

Wissenschaftsgeleitete Forschungsinfrastrukturen für die Geistes- und Kulturwissenschaften in Deutschland

3. Workshop: Politische Perspektive

Prof. Dr. Martin Grötschel, BBAW
Prof. Dr. Erhard Hinrichs, CLARIN-D
Prof. Dr. Wolfram Horstmann, DARIAH-DE
Prof. Dr. Christof Schöch, DHd-Verband

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Rollen der Anbieter technischer Infrastruktur

Martin Grötschel
Berlin-Brandenburgische
Akademie der Wissenschaften

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

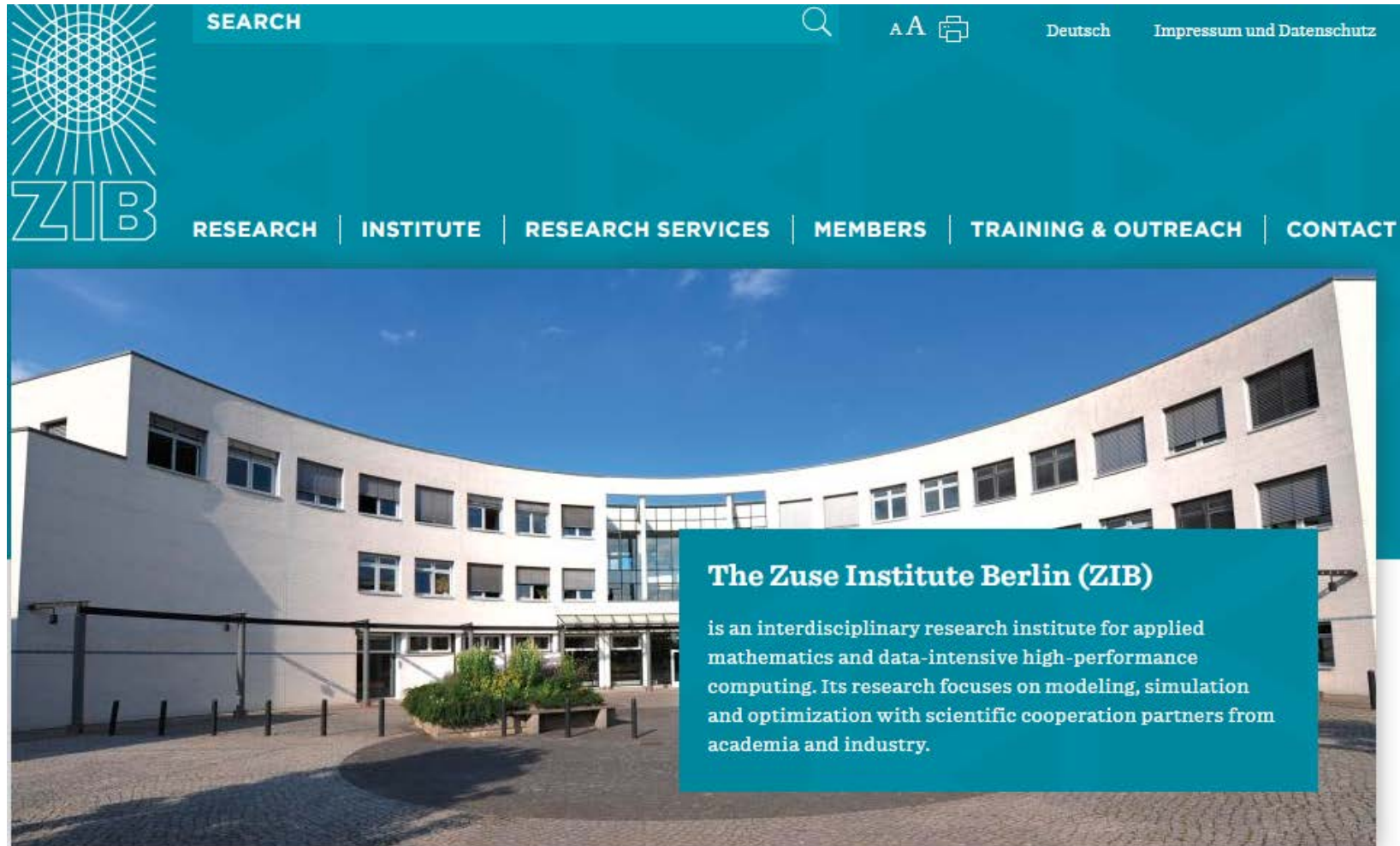
Meine Rollen: (1) Vertretung der Union der deutschen Akademien der Wissenschaften

- Bayerische Akademie der Wissenschaften, Sitz: München
- Berlin-Brandenburgische Akademie der Wissenschaften, Sitz: Berlin und Potsdam
- Akademie der Wissenschaften zu Göttingen
- Heidelberger Akademie der Wissenschaften
- Akademie der Wissenschaften und der Literatur Mainz
- Nordrhein-Westfälische Akademie der Wissenschaften und der Künste, Sitz: Düsseldorf
- Sächsische Akademie der Wissenschaften zu Leipzig
- Akademie der Wissenschaften in Hamburg

Die Nationale Akademie der Wissenschaften Leopoldina ist nicht Mitglied, gleichwohl besteht eine enge Zusammenarbeit.

Meine Rollen: (2) Ich war 25 Jahre VP/P eines „Supercomputer-Rechenzentrums“

Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)



Anbieter technischer Infrastruktur

Unter dieser Bezeichnung treten eine Vielzahl von Einrichtungen auf, z. B.:

- Rechenzentren von Universitäten und wiss. Instituten
- Super Computer/High Performance Rechenzentren
- Forschungsdatenzentren bestimmter Einrichtungen für spezielle Themen (z.B. Netzwerk von 31 vom RatSWD akkreditierten Forschungsdatenzentren)
- Fachinformationsdienste für die Wissenschaft (FIDs) (Weiterentwicklung der Sondersammelgebiete), DFG-Projektförderung, an Bibliotheken angesiedelt
- Leistungsstarke Bibliotheken
- Kommerzielle Anbieter (Cloud und andere Services)
- ...

KURZ ab jetzt: **Rechenzentren** (auch wenn das nicht überall passt)

Fragen an die Rechenzentren (bezüglich der NFDI-Initiative)

Themen:

- Was sollen/müssen Rechenzentren im Rahmen eines NFDI-Konsortiums leisten?
- Was wollen die Fachwissenschaftler von Ihnen?
(Heikle Frage, weil die Antworten schnell zu einem Wunschkonzert ausarten können)
- Was erwarten/wünschen die Rechenzentren selbst?
- Allgemeine Dienste – unabhängig vom Fachgebiet?
- Spezialdienste für die Geisteswissenschaften?

Zunächst ein Blitzüberblick:

Allgemeine Leistungen von Rechenzentren

Basistechnologien und grundlegende Dienste von allen:

- unterstützende Dienstleistungen in Rechen- und Speichertechnologie, sichere Speicher, Cloud-Technologien, zuverlässige Zugriffsmechanismen, damit zusammenhängende Software-Entwicklungen (und ein leistungsfähiges Forschungsnetzwerk)

Je nach Ausrichtung von Rechenzentren (Beispiele):

- Methoden der Künstliche Intelligenz & Maschinelles Lernen
- Bild-, Musik- oder Videodatentechnologien
- Metadaten & Wissenssysteme, die das schnelle und zielgenaue Auffinden von Information in großen Datensystemen und Kontexten ermöglichen (wesentlich für den Aufbau und nutzerfreundlichen Betrieb einer Nationalen Forschungsdateninfrastruktur (NFDI))

Zunächst ein Blitzüberblick:

Rechenzentren forschen auch

Ziel ist die sichere und vertrauenswürdige Erfassung, Speicherung, Verarbeitung, Übertragung und Nutzung von Daten/Information.

Grundsätzliche Forschungsthemen

(Beispiele, unabhängig von GW-NFDI-Konsortien):

- Speichertechnologien
- Datenbank- und Rechner-Architekturen
- Daten-Lebenszyklus
- Suchmethoden
- Big Data Analytics
- Datenschutz (rechtlich und technisch)
- Zugriffsmechanismen (z. B. integrativer Ansatz zur Informationsnutzung in natürlichen, technischen, kognitiven und gesellschaftlichen Systemen)
- Nutzerorientierung

Grundsätzliche Fragen von NFDI-Konsortien – unabhängig von Geistes- und Kulturwissenschaften

Über die Einbindung der Rechenzentren in ein NFDI-Konsortium bestehen deutliche Meinungsunterschiede.

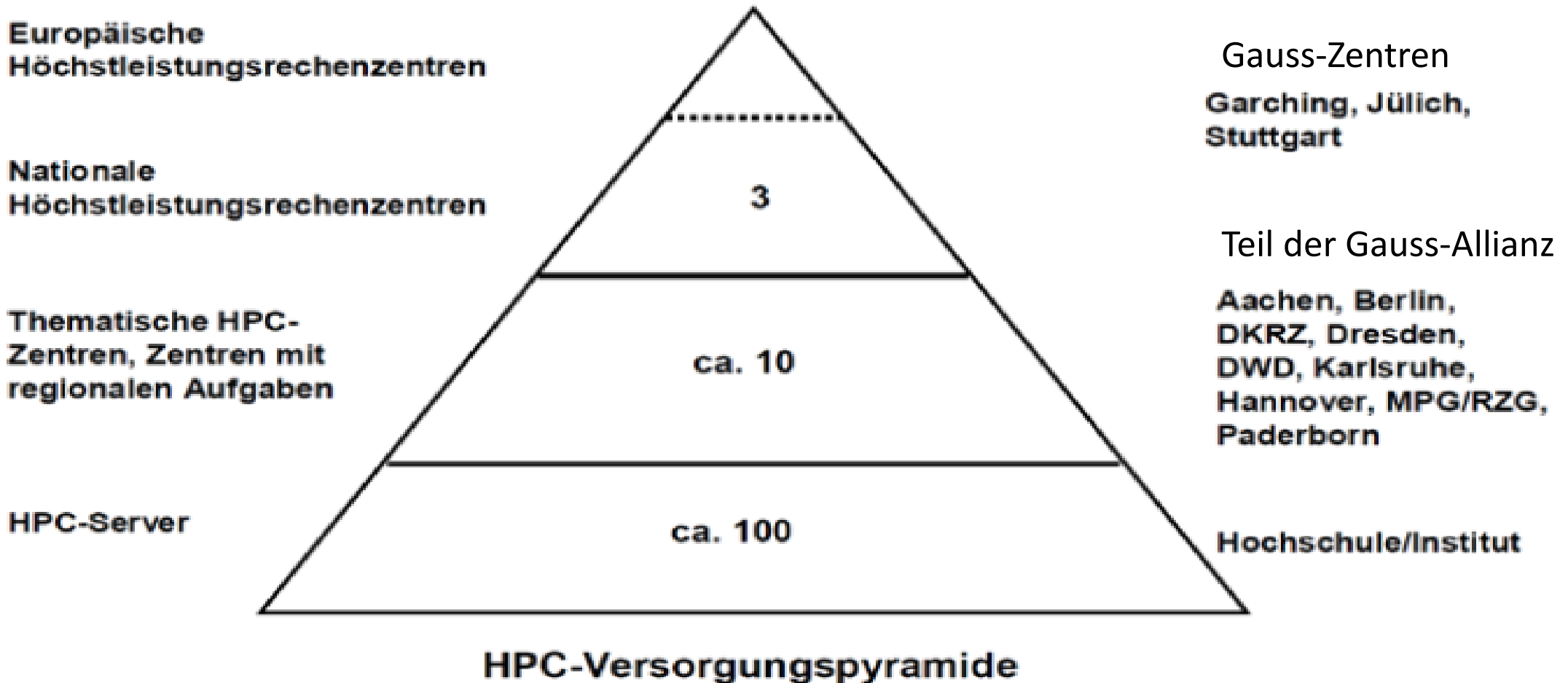
- Sollen Rechenzentren nur Dienstleister sein, bei denen man Dienstleistungen „einkaufen“ kann?
- Sollen Rechenzentren wissenschaftliche Partner sein?
- Wie sehen das die Rechenzentren selbst?
- Sollen auch kommerzielle Rechenzentren als Partner in Betracht gezogen werden?

Wie steht es um die Leistungsfähigkeit akademischer Rechenzentren?

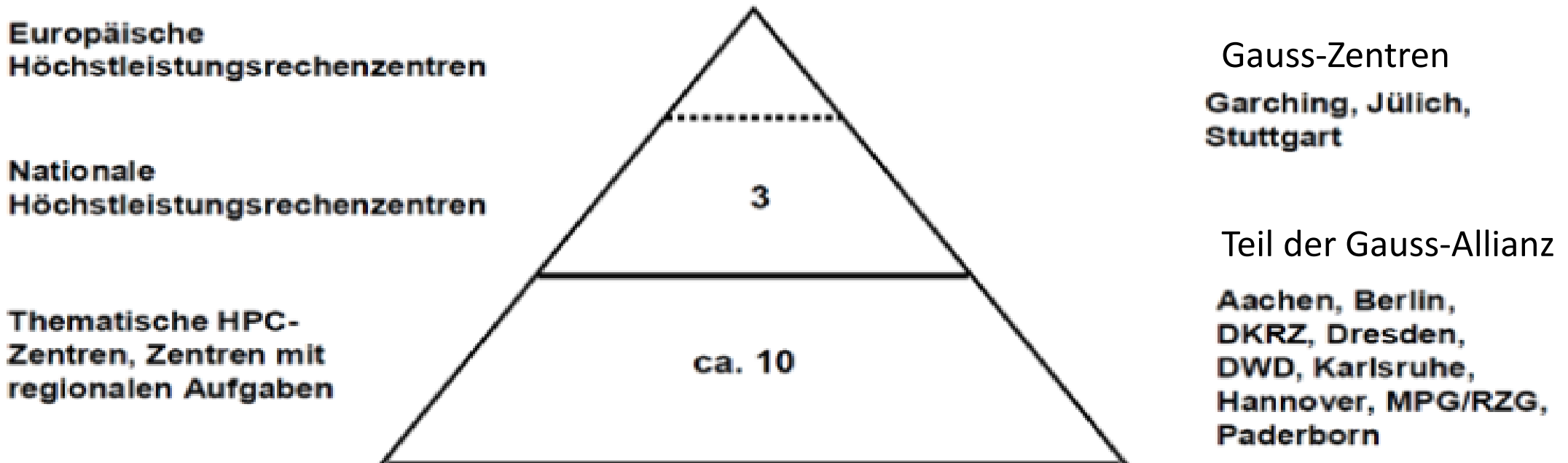
Rechenzentren gibt es an jeder Universität und an jeder wissenschaftlichen Einrichtung.

- Die Leistungsfähigkeit dieser Rechenzentren wird in Bezug auf einige Aspekte der geplanten Forschungsdateninfrastruktur in der Regel stark überschätzt.
- Für die gesamte NFDI müssen für mehrere der zu lösenden Aufgaben „starke“ Partner gefunden werden.
- Gibt es solche starken Partner?

Starke Partner, die Situation in Deutschland: High Performance Computing-Versorgung



Starke Partner, die Situation in Deutschland: High Performance Computing-Versorgung



Meine Meinung: Supercomputing ist für die NFDI nicht notwendig, aber Höchstleistungsrechenzentren haben die (teure) technische Infrastruktur, die für einen Teil der NFDI-Aufgaben essentiell ist.

HPC-Versorgungspyramide

Ein Beispiel für ein Supercomputing-Zentrum



Leibniz-Rechenzentrum
der Bayerischen Akademie der Wissenschaften

**IT-Dienste für
Forschung und Lehre
in München, Bayern,
Deutschland und Europa**

LRZ-Reklame: Cloud Storage im Münchner Wissenschaftsnetz

Die Daten werden auf skalierbaren, leistungsstarken Systemen gespeichert. Die unterbrechungsfreie und verlustsichere Bereitstellung der Daten wird durch Snapshots, Spiegelung und mehrstufige Sicherung gewährleistet. Mit Hilfe der Snapshots können Sie Daten selbst wiederherstellen, z.B. bei Befall durch einen Verschlüsselungstrojaner oder versehentlichem Löschen/Überschreiben.

Die zentrale Datenhaltung am LRZ bietet noch weitere Vorteile:

- effektive Ressourcennutzung
- kaum Administrationsaufwand vor Ort, da die Administration durch geschultes und zertifiziertes LRZ-Fachpersonal erfolgt,
- eine schnelle Fehlerbehebung durch Premium-Supportverträge mit der Herstellerfirma.

Ein Beispiel für ein HPC-Zentrum mit regionalen Aufgaben: HLRN (Berlin & Göttingen)

Der **Norddeutsche Verbund für Hoch- und Höchstleistungsrechnen** (HLRN) betreibt an den Standorten Zuse-Institut Berlin (ZIB) und Georg-August-Universität Göttingen ein verteiltes Supercomputersystem für Wissenschaft und Forschung in **sieben norddeutschen Bundesländern**.

Die Nutzer des HLRN werden durch ein überregionales Kompetenz-Netzwerk von Fachberatern, an dem alle Institutionen des HLRN beteiligt sind, unterstützt.

Mit seinen beiden Betreiberzentren ist der HLRN-Verbund in der **Gauss-Allianz** vertreten.

Die Gauss-Allianz hat 19 Mitglieder.

<https://www.hlrn.de/home/view>

ZIB-Reklame:

http://www.zib.de/de/research_services/supercomputing

Large Storage Capacities for HPC & Big Data Analytics

In HPC, today's complex computational workflows including pre- / post-processing and computational steps require large on-line storage capacities. Even for increasingly on-the-fly computational methods the input and output data sets for individual jobs on HLRN-III hit the Tera byte scale.

With the on-going unification of requirements in the HPC and Big Data world, these large-scale storage capacities providing an optimal bandwidth to data sets become vital for current and future workloads.

Storage Infrastructure for HLRN-III at a glance:



3.7 PByte

on-line storage capacity is available in two globally accessible parallel file systems (Lustre) offering a high bandwidth access to application data in batch jobs.



0.5 PByte

on-line storage capacity is provided in a globally accessible NAS appliance for permanent project data and program code development.



Peta-scale

tape library for archiving large data sets is operated independently by ZIB.



Sicheres Datenarchiv



Lampertz Sicherheitsraum (EN 1047-2)
Sicherung gegen
Feuer, Wasser, Rauch und Diebstahl



Foto: J. Keiper

CO₂ Feuerlöschanlage



Wissenschaftsratsempfehlung 2015: **NHR**

Mit den vorliegenden Empfehlungen legt der Wissenschaftsrat die Grundlagen für die **langfristige Strukturentwicklung des HPC in Deutschland**.

- Die Kernempfehlung besteht darin, einen **Verbund von Kompetenzzentren** für Nationales Hoch- und Höchstleistungsrechnen (**NHR**) im Sinne einer langfristig stabilen, gut koordinierten und deutschlandweit einfach zugänglichen Forschungsinfrastruktur einzurichten.
- Die strategische Steuerung des NHR sollte ein externer, **koordinierender Ausschuss** übernehmen.
- Operative und technische Fragestellungen sollten dagegen in einem **gemeinsamen internen Lenkungsausschuss** der Zentren behandelt und entschieden werden. Für die Auswahl von Anträgen sollte außerdem ein gemeinsamer Nutzungsausschuss gebildet werden, der die zentrenspezifischen Auswahlausschüsse koordiniert.

<https://www.wissenschaftsrat.de/download/archiv/4488-15.pdf>

Wissenschaftsratsempfehlung 2015: Aufgaben von NHR-Zentren

- 1 – Nationale Kompetenzzentren:
- 2 – Offener Zugang für alle Wissenschaftlerinnen und Wissenschaftler in Deutschland:
- 3 – Einheitlicher Zugang:
- 4 – Flächendeckende Bereitstellung von HPC-Ressourcen:
- 5 – Koordination zwischen den Ebenen:
- 6 – Koordination zwischen Zentren:
- 7 – Strukturen: Zum Zwecke der Koordination und zur Sicherstellung ihrer Arbeitsfähigkeit als nationale Forschungsinfrastruktur nutzen die NHR-Zentren gemeinsame Strukturen und Organisationseinheiten.
- 8 – Methodenkompetenz und Beratung:
- 9 – Pflege von Softwarecodes: NHR-Zentren sind in der Pflege und bedarfsgerechten Weiterentwicklung wichtiger Softwarecodes aktiv, um deren fortgesetzte und ressourceneffiziente Einsatzfähigkeit auf neuen Systemen zu gewährleisten.

Wissenschaftsratsempfehlung 2015: Mehrwert des NHR

- 1 – Höhere Nutzerfreundlichkeit:
- 2 – Gestärkte Methodenkompetenz:
- 3 – Verbesserte Aus- und Weiterbildung:
- 4 – Verbesserte Kosteneffizienz:
- 5 – Verbesserte Energieeffizienz:
- 6 – Bessere Qualitätssicherung:
- 7 – Gemeinsame Strukturen: Eine nationale Koordination der Zentren auf Ebene 1 und 2 mit Strukturen für die gemeinsam anzugehenden Aufgaben ist der Schlüssel zu mehreren oben genannten Punkten. Durch ein koordiniertes NHR besteht damit auch ein Ansprechpartner für die internationale Zusammenarbeit.

Zusammenfassung

- Die NHR-Struktur entsteht gegenwärtig. Es ist sinnvoll, diese bei der Planung von NFDI-Konsortien zu berücksichtigen.
- Ich glaube nicht, dass man (starke) Rechenzentren als reine Dienstleister gewinnen kann (oder sollte).
- Rechenzentren müssen zur Einrichtung von NFDI-Konsortien sorgfältig ausgewählt werden.
- Die beteiligten Rechenzentren müssen sich mit den Besonderheiten der GW-Anforderungen befassen wollen und partnerschaftlich in allen Belangen mitwirken.
- In Deutschland sind die Voraussetzungen dafür vorhanden.

Geeignete Partner finden: Über Kriterien nachdenken

- Rechenzentren unter NFDI-Gesichtspunkten zertifizieren?
- Die starken Rechenzentren werden mehrere NFDI-Konsortien beheimaten (müssen).
- Soll eine wettbewerbliche Auswahl unter zertifizierten Rechenzentren erfolgen?
- Regionale, nationale, europäische und internationale Aspekte?
- Bereitschaft zu einer funktionalen Vernetzung?
- Kommerzielle Anbieter berücksichtigen?
- **Bedeutung von Rechenzentren für GW-NFDI-Konsortien:**
 - **Bereitstellung nicht nur von technischer Infrastruktur sondern auch von technischen Services und Beratung,**
 - **Willen zur Erstellung von Angeboten von generischen Diensten**

Geeignete Partner finden:

Wichtige technische Aspekte (kleine Auswahl)

- Vertrauenswürdige Software-Unterstützung von Datenerhebung, -vermittlung und -lieferung
- Langfristiges Forschungsdatenmanagement (Datenkuratierung)
- Sicherstellung von Interoperabilität
- Standardisierung von Metadaten und Speicherformaten
- Spektrum von Services
 - Cloubasierte Infrastrukturangebote
 - Software as a service
 - Function as a service
 - NFDI-Konsortien-unabhängige skalierbare Tools

Geeignete Partner finden:

Wichtige technische Aspekte (kleine Auswahl)

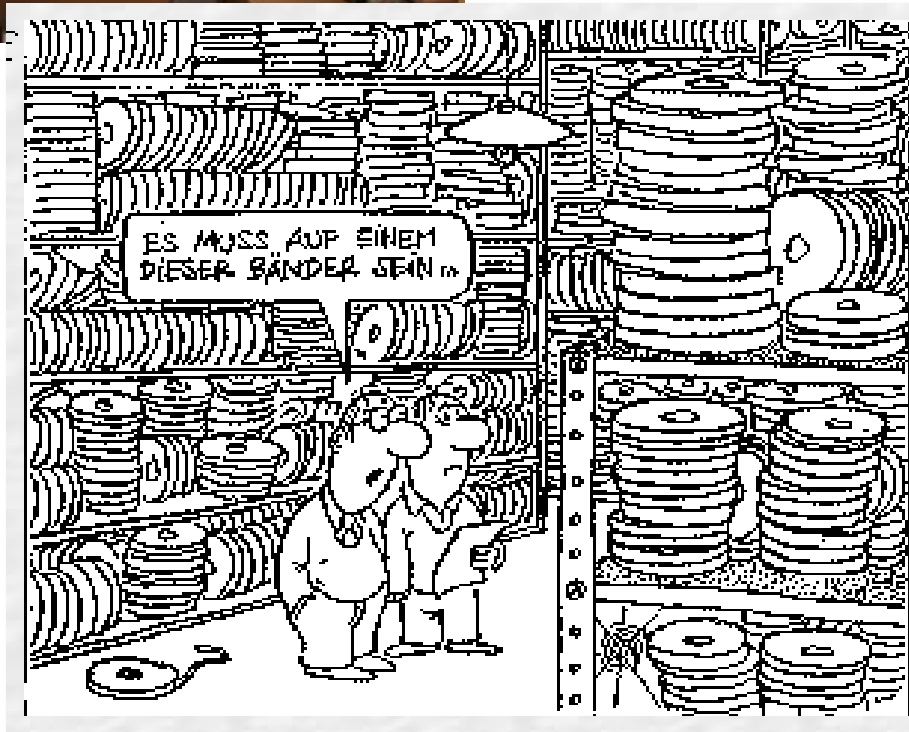
- Vertrauenswürdige Software-Unternehmen
- -vermittlung und -lieferung
- Langfristige Partnerschaft
- Die Auswahl von Anbietern technischer Infrastruktur darf nicht anhand ihrer Reklamefolien sondern muss danach erfolgen, inwieweit sie das für ein GW-NFDI-Konsortium geeignete Service-Spektrum bereitstellen
- Cloud-basierte Infrastrukturangebote
- Software as a service
- Function as a service
- NFDI-Konsortien-unabhängige skalierbare Tools



Stadtarchiv Köln



Anna Amalia Bibliothek Weimar



NFDI:
Bewusste Vermeidung von
und Absicherung gegen
„Katastrophen“ notwendig



Die Auswahl von Anbietern technischer Infrastruktur muss – in Verbindung mit fach- und methodenspezifischen Überlegungen – bei der Planung von NFDI-Konsortien mit großer Aufmerksamkeit berücksichtigt werden.



Amalia Bibliothek Weimar

NFDI:

Bewusste Vermeidung von
und Absicherung gegen
„Katastrophen“ notwendig

Das Angebot der Akademien

- Die Akademien haben Kompetenzen in der Verarbeitung zahlreicher Forschungsdatentypen. Schwerpunkte: Editions-, Sammlungs-, Erschließungs- und Wörterbuchprojekte. (Juni-Vortrag Schrade)
- Die Akademien sind – national und international stark vernetzt – in der praktischen Forschung verankert, haben in ihren Schwerpunkten große und wertvolle Datenbestände und entwickeln Software-Werkzeuge zu deren „Bearbeitung“. Sie sind willens Daten und Software nach den FAIR-Prinzipien verfügbar zu machen.
- Die Akademien repräsentieren auch viele „kleine Fächer“.

Das Angebot der Akademien

- Wunsch: Beteiligung an einem GW-NFDI-Konsortium, das Editionen, Textsammlungen und Wörterbücher umfasst.
(Vorträge Rapp & Geyken heute)
- Bereitschaft zur nachhaltigen Zusammenarbeit mit Partnern in diesen Themenbereichen zur koordinierten Bereitstellung von nutzerfreundlichen Diensten und Beratung.
- Die Akademien bieten institutionelle Konstanz.
- Erfahrung in der Zusammenarbeit mit Rechenzentren ist vorhanden.

Das Angebot der Akademien

- Die Akademien bevorzugen eine in den GW-Fachkulturen begründete aber fachübergreifende Konsortialbildung
(Verantwortung für kleine Fächer, Beispiel: Marx-Engels-Gesamtausgabe)
- Die Akademien möchten sich an einem NFDI-Konsortium beteiligen, das sich:
 - auf die „Behandlung“ von bestimmten Datentypen mit hoher GW-Relevanz einigt und
 - darauf aufbauend, entlang der FAIR-Prinzipien, fach- und methodenspezifische Verfahren, Dienste und Standards innerhalb der praxis-bezogenen Reichweite des Konsortiums entwickelt.
- Die Akademien sind bereit, im Rahmen der NFDI-Struktur mitzuwirken:
 - an der Etablierung generischer Dienste und fachübergreifender Standards,
 - am Qualitätsmanagement und
 - an der Definition von Schnittstellen und weiteren technischen Grundlagen für das Forschungsdatenmanagement.

Weiteres Vorgehen?

- Der Planungsprozess für NFDI-Konsortien in den Geistes- und Kulturwissenschaften muss jetzt konkret begonnen werden.
- Die „Spieler“, die an der Planung eines NFDI-Konsortiums und dessen Betrieb mitwirken wollen, müssen ihre Intentionen und Ziele möglichst bald offenlegen.
- Wenn der genaue Rahmen der Ausschreibung bekannt gegeben wird, ist es vermutlich zu spät für eine Teilnahme an der ersten Ausschreibungsrunde.
- Konkrete Interessensbekundungen in der Abschlussdiskussion heute Abend?

Vielen Dank für Ihre Aufmerksamkeit

Rollen der Anbieter technischer Infrastruktur

Martin Grötschel

Berlin-Brandenburgische
Akademie der Wissenschaften

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung