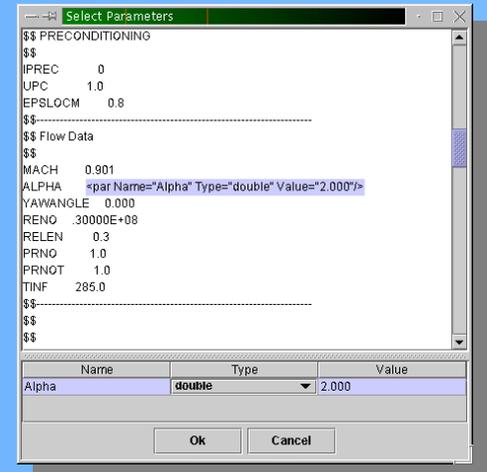


Komfortabelste Art des Zugangs für Simulationsrechnungen:

- Direkter Datenzugriff von der TENT-GUI aus
- Automatischer Transfer der Simulationsdateien von/zu dem Compute Server
- Alle Simulationsparameter zu einem Simulationsergebnis werden als Metadaten gespeichert
- Workflow-Konfigurationen werden automatisch gespeichert und sind später wieder abrufbar

Option 3: Zugriff durch andere Anwendung

Die Programmierschnittstelle zum Datenbestand ist nicht an TENT gebunden.



Volle Grundfunktionalität über Standard-Schnittstellen

Strukturierte Anwendungsdaten abgelegt mit XML

Frei wählbares Datenmodell

Entwicklungsstand:

Eine erste Lösung wurde im DLR-Projekt **IMENS** realisiert (Option 2):

Hierarchische Projektstruktur: Organisation von Workflows und zugehörigen Files

Datenzugriffskontrolle: In TENT client-seitig realisiert

Gegenwärtig in Entwicklung: Stand-alone GUI (TENT-unabhängig)
Erste Version im August 2002

Übersicht

DLR-Anforderungen an ein Datenmanagementsystem

Verfügbare Lösungsansätze

SISTEC-Konzept und seine Realisierung

Zusammenfassung



Zusammenfassung:

Internetstandards: mächtige Bausteine für Datenmanagement

Grundfunktionalität bereitgestellt durch:

- **Open-Source** – Software (z.B. Apache-Server)
- **Standardbausteine** (z.B. XML-Datenbanken mit WebDAV)

Begrenzte Eigenentwicklungen:

- Einbindung in vorhandene Client-Umgebungen (z.B. TENT)
- Client-GUI zum freien Datenzugriff

Bei SISTEC:

- Prototyp realisiert im IMENS-Projekt (in TENT)
- TENT-unabhängiger Client in wenigen Monaten verfügbar

