

# Verteiltes Rechnen mit UNICORE - Gegenwart und Zukunft

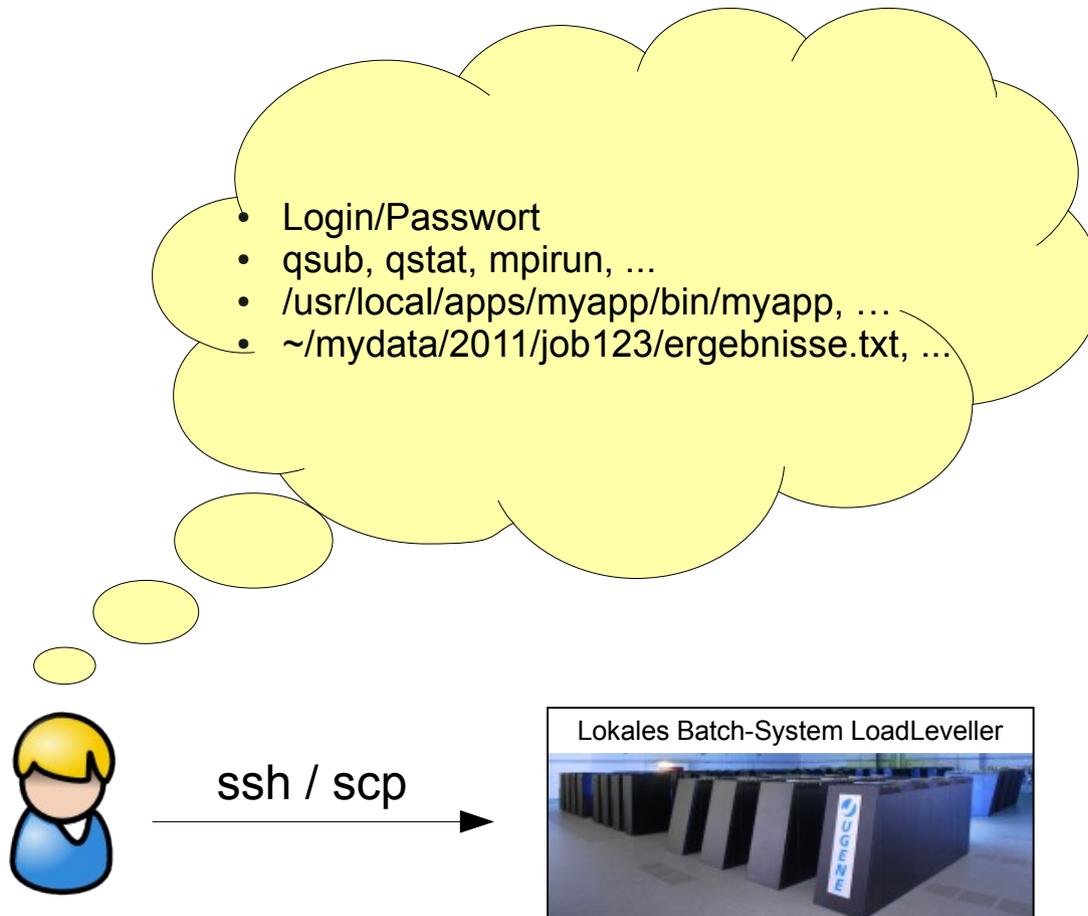
Dr. Bernd Schuller  
Federated Systems and Data division  
Jülich Supercomputer Centre  
Forschungszentrum Jülich GmbH

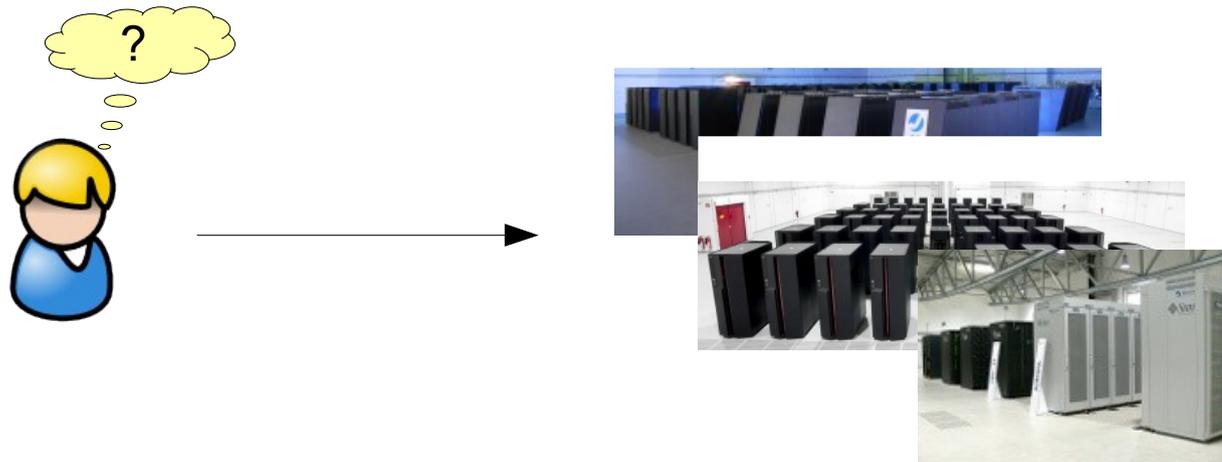
23. Juni 2011  
ZIH, Dresden

# Überblick

- Verteilte Ressourcen
  - Cluster, Clouds, Grid-Infrastrukturen ...
- Was ist UNICORE?
  - Jobs und Daten, Anwendungen, Workflows, Sicherheit
  - Anwendungsbeispiele
- Aktuelle Entwicklungen
  - Daten, Metadaten und Datentransfer
  - Virtuelle Systeme und Clouds
  - Kollaboratives Arbeiten

# Verteiltes Rechnen ...

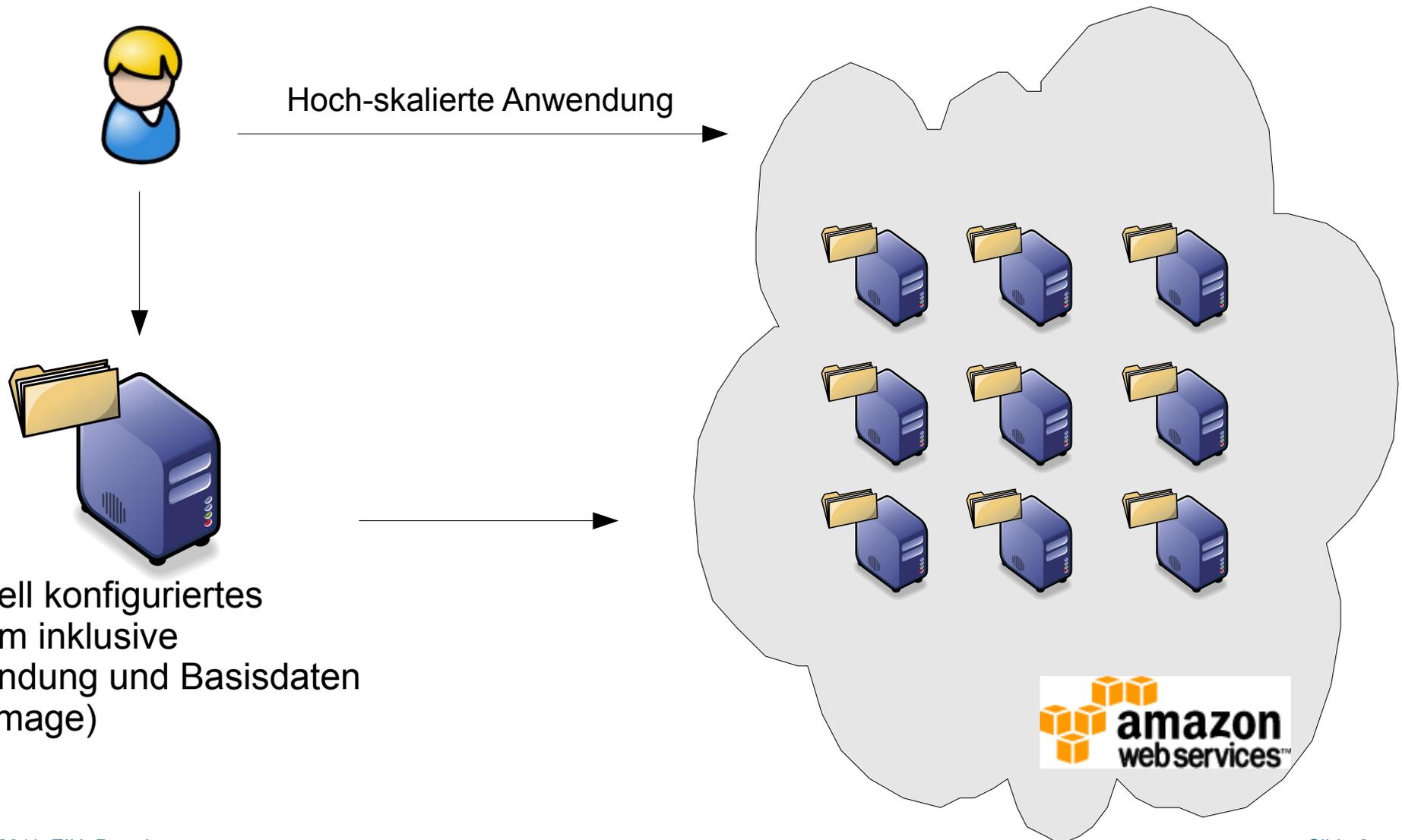




Wie kann ich ...

- ... mehrere, verschiedene Systeme möglichst einfach nutzen,
- ... meine Eingabedaten und Ergebnisse verwalten?
- ... und dies auch System-übergreifend? Workflows?
- Dies war die ursprüngliche Motivation für UNICORE (1997)

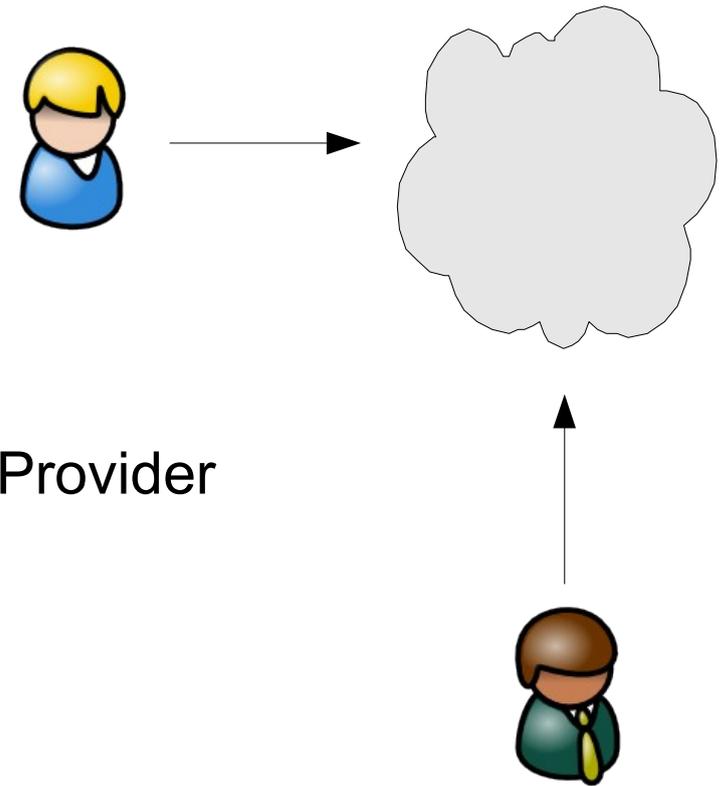
# Anwender-spezifische Umgebungen



## ... nicht schlecht, aber

Wie kann ich als Anwender ...

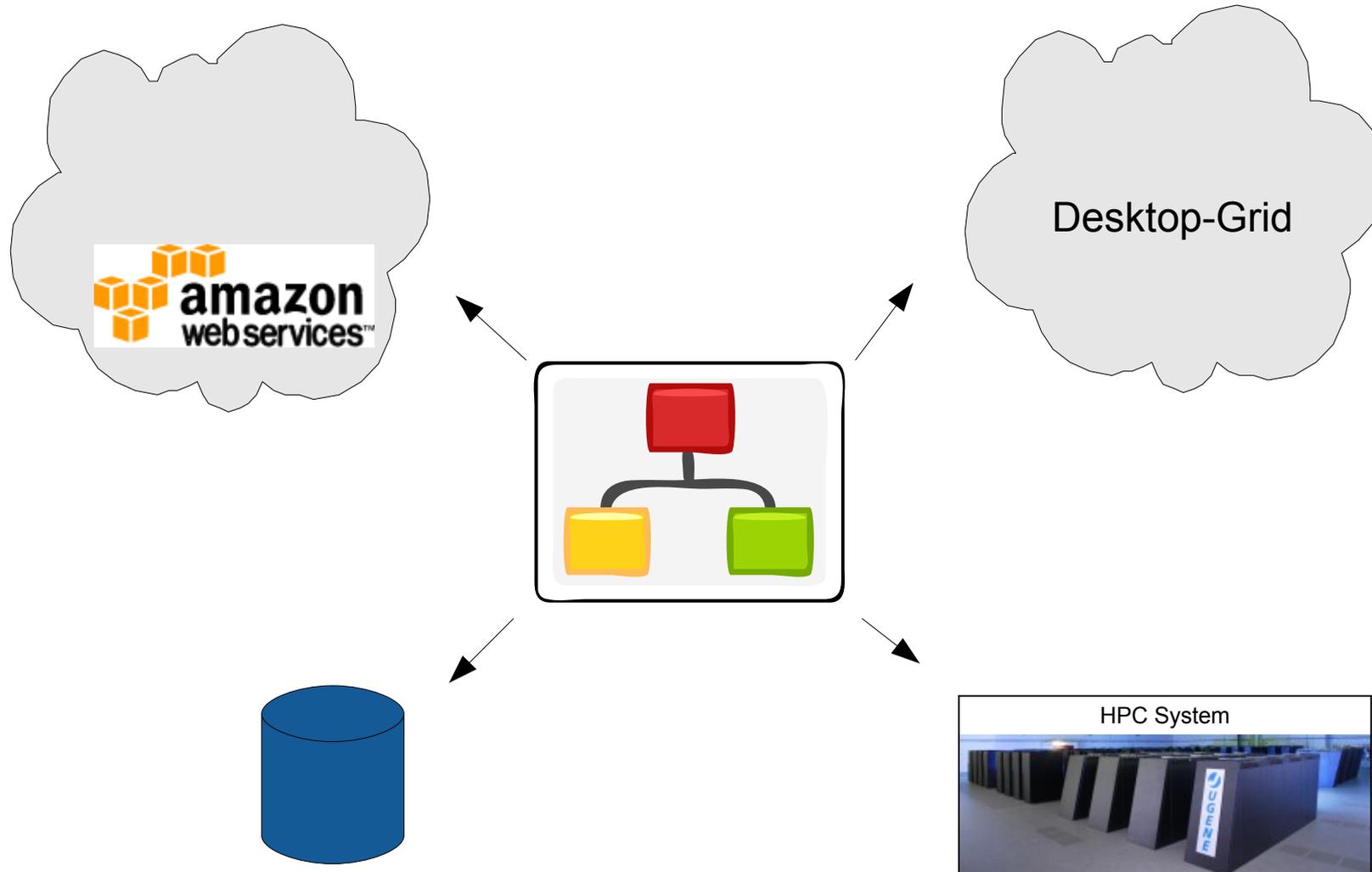
- ... meine Eingabedaten und Ergebnisse verwalten?
- ... meine Anwendungen unabhängig vom Provider gestalten?



Als möglicher Provider solcher Dienste ...

- ... meinen Kunden die Verwendung möglichst einfach machen?
- ... mein Angebot unabhängig von der Technologie gestalten?
- ... wie geht z.B. Sicherheit, Accounting, Logging, Monitoring?

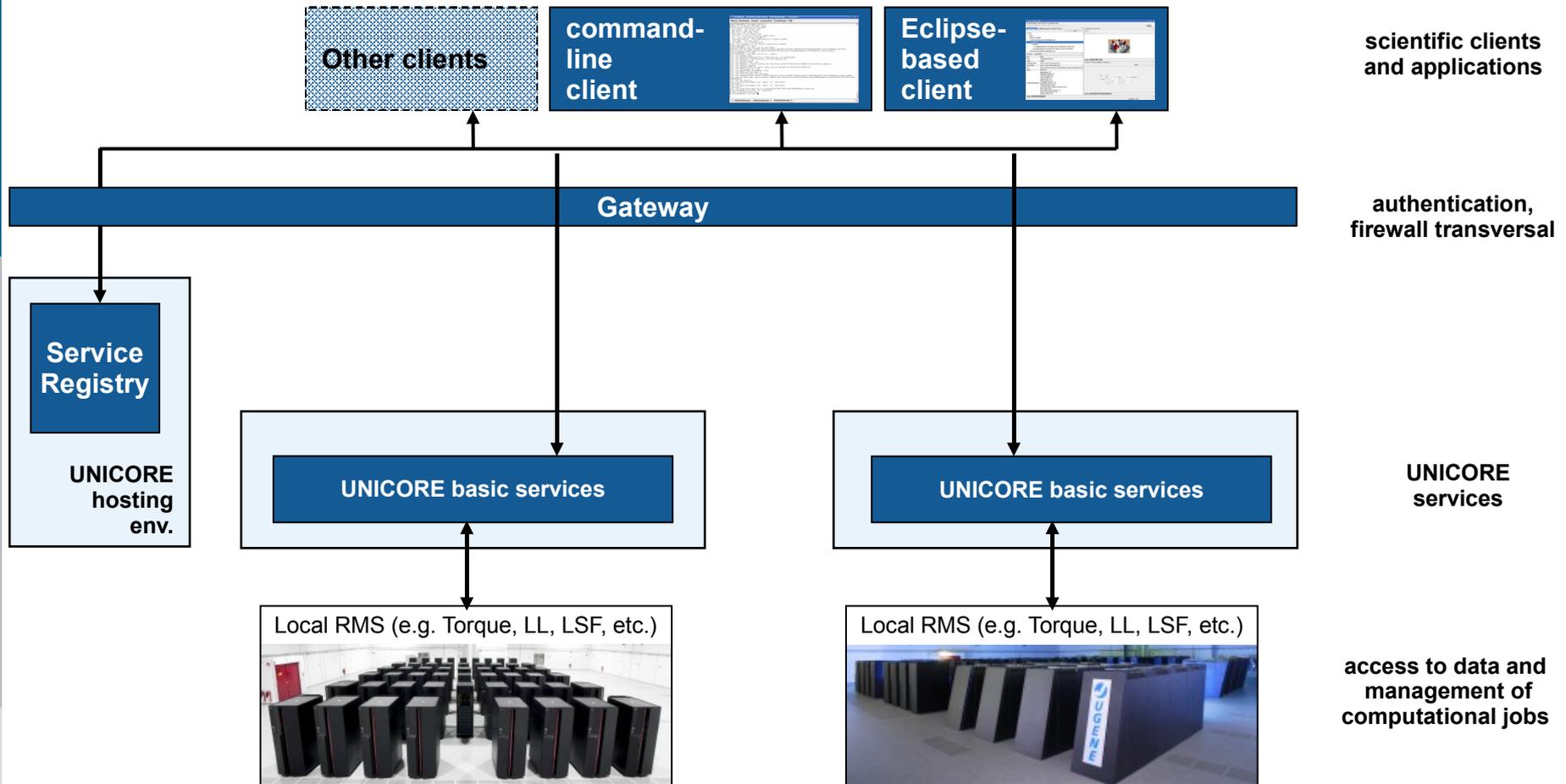
## ... und was ist mit Workflows?



# UNICORE

... ist eine integrierte Middleware für sicheren Zugang zu Rechenressourcen und Daten

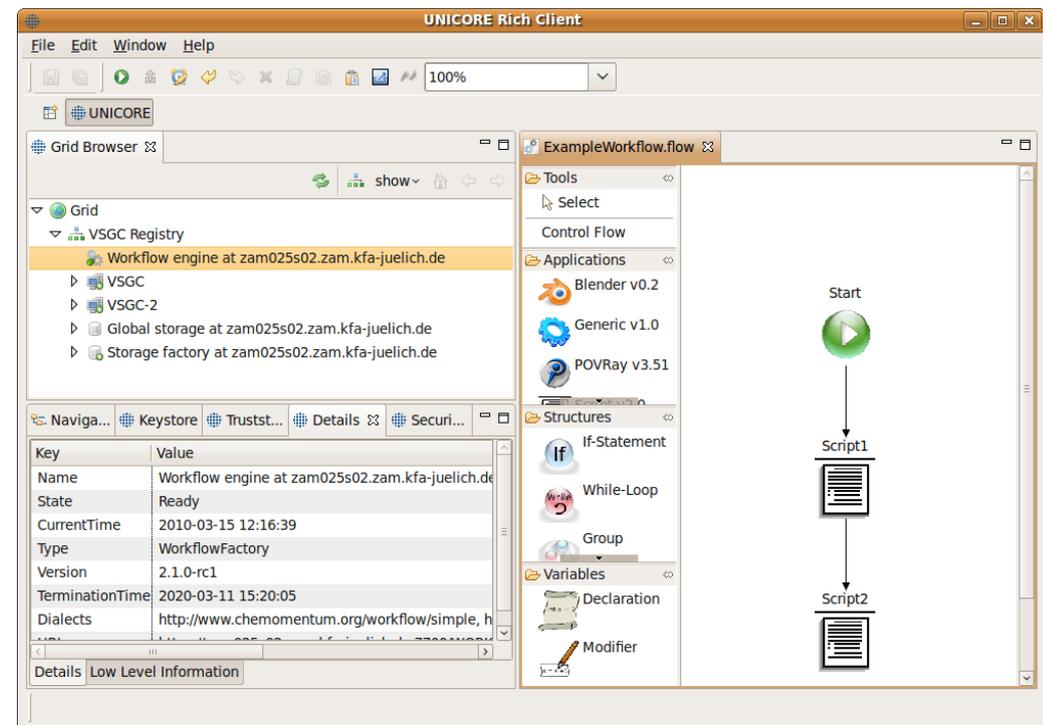
- Einfache Integration in existierende Umgebungen
- Implementiert in Java und Perl
  - UNIXe, MacOS X, Windows
- Unterstützt Torque, LoadLeveller, LSF, SGE, ...
- Effizienter Support, regelmässige Releases
- **Open source unter BSD Lizenz, siehe <http://www.unicore.eu>**



## Clients

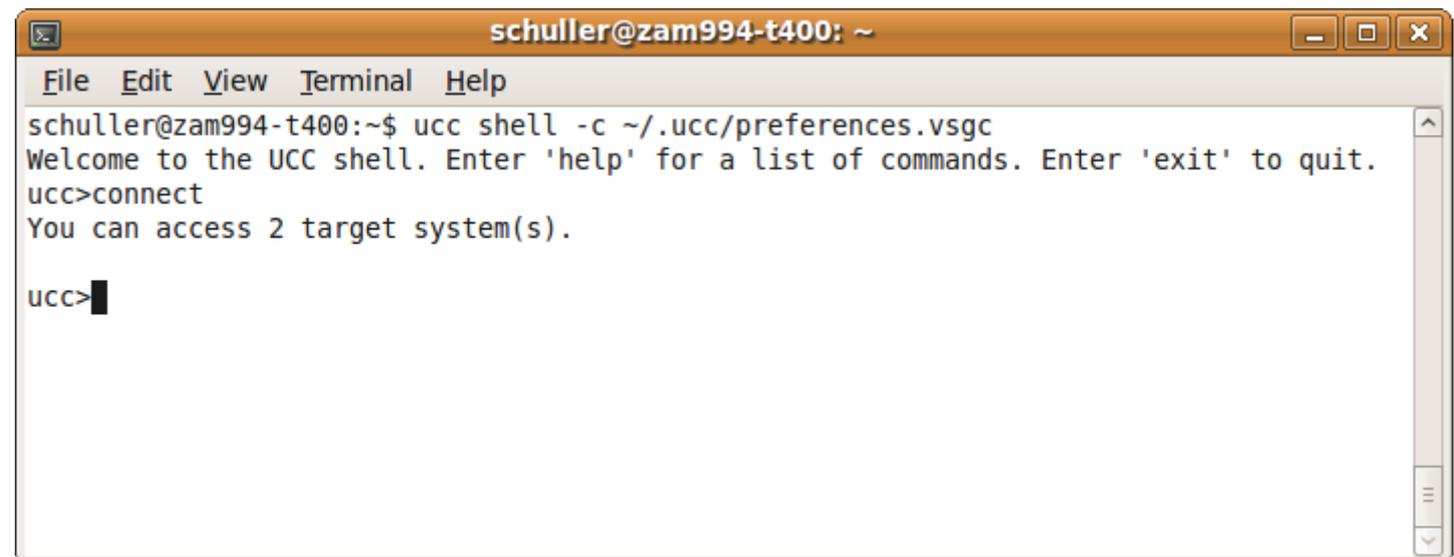
## Der URC – UNICORE Rich client

- Erstellen und Verwalten von Jobs und Workflows
- Integriertes Datenmanagement
- Basiert auf dem Eclipse Framework
- Erweiterbar



## Der UCC - UNICORE Commandline Client

- Bietet Zugang zur vollen UNICORE Funktionalität
- „Batch-Modus“ für high-throughput Szenarien
- Erweiterbar durch
  - neue Kommandos
  - selbstprogrammierte Skripten



```
schuller@zam994-t400: ~  
File Edit View Terminal Help  
schuller@zam994-t400:~$ ucc shell -c ~/.ucc/preferences.vsgc  
Welcome to the UCC shell. Enter 'help' for a list of commands. Enter 'exit' to quit.  
ucc>connect  
You can access 2 target system(s).  
ucc>
```

## Sicherheit

## Sicherheit - Überblick

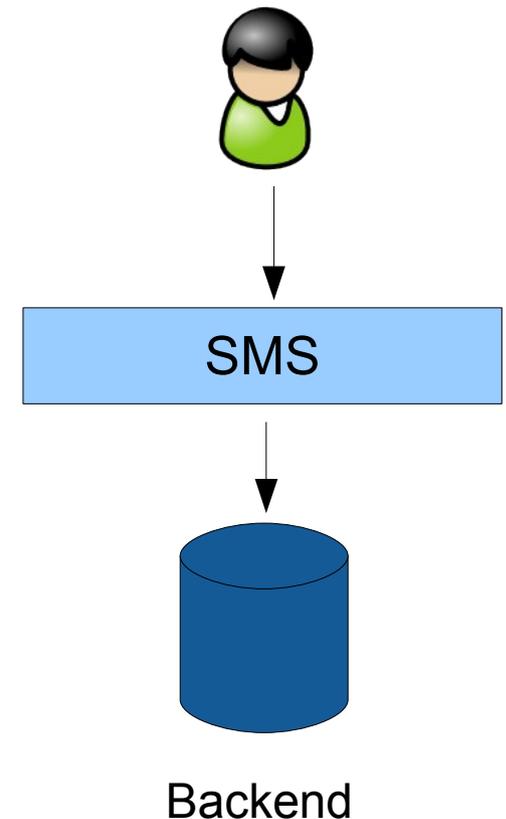
Basiert auf offenen Industrie-Standards

- **X.509** Zertifikate für Benutzer und Server
- **SSL** für die Kommunikation
- Signierte **SAML**-Dokumente (Security assertion markup language)
  - Sichere Delegation
- Verschiedene Quellen für Benutzerattribute zur Autorisierung
- Autorisierung basierend auf **XACML 2.0** Policies
- Erweiterbar, z.B. optionale, eingeschränkte Unterstützung von Proxy-Zertifikaten

# Datenmanagement

## Anbindung von Speichersystemen

- Der UNICORE Storage Management Service („SMS“) ist die zentrale Abstraktion: eine Filesystem-artige Sicht auf Daten
- Typische Operationen
  - mkdir, delete, ls, etc
- Erlaubt das Anstossen von File-Transfers
  - Import/Export von Daten von/zur lokalen Maschine
  - Senden/Empfangen von Daten anderer Server
  - Verschiedene Protokolle möglich

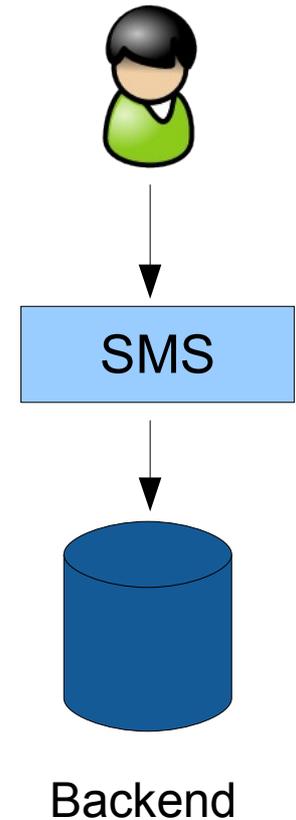


## Existierende SMS-Implementierungen

- Filesystem
- Apache HDFS (Hadoop distributed file system)



- iRODS (Prototyp)



## Integriertes Datenmanagement im UNICORE Rich Client

- Files erzeugen
- Drag & drop vom/zum Desktop
- Copy & paste

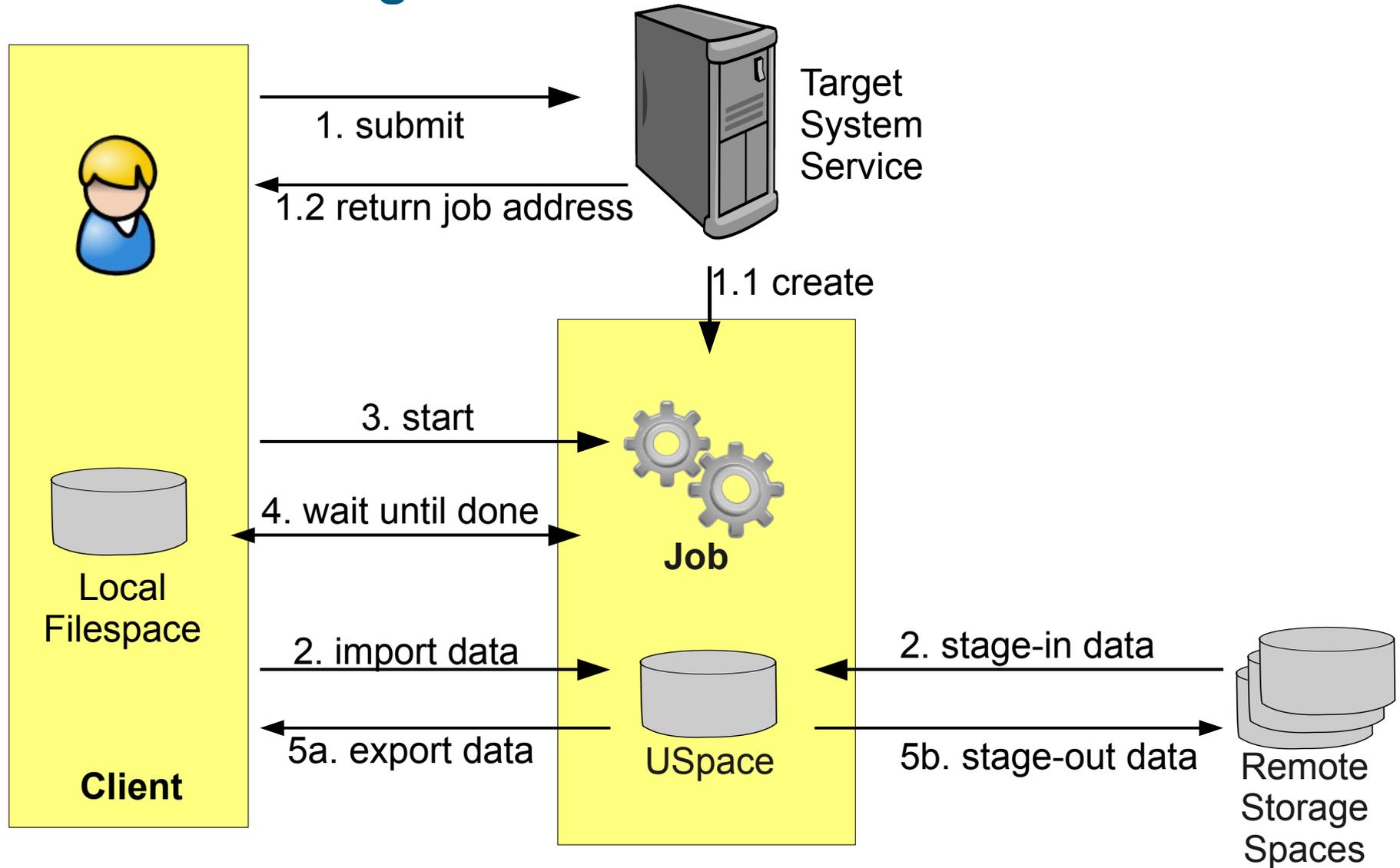
The screenshot displays the UNICORE Rich Client interface. On the left, a 'Grid Browser' window shows a hierarchical file structure under 'VSGC-2\_TargetSystem Home', including files like '.bash\_history', '.bashrc', and '.emacs'. A context menu is open over the browser, listing actions such as 'add Registry', 'create file', 'create directory', and 'destroy'. Below the browser is a 'Navigator' window showing a project structure with folders like 'input files' and 'jobs', and a file named 'Blender2.flow'. The main workspace on the right shows a workflow diagram for '\*Blender2.flow'. The workflow starts with a 'Start' node, followed by a 'Declaration1' node (int x = 3). This leads into an 'Iteration' block containing a 'ForEach1' loop. Inside the loop, there is a 'Modifier1' node (x = x + 1) and a 'Blender1' node (Blender logo). The workflow concludes with a 'Script1' node. A toolbar on the left side of the workspace provides various workflow components like 'Control Flow', 'Applications', 'Structures', and 'Variables'.

## Jobs

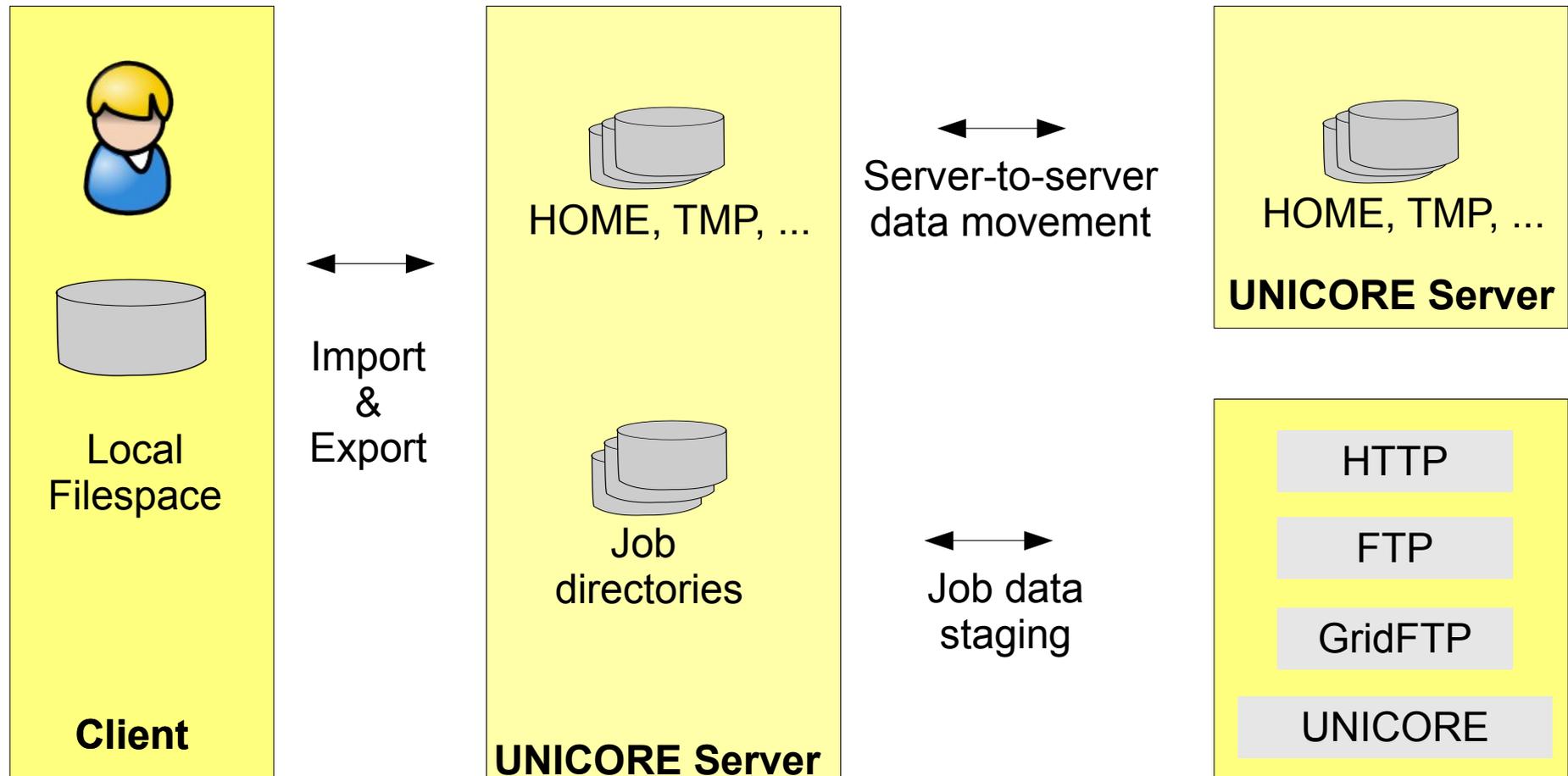
## Anbindung von Compute-Ressourcen

- „Target system service“ (TSS) und „Job management service“ (JMS) als Abstraktion
- Submission (TSS) und Management (JMS) von Einzeljobs
  - Anwendung, oder direkt das Executable
  - Argumente, Environment-Variablen, stderr, stdout, ...
  - Eingabedaten und Resultate
  - Angeforderte Ressourcen (Cores, Speicher, ...)

## Job-Ausführung



## Storages, Jobs und Daten



# Anwendungs-Integration

## Anwendungs-Integration

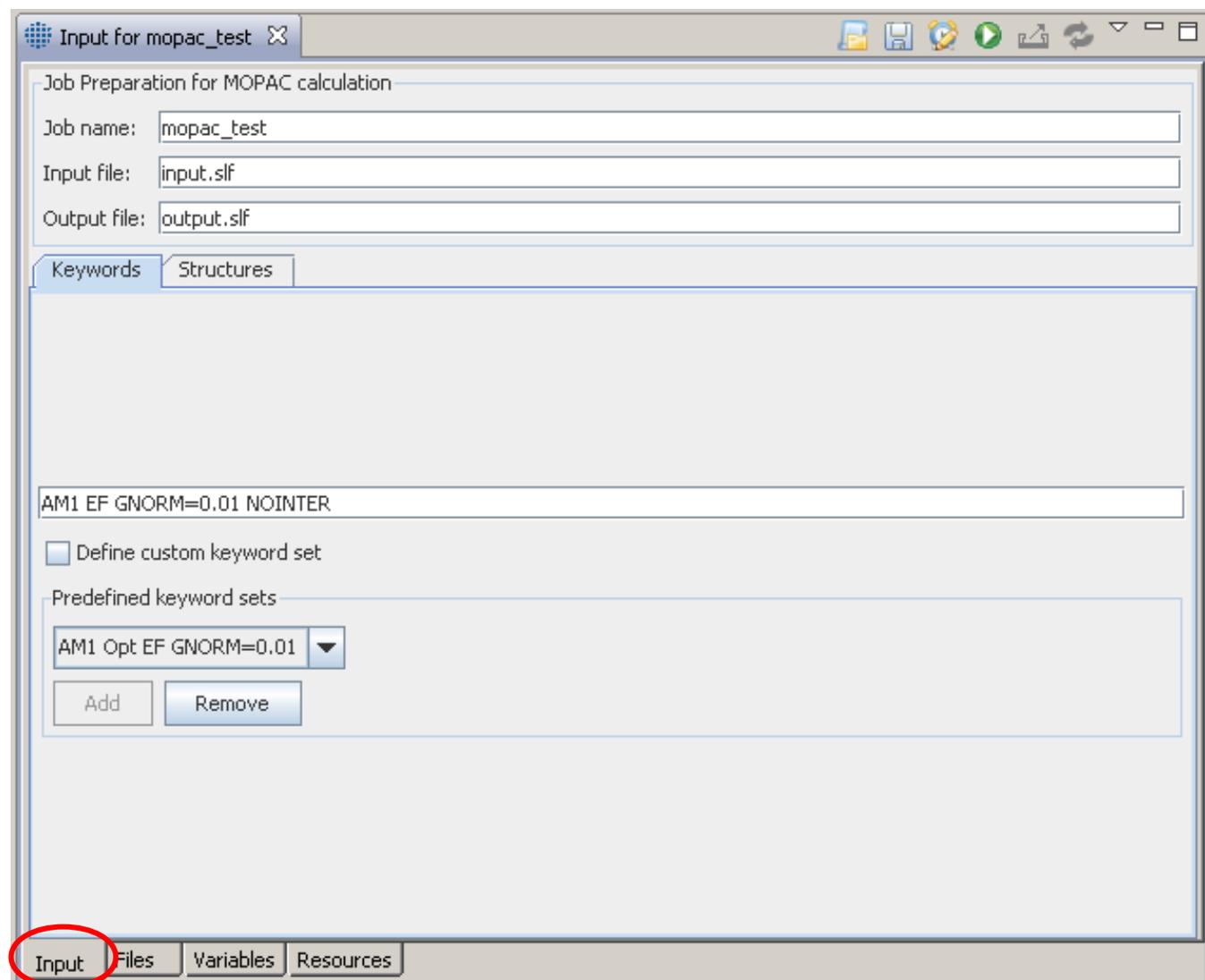
- Anwendung = Software-Paket, das im Batch-Betrieb ausführbar ist
- Installation auf dem Cluster durch einen Administrator
- Konfiguration im UNICORE-Server
- Unterstützung im grafischen Klienten
  - Anwendungs-spezifisches Plugin (ev. Entwicklungsaufwand)
  - „Generic“ Plugin (eingebaut)

## Beispiel: MOPAC

- Molecular Orbital PACkage
- Ein semi-empirisches Quantenchemie-Programm.
- Typische Anwendung: Optimierung von 3D – Molekülstrukturen

## Beispiel: MOPAC - 2

### Anwendungsspezifisches Plugin



## Beispiel: MOPAC - 3

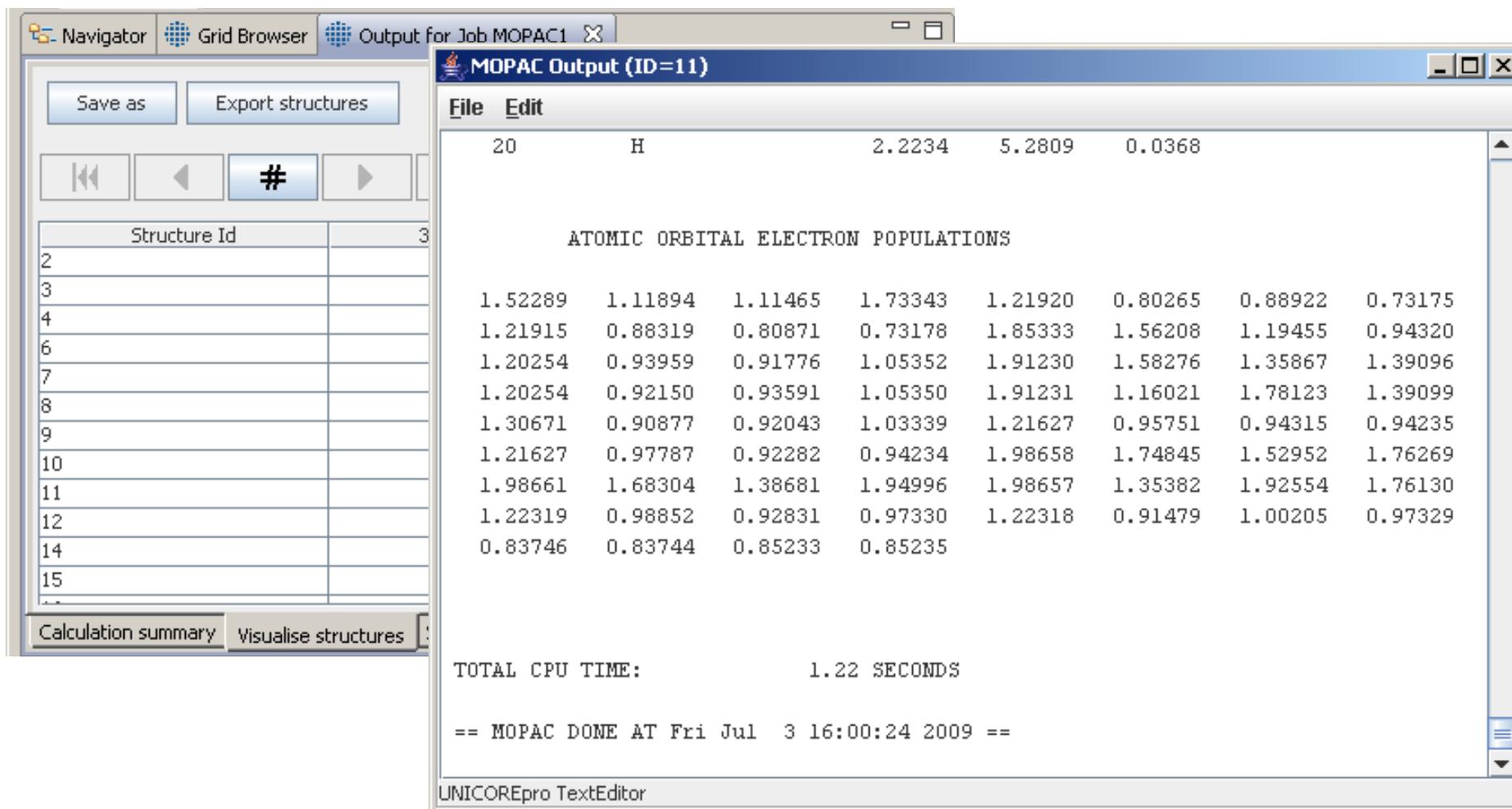
Anzeige der Ergebnisse (Anwendungs-spezifisches Plugin)

The screenshot shows a software interface with a table of results and a 3D molecular structure viewer. The table has three columns: Structure Id, 3D structure, and MOPAC output. Row 11 is highlighted. The 3D viewer shows a ball-and-stick model of a molecule with a central blue nitrogen atom, a yellow sulfur atom, and several green oxygen atoms, along with a grey carbon skeleton and white hydrogen atoms.

Structure Id	3D structure	MOPAC output
2	View	View
3	View	View
4	View	View
6	View	View
7	View	View
8	View	View
9	View	View
10	View	View
11	View	View
12	View	View
14	View	View
15	View	View

## Beispiel: MOPAC - 4

Anzeige der Ergebnisse (Anwendungs-spezifisches Plugin)



Navigator Grid Browser Output for Job MOPAC1

Save as Export structures

Structure Id 3

2

3

4

6

7

8

9

10

11

12

14

15

Calculation summary Visualise structures

MOPAC Output (ID=11)

File Edit

20 H 2.2234 5.2809 0.0368

ATOMIC ORBITAL ELECTRON POPULATIONS

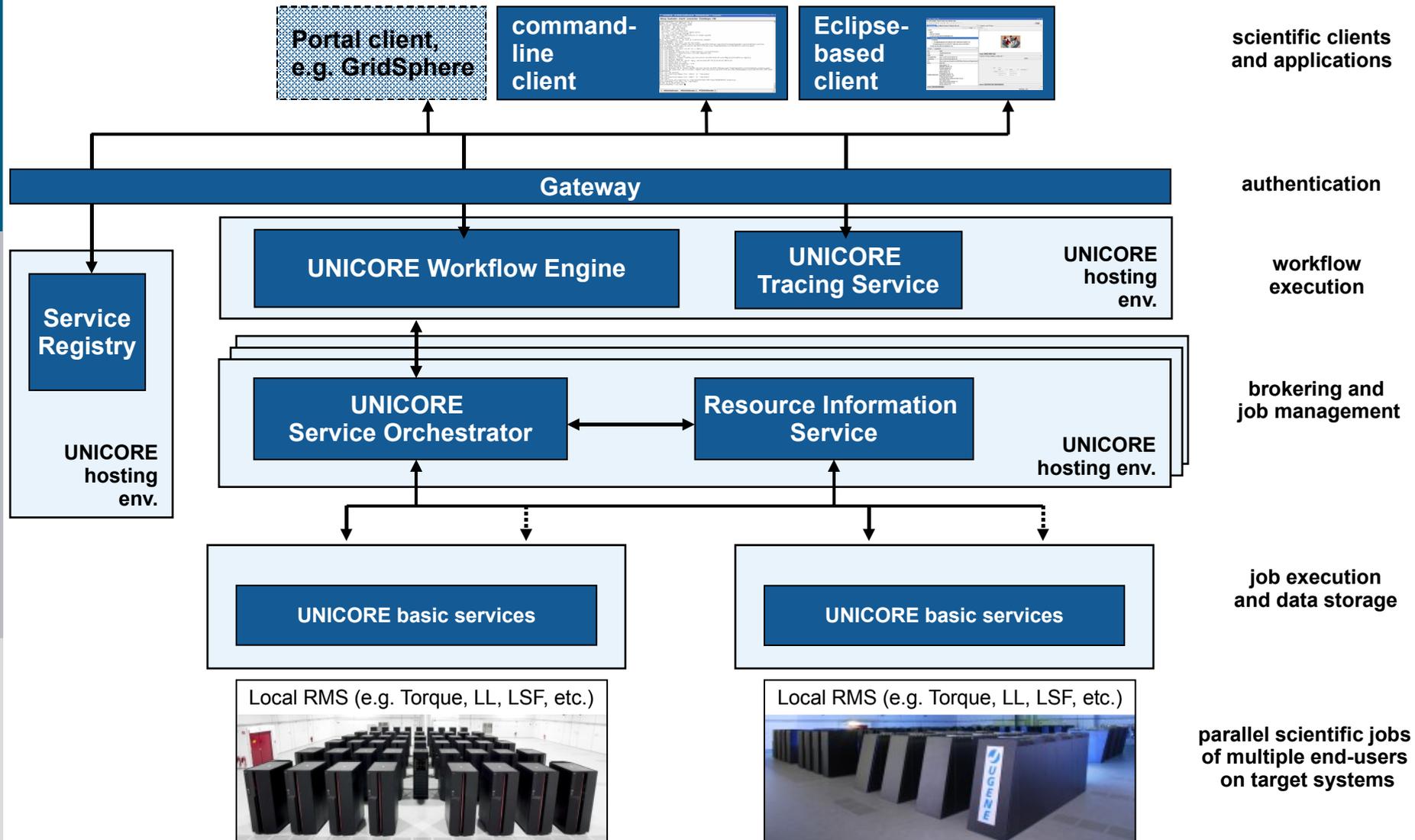
1.52289	1.11894	1.11465	1.73343	1.21920	0.80265	0.88922	0.73175
1.21915	0.88319	0.80871	0.73178	1.85333	1.56208	1.19455	0.94320
1.20254	0.93959	0.91776	1.05352	1.91230	1.58276	1.35867	1.39096
1.20254	0.92150	0.93591	1.05350	1.91231	1.16021	1.78123	1.39099
1.30671	0.90877	0.92043	1.03339	1.21627	0.95751	0.94315	0.94235
1.21627	0.97787	0.92282	0.94234	1.98658	1.74845	1.52952	1.76269
1.98661	1.68304	1.38681	1.94996	1.98657	1.35382	1.92554	1.76130
1.22319	0.98852	0.92831	0.97330	1.22318	0.91479	1.00205	0.97329
0.83746	0.83744	0.85233	0.85235				

TOTAL CPU TIME: 1.22 SECONDS

== MOPAC DONE AT Fri Jul 3 16:00:24 2009 ==

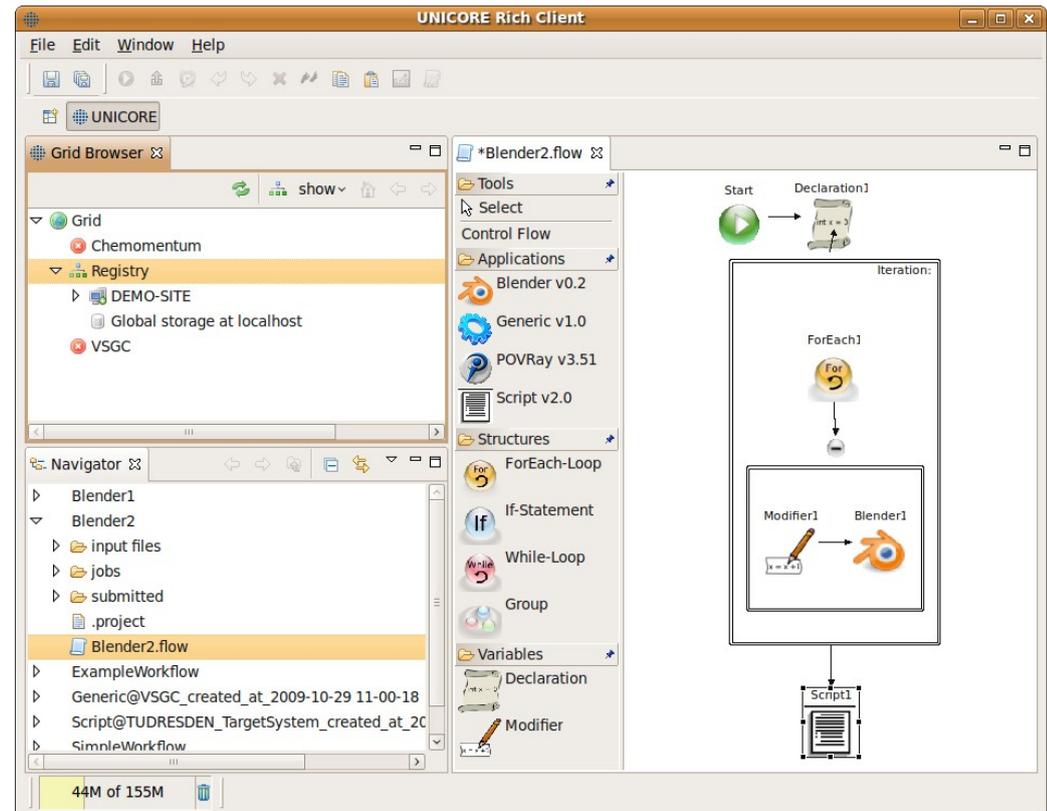
UNICOREpro TextEditor

## Workflows



## Workflow features

- Simple graphs (DAGs)
- Workflow variables
- Loops and control constructs
  - while, for-each, if-else
- Conditions
  - Exit code, file existence, file size, workflow variables
- Clients
  - UNICORE Rich client
  - Commandline client

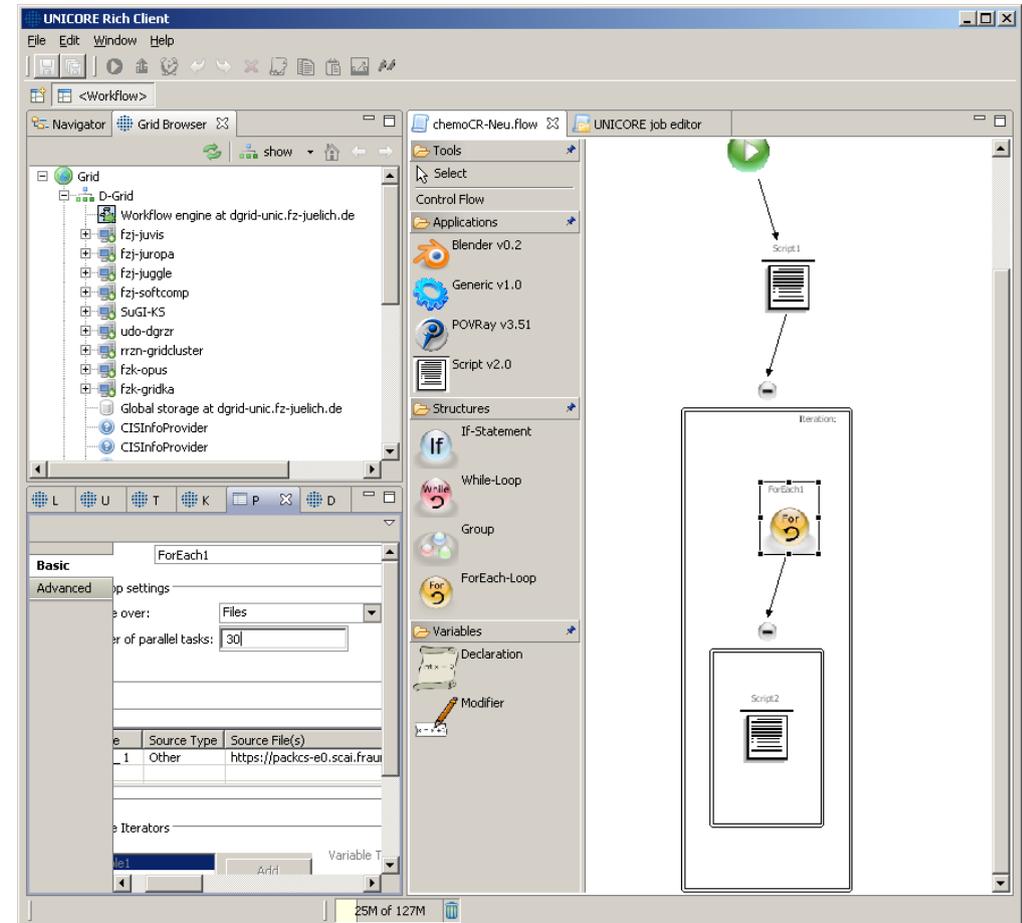


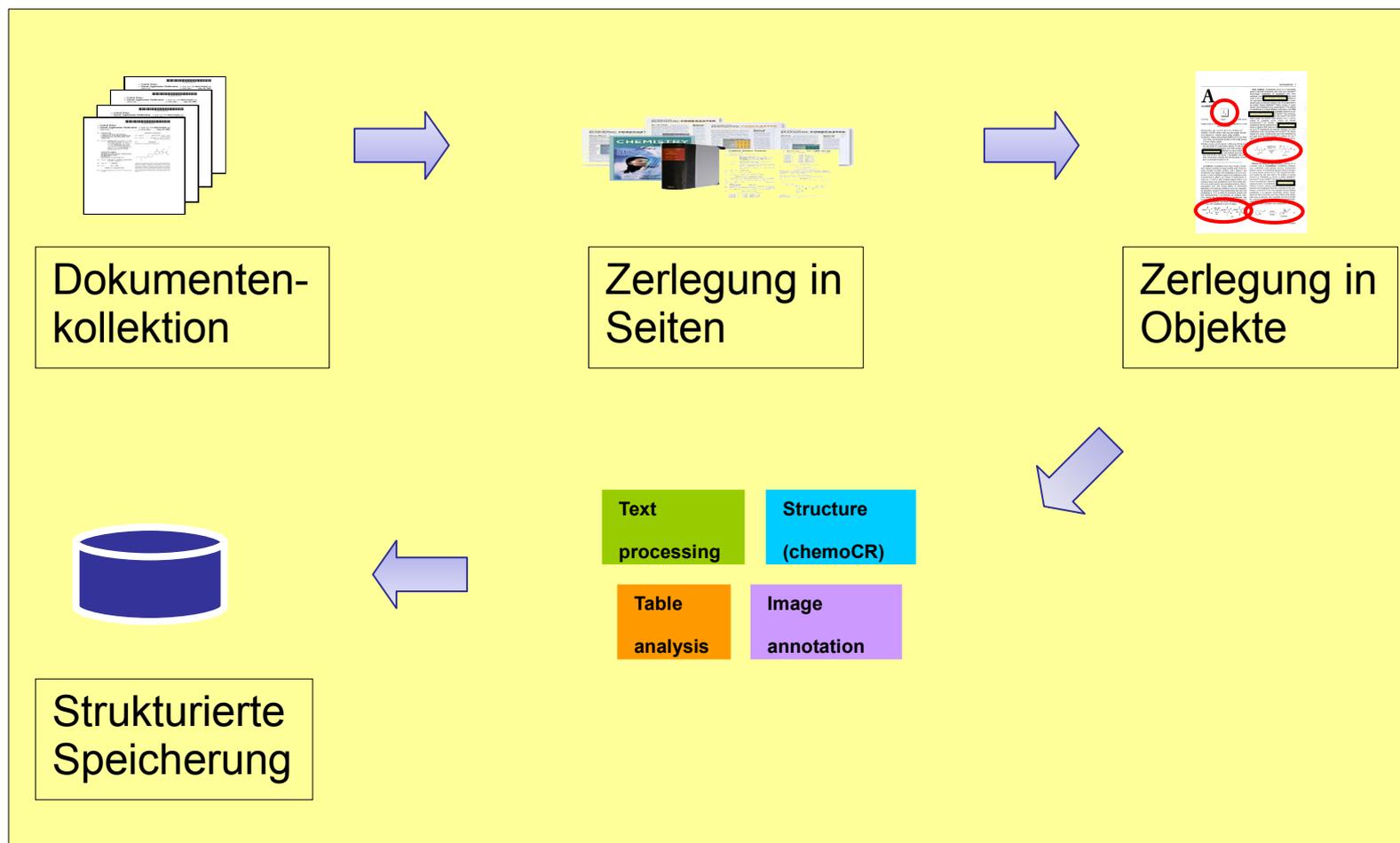
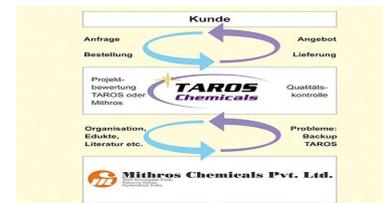
# Anwendungsbeispiel: The grand patent challenge

- Annotation aller chemischen Struktur-Beschreibungen in Patent-Dokumenten
- Verfahren
  - Identifikation von Molekülbeschreibungen in Dokumenten
    - *SMILES: CC(C)CC(=O)*
    - *2D-Zeichnungen oder Bilder*
    - *Molekülname*
  - Eingabe: Mehrseitige Dokumente (pdf, tif, ...)
  - Ausgabe: Datenbank
- High throughput Anwendung: Dieselbe Programm-Folge bearbeitet viele Dokumente unabhängig voneinander

## Realisierung mittels UNICORE Workflows

- 16000 Dokumente von durchschnittlich 14,5 Seiten
- 747000 Moleküle rekonstruiert
- 100 Stunden Rechenzeit
- auf „Juggle“ Cluster (D-Grid)





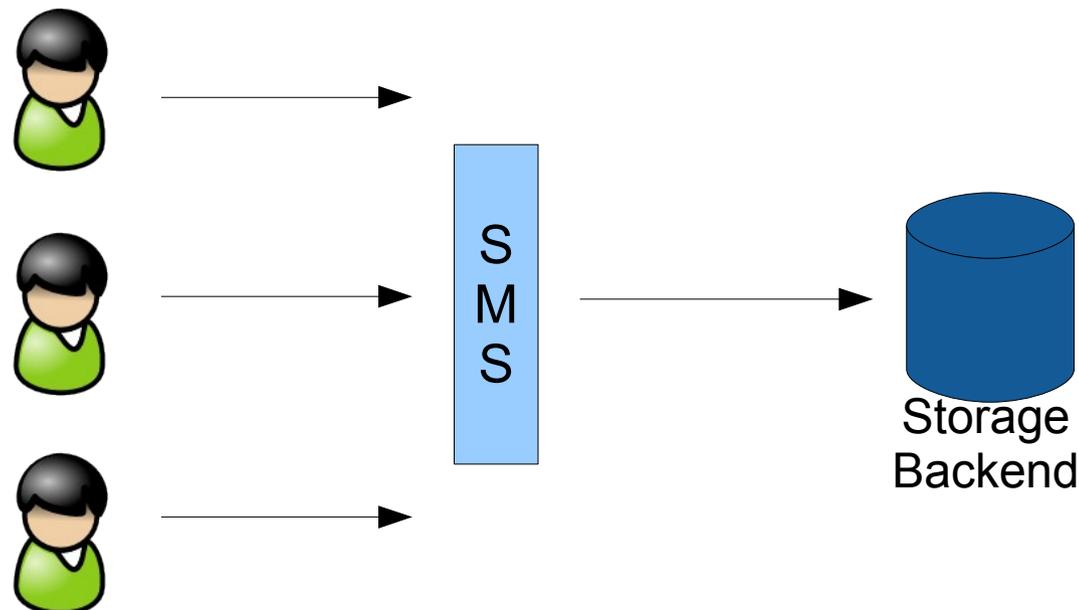
## Aktuelle Entwicklungen

- Daten, Metadaten und Datentransfer
- Virtuelle Systeme
- Kollaboratives Arbeiten

## Skalierbare Speicherdienste

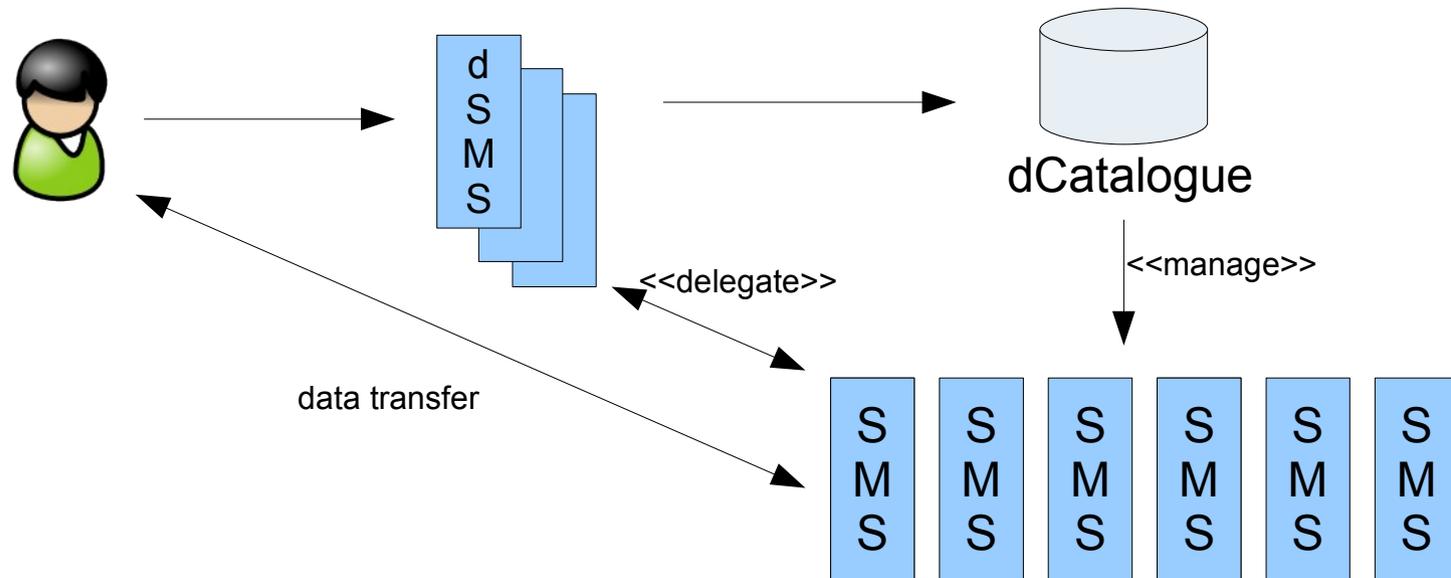
## Skalierbare Speicherdienste: Motivation

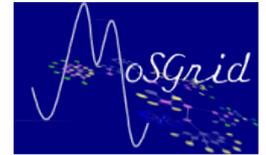
- Zentrale Abstraktion in UNICORE ist der Storage management service (SMS)
- Mögliche Engpässe bei grossen oder häufig benutzten Speichersystemen



# Ein Ansatz: distributed SMS

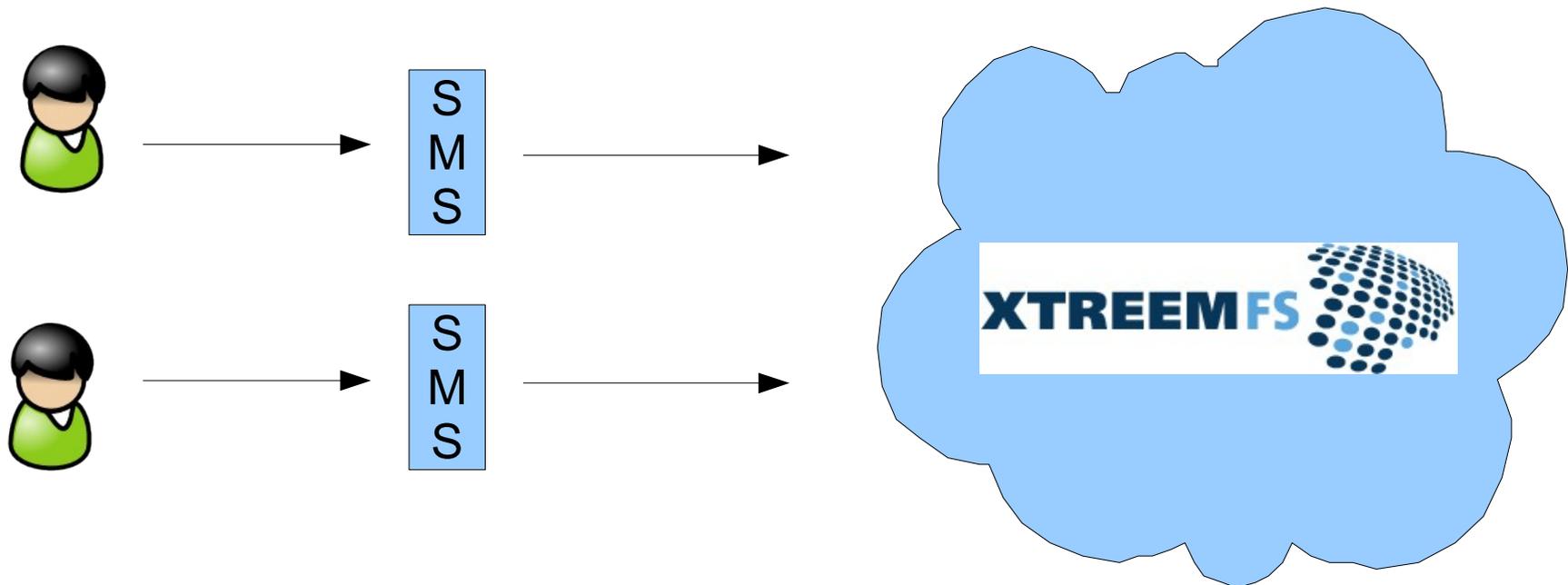
- Distributed Storage management service (dSMS):
  - Selbes Interface wie SMS, aber delegiert Operationen an “echte” SMS
  - Ermöglicht “beliebig viele” Zugangspunkte
  - Datentransfer direkt von/zu den SMS
  - Ähnliche Architektur wie z.B. Apache Hadoop
- File-Katalog(dCatalogue), verwaltet Dateien und angebundene SMS





# Alternative: XtremFS und UNICORE

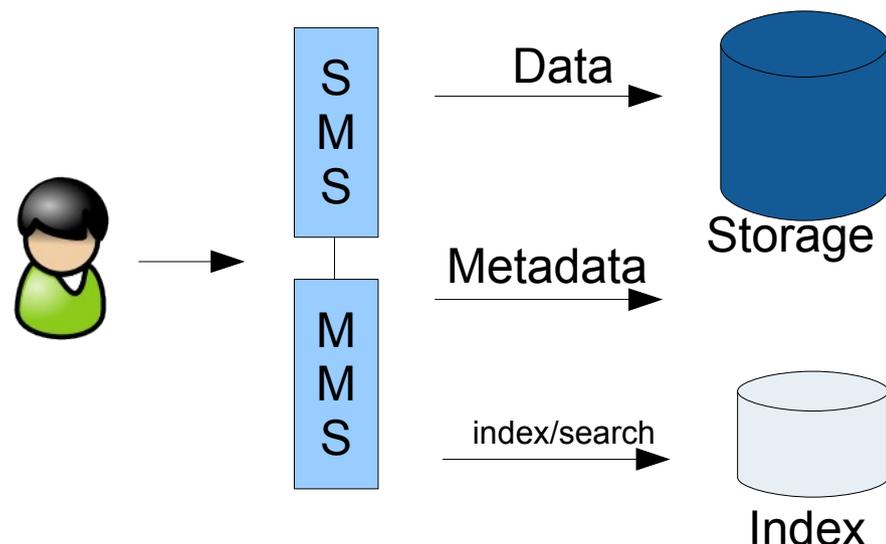
- XtremFS
  - *Verteiltes Filesystem (WAN/Internet), lokal gemountet (FUSE)*
  - *Bietet globalen Namensraum*
- Mehrere Einstiegspunkte über mehrere SMS-Instanzen



## Metadaten

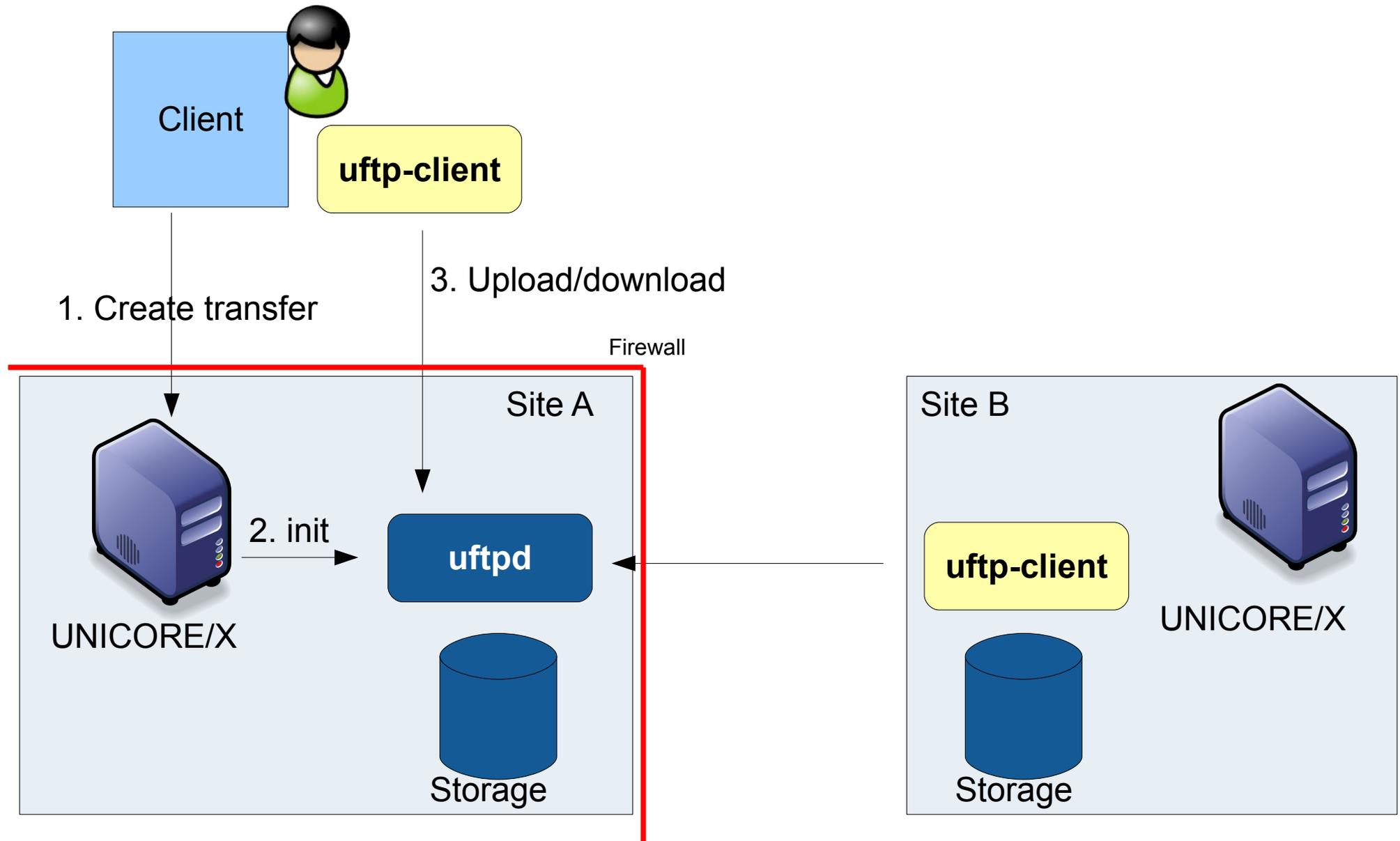
# Integriertes Metadaten-Management

- Dezentral: Je ein „metadata management service“ (MMS) pro Speicher (SMS)
- Schema-frei: Metadaten als Schlüssel/Wert-Paare
- Metadaten werden direkt „neben“ den Daten gespeichert
- Benutzer-definiert und (halb-)automatische Extraktion



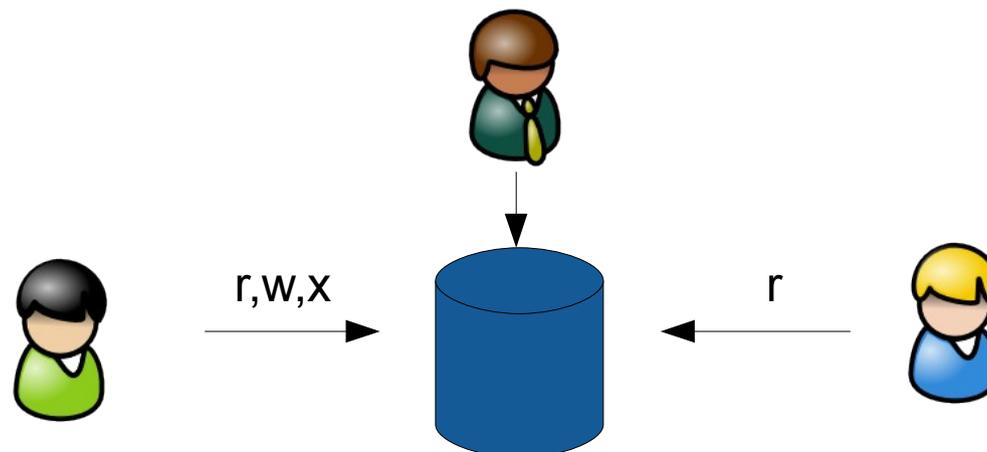
```
schuller@zam994-t400:/$ ucc-vsgc ls -l u6://VSGC-2/
-rw-          1670701 2011-02-24 09:26 /Documents/ref
{
  "Content-Type": "application/pdf",
  "Creation-Date": "2010-09-23T16:25:05Z",
  "Last-Modified": "2010-09-23T16:25:11Z",
  "created": "Thu Sep 23 18:25:05 CEST 2010",
  "creator": "Adobe InDesign CS5 (7.0)",
  "producer": "Adobe PDF Library 9.9",
  "resourceName": "/Documents/refcard-hadoop.pdf",
  "trapped": "False",
  "xmpTPg:NPages": "6"
}
```

# File-Transfer mit UFTP



## **Kollaboratives Arbeiten**

- Bisher:
  - für Dateien: meist Unix-artig (-rwxr--r--)
  - für Dienste: in UNICORE sind nur recht einfache Regeln möglich (z.B. Zugriff für alle Benutzer, nur für eine Gruppe, oder gar nur für den Besitzer)
- In „traditionellen“ Grids (EGEE / EGI, aber auch D-Grid)
  - „Grosse“ virtuelle Organisationen (VOs), abgebildet auf Unix-Gruppen. Hoher Erstellungsaufwand.

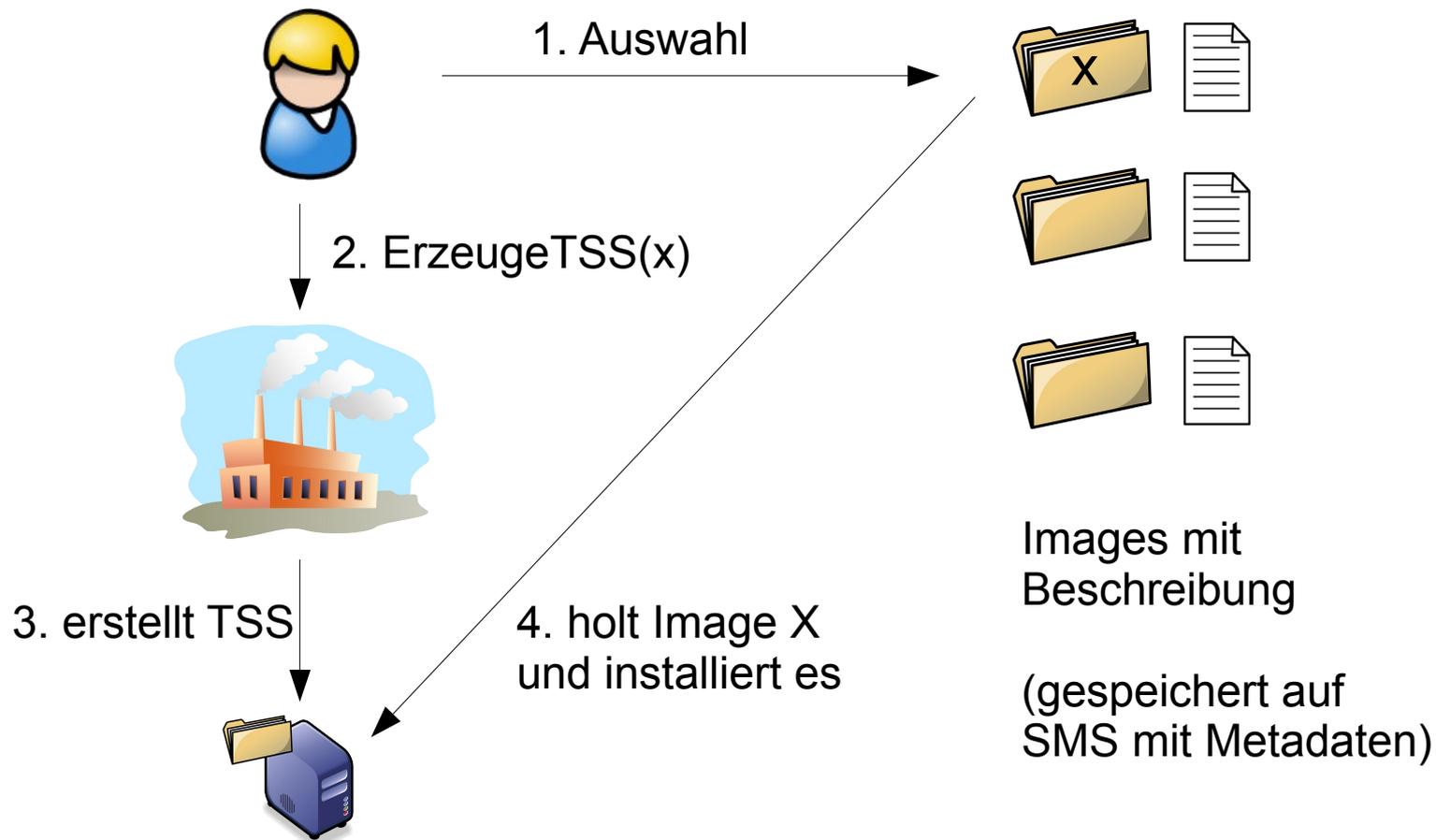


## Ziel: flexible Kollaborationen, Vorbild: das Web 2.0!

The screenshot displays a YouTube video player with the following elements:

- Search Bar:** "Data wave" with a search button and links for "Suchen", "Kategorien", and "Video hochladen".
- Video Title:** "Scientific Evaluation of an accurate measurements last year".
- Channel:** "MediaArtTube" with "300 Videos" and an "Abonnieren" button.
- Video Content:** A 3D pie chart with segments labeled 8, 5, 10, 4, 10, and 6. The video progress bar shows 0:12 / 1:02.
- Right Sidebar:**
  - Warning:** "76% of your data in this storage is unused since 90 days - remove data?"
  - Available Resources (Filter #443):** STORAGE: (RZG), STORAGE: (JUELICH), STORAGE: (RZG), sensor ship #4455, sensor airplane #4711.
  - Available Services (Filter #43):** Data-mining service, Workflow service.
  - 10:46 - CEST Meetings Today:** 9:00 Steering Group, see Meeting Minutes.
- Bottom Left:** "Mag ich" button, "7.557.864 Aufrufe" (views), and icons for atom, DNA, and molecule.
- Comments:**
  - Prof. Dr. Known:** "Correct. I used this great data in a recent publication supporting this claim, you can see it here: [PaperBerlinTraffic](#)" (posted 2 hours ago).
  - Citizen from Berlin:** "I heard that you have used my data from Berlin to claim that the traffic evolution in this region is..." (posted 9 hours ago).
- Most Viewed Today:**
  - MPI toolset usage:** von DieseSchuleRock
  - Paper about aging:** von aberason
  - Device #46 data:** von ElectronicArtsDE

# Unterstützung von Anwender-spezifischen Umgebungen



.... mit kleinen Erweiterungen in UNICORE realisierbar!

**... wie schnell sind eigentlich Datentransfers?**

## UFTP: ein hoch-performanter Datentransfer

- UNICORE FTP (durchaus auch als Gegenpol zu Globus GridFTP)
- Performant (mehrere parallele TCP-Verbindungen)
- Client-Server, Server-Server
- Sicher, einfach einzurichten und zu verwenden
- Öffnen der benötigten Ports in der Firewall („FTP“ Protokol)
- Plattformunabhängig: Java
- Verfügbar mit UNICORE Version 6.4.1
- ... aber sicher werden Filetransfers ein Thema bleiben. Vielleicht sollte man sie ganz vermeiden?

- Aktuelle und zukünftige Anwendungen ...
  - erfordern die Verwendung immer mehr diversifizierter Systeme
  - verlangen ein ausgefeiltes Daten- und Rechte-Management
  - realisieren neue Szenarien jenseits des traditionellen Cluster- und Gridcomputing
  
- UNICORE bietet ...
  - eine solide Grundlage mit gleichzeitig hoher Erweiterbarkeit
  
- Aktuelle Themen sind unter anderem...
  - Verbesserungen und neue Funktionalität im Datenmanagement
  - Realisierung einer flexibleren Rechteverwaltung

- Danke
  - ICM: Krzysztof Benedyczak, Tomas Rekawek, ...
  - JSC: Sandra Bergmann, Jason Daivandy, Bastian Demuth, Björn Hagemeyer, Valentina Huber, Daniel Mallmann, Shiraz Memon, Michael Rambadt, Morris Riedel, Mathilde Romberg, Jedrzej Rybicki
  - ZIH: Richard Grunzke, Christian Loeschen
  - ...