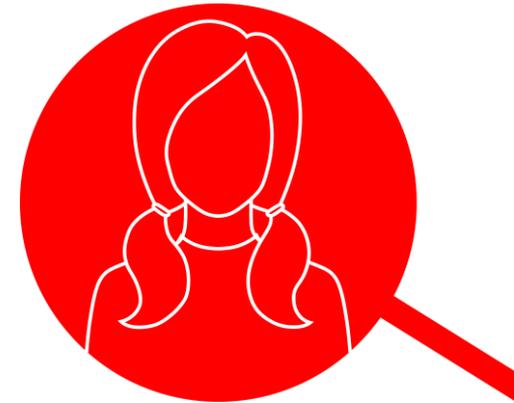


NFDI4ing - DFG 45 - Community Meeting

High Performance Measurement and Computation (HPMC)

NFDI4ing Archetyp DORIS

04.10.2021



HPMC-Forschungsdaten

Was sind HPMC Forschungsdaten?

Forschungsdaten sind Daten, die während eines Forschungsprozesses entstehen oder dessen Ergebnis sind

High Performance **M**asurement

- Messwerte
- Metadaten (Hardware, Methode etc.)



Auswertung der Messdaten mit Hilfe von HPC

High Performance **C**omputing

- Skripte / Code (?)
- Inputfile, Outputfile, Logfile
- Rohdaten
- Prozessierte Daten
- Metadaten (Software, Hardware, Methode etc.)
- Daten für Sekundärforschung (z.B. Energieverbrauch oder Temperatur im HPC)

Vortrag: Beteiligte



Prof. Dr.-Ing. Christian Stemmer, Lehrstuhl für Aerodynamik und Strömungsmechanik TUM, Forschungsprojekt NFDI4Ing



Benjamin Farnbacher, Lehrstuhl für Aerodynamik und Strömungsmechanik TUM, Forschungsprojekt NFDI4Ing



Nils Hoppe, Lehrstuhl für Aerodynamik und Strömungsmechanik TUM, Forschungsprojekt NFDI4Ing



Vasiliki Sdralia, Munich Data Science Institut TUM, Lehrstuhl für Aerodynamik und Strömungsmechanik TUM, Forschungsprojekt NFDI4Ing



Kontakt: info-doris@nfdi4ing.de

Website: <https://nfdi4ing.de/archetypes/doris/>

Newsletter: https://lists.tu-darmstadt.de/mailman/listinfo/nfdi4ing_taskarea_doris

DORIS: Team

[https://nfdi4ing.de/archetypes/doris/
info-doris@nfdi4ing.de](https://nfdi4ing.de/archetypes/doris/info-doris@nfdi4ing.de)

TUM: LS für Aerodynamik und Strömungsmechanik (Koordination)

Prof. Dr.-Ing. Christian Stemmer | christian.stemmer@tum.de (Sprecher)
Benjamin Farnbacher | benjamin.farnbacher@tum.de
Nils Hoppe | nils.hoppe@tum.de

High-Performance Computing Center Stuttgart (HLRS)

Dr.-Ing. habil. Martin Thomas Horsch | martin.horsch@hlrs.de
Dr.-Ing. Björn Schembera | bjorn.schembera@hlrs.de

Leibniz-Rechenzentrum (LRZ)

Dr. Stephan Hachinger | stephan.hachinger@lrz.de
Johannes Munke | johannes.munke@lrz.de

RWTH Aachen: Lehrstuhl für Strömungslehre und Aerodynamisches Institut (AIA)

Dr.-Ing. Simon Loosen | s.loosen@aia.rwth-aachen.de

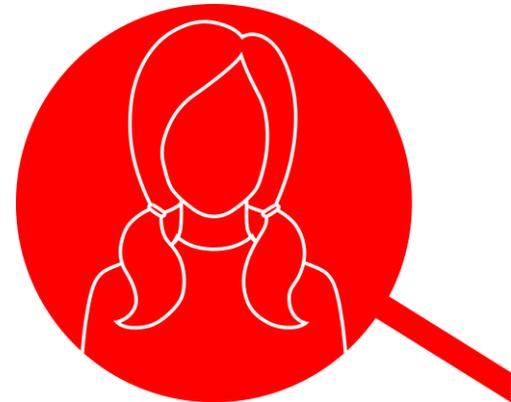
TUM: Munich Data Science Institute (MDSI) / LS für Aerodynamik und Strömungsmechanik

Vasiliki Sdralia | vasiliki.sdralia@tum.de

TUM: Universitätsbibliothek (UB)

Dr. Christine Wolter | christine.wolter@ub.tum.de

... dazu Pilotnutzer aus der Anwendung



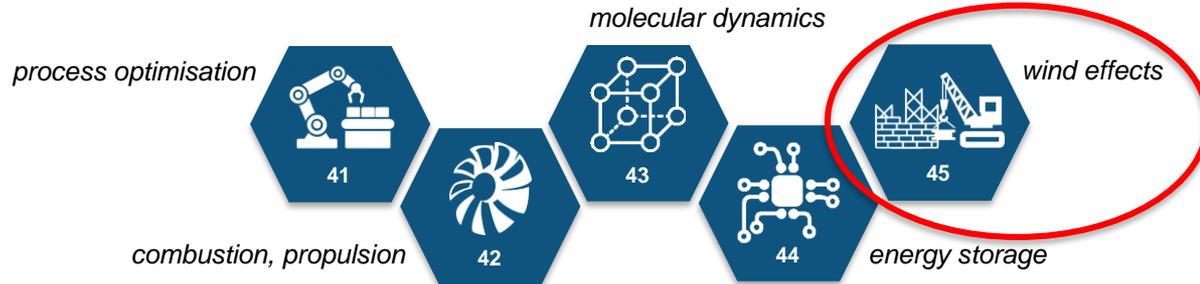
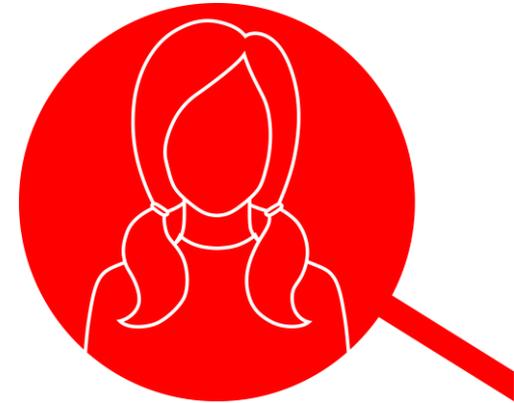
Archetyp DORIS

... I'm an engineer conducting and post-processing high-resolution and **high-performance measurements and computation** (simulation) **with very large data** on HPC systems.

The data sets I work with are extremely large and as such are largely immobile. This mandates tailored, hand-made software."

My needs are

- Enable **exchange** of **huge** high-quality **datasets**.
- Provision of HPC-data to foster **wide-spread usage**.
- Drive NFDI-wide **new methodological** sharing approaches.



DORIS's patron is
Christian Stemmer

SUPERMUC Nutzung nach Fächern	Jobs		Systemzeit	
	Anzahl	%	CPU-Std.	%
Computational Fluid Dynamics	18538	2.9	242108463.3	30.5
Physics - High Energy Physics	510613	79.1	159499564.2	20.1
Astrophysics/Cosmology	6243	1.0	141512980.3	17.8
Biophysics/Biology/Bioinformatics	4478	0.7	71573902.4	9.0
Chemistry	7487	1.2	39442036.4	5.0
Plasmaphysik	3891	0.6	34252924.3	4.3
Geophysics	4304	0.7	28257114.6	3.6
Physics - Solid State	1729	0.3	28080372.3	3.5
Engineering - others	2491	0.4	16447113.2	2.1
Meteorology/Climatology/Oceanography	11133	1.7	13684019.6	1.7
Informatics/Computer Sciences	22236	3.4	7418101.6	0.9
Support/Benchmarking	47943	7.4	3503220.6	0.4
Enviromental Sciences	484	0.1	2925216.2	0.4
Mathematics	2267	0.4	2592946.2	0.3
Material Science	545	0.1	1976808.3	0.2
Engineering - Electrical Engineering	480	0.1	846339.9	0.1
Crystallography	370	0.1	637657.9	0.1
Leibniz-Rechenzentrum	17	0.0	13.3	0.0
Summe	645249	100.0	794758794.5	100.0

HPMC-Forschungsdaten

Besonderheiten

- Daten werden in HPC-Zentren in personalisierten Accounts erzeugt und gespeichert → keine Indizierung durch Repositorien oder Suchmaschinen
- Spezielle Hard- & Software zum Erzeugen, Lesen und Bearbeiten notwendig
- Terabyte bis Petabyte → Daten nicht mobil
- „Daten“ bestehen aus diversen Komponenten (Code, Inputfile, Rohdaten, postprozessierte Daten, Metadaten etc.)
- Kein einheitliches / etabliertes Metadatenschema
- Keine Best-Practice-Beispiele für Datenmanagementpläne, Terminologien etc.

Umsetzung der FAIR-Prinzipien

Findable: Speicherung in persönlichen Accounts, keine Metadaten



Accessible: kein Zugriff durch Externe, unzureichende Transfertools



Interoperable: abhängig von Formaten und Metadaten



Reusable: Rechenzeit an HPC-Zentren oder Containervirtualisierung

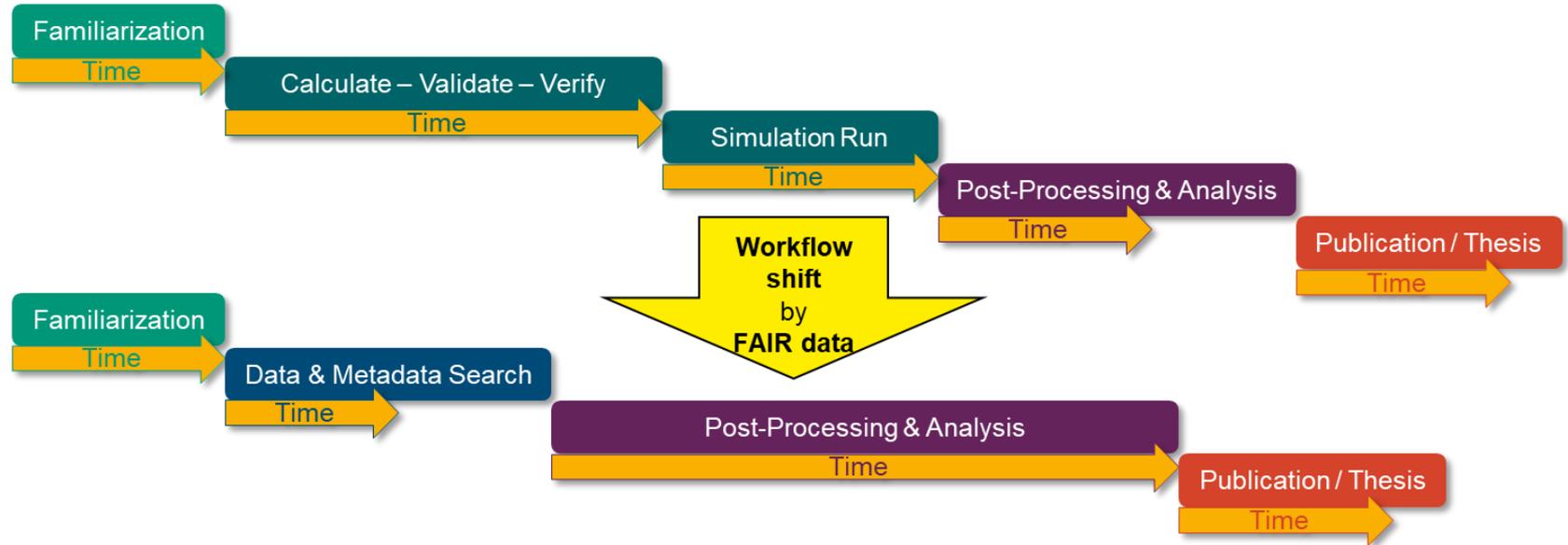


HPMC-Forschungsdaten

Warum Forschungsdatenmanagement für HPMC-Daten?

- Integrität / Erfüllung der DFG-Kriterien
- Sekundärforschung (z.B. Optimierung der Energienutzung eines Supercomputers)
- Neuer Erkenntnisgewinn basierend auf vorhandenen Daten

HPMC-Forschungsdaten



DORIS: Aufgaben und Maßnahmen

Maßnahmen gemäß Projektantrag

- D-1: Accessibility and access rights, data security and sovereignty
- D-2: Support for third-party users & community-based training, provision of post-processing algorithms and modules
- D-3: Metadata definitions & terminologies, support to data-generating groups
- D-4: Storage & archive for very large data
- D-5: Reproducibility on large-scale high-performance systems

Updates / DORIS-Newsletter:

https://lists.tu-darmstadt.de/mailman/listinfo/nfdi4ing_taskarea_doris

DORIS: Aufgaben und Maßnahmen

Daran wird aktuell gearbeitet

- Rechtliche Aspekte, Datensicherheit und -Souveränität für (HPMC-) Daten
- Mitarbeit an einem einheitlichen NFDI4Ing Metadatenchema (Ontologie)
- Entwurf eines best-practice Metadatenchemas für HPMC-Daten (Sub-Ontologie)
- Entwicklung eines Metadata-Crawlers zur automatisierten Auslese von Metadaten und Eingabe in Ontologie

- Test von Interfaces, Front-Ends und Repositorien

- Testprojekte in den großen Rechenzentren (LRZ, HLRS, JSC)
- Test von Transfertools und Transferprotokollen
- Test von Reproduktionsmöglichkeiten, Container- und Postprocessing-Tools

- Lehre, Training und Support

DORIS: Aufgaben und Maßnahmen

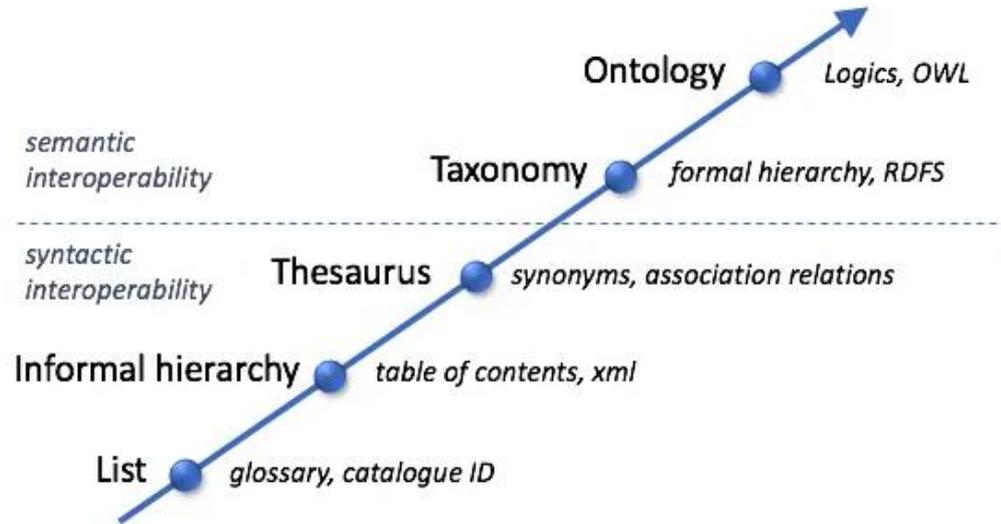
Rechtliche Aspekte, Datensicherheit und -Souveränität für (HPMC-) Daten

- Konzept zu Vertrauensschutz
- Benutzerzugangsverwaltung und Datensicherung
- Umsetzung von ISO/IEC 27001 - *Information technology – Security techniques – Information security management systems – Requirements*
- NFDI-Working group “Copyright and Licenses” / section “Legal Aspects”



DORIS: Aufgaben und Maßnahmen

Metadaten und Ontologie



https://i1.wp.com/emmc.info/wp-content/uploads/2019/05/alternatives_2.png?w=600&ssl=1 (04.05.2021)

DORIS: Aufgaben und Maßnahmen

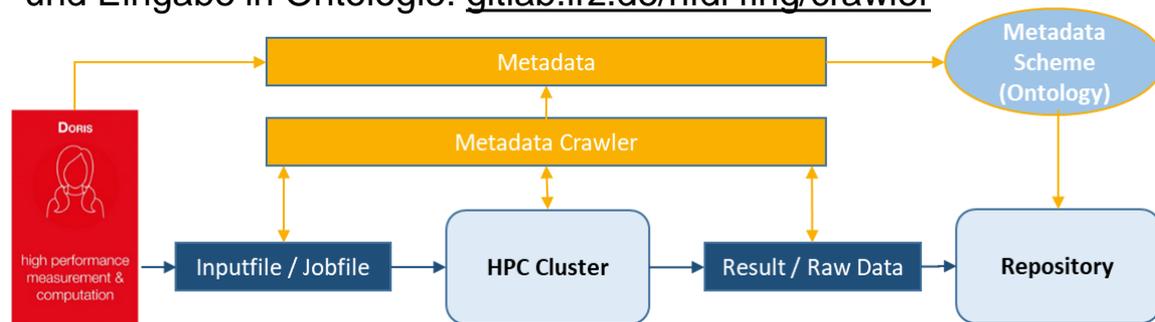
Metadaten und Ontologie

- Best-Practice Metadaten schemata für HPMC-Daten
- Entwicklung einer NFDI4Ing-Ontologie / Einbindung einer **HPMC-Subontologie**
- NFDI4Ing Ontologie (nicht öffentlich): erster öffentlicher Entwurf „coming soon“
<https://git.rwth-aachen.de/nfdi4ing/metadata4ing>

DORIS: Aufgaben und Maßnahmen

Metadaten und Ontologie

- Best-Practice Metadaten schemata für HPMC-Daten
- Entwicklung einer NFDI4Ing-Ontologie / Einbindung einer **HPMC-Subontologie**
- NFDI4Ing Ontologie (nicht öffentlich): erster öffentlicher Entwurf „coming soon“
<https://git.rwth-aachen.de/nfdi4ing/metadata4ing>
- Entwicklung eines **Metadata-Crawlers** zur automatisierten Auslese von Metadaten und Eingabe in Ontologie: gitlab.lrz.de/nfdi4ing/crawler



DORIS: Aufgaben und Maßnahmen

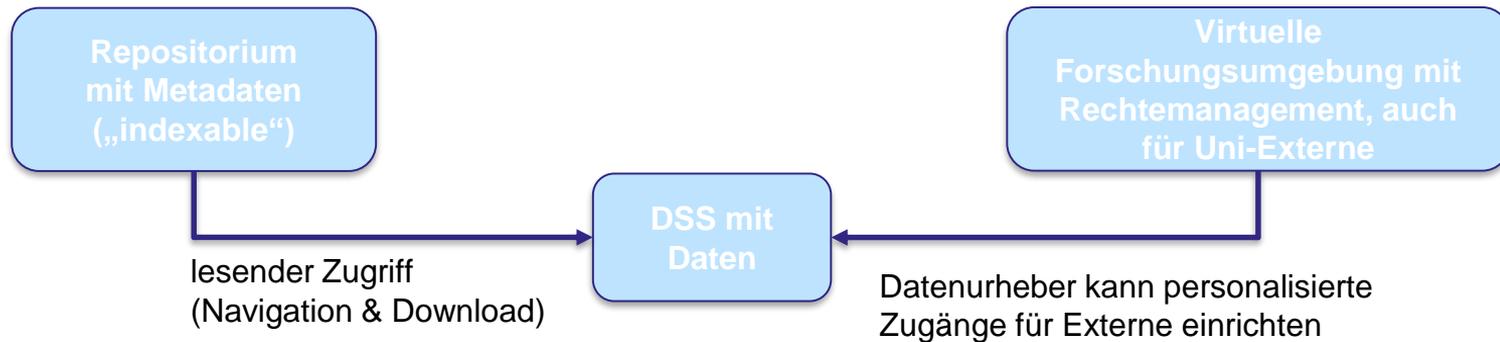
Metadaten und Ontologie

- Best-Practice Metadaten schemata für HPMC-Daten
- Entwicklung einer NFDI4Ing-Ontologie / Einbindung einer **HPMC-Subontologie**
- NFDI4Ing Ontologie (nicht öffentlich): erster öffentlicher Entwurf „coming soon“
<https://git.rwth-aachen.de/nfdi4ing/metadata4ing>
- Entwicklung eines **Metadata-Crawlers** zur automatisierten Auslese von Metadaten und Eingabe in Ontologie: gitlab.lrz.de/nfdi4ing/crawler
- **Mailingliste „HPC Metadata“ (LRZ):** https://lists.tu-darmstadt.de/mailman/listinfo/nfdi4ing_hpc-metadata

DORIS: Aufgaben und Maßnahmen

Interfaces, Front-Ends und Repositorien

- Beispiel MediaTUM (TU München) <https://mediatum.ub.tum.de/>
- **“Storage to front-ende interface”**: Verknüpfung Speicher (DSS am LRZ) mit Repositorium (MediaTUM) oder virtueller Forschungsumgebung (TUM Workbench)



Ziel: Speicherung von Metadaten in Repositorium, Speicherung von Forschungsdaten an Rechenzentrum → Nutzer kommt zu den Daten

DORIS: Aufgaben und Maßnahmen

Tests und Testprojekte

- **Testprojekte** in Stuttgart (HLRS), Jülich (JSC) und Garching (LRZ)
- **Transfertools** mit Rechtemanagement,
 - Bsp.: Globus Online (LRZ), GridFTP (HLRS, DKRZ)
- **Virtualisierung** via Container
 - Bsp.: Singularity oder Charliecloud
- In Planung: Best-Practice-Guide und Positionspapier zu Reproduzierbarkeit

DORIS: Aufgaben und Maßnahmen

Lehre, Training und Support

- (TUM-) Workshop im Juli
- „Großer“ Workshop ca. Ende 2021
- Regelmäßig Beiträge NFDI4Ing-Veranstaltungen
- Beitrag zur Lehre (TUM) ab SS 2022
- Software mit „How To“ (to come...): <https://gitlab.lrz.de/nfdi4ing>



→ *best-practices to come ...*

Website: <https://nfdi4ing.de/archetypes/doris/>

Newsletter: https://lists.tu-darmstadt.de/mailman/listinfo/nfdi4ing_taskarea_doris