

Project Presentation
“Publikation Umweltdaten”
(Publication of environmental data)

Prof. Dr. Andreas Hense
andreas.hense@h-brs.de

Deutsche
Forschungsgemeinschaft
DFG

universität**bonn**



Hochschule
Bonn-Rhein-Sieg

Max-Planck-Institut
für Meteorologie



Max-Planck-Institut
für Meteorologie



Dr. Michael
Lautenschlager



universität
bonn

Prof. Dr.
Andreas N.
Hense



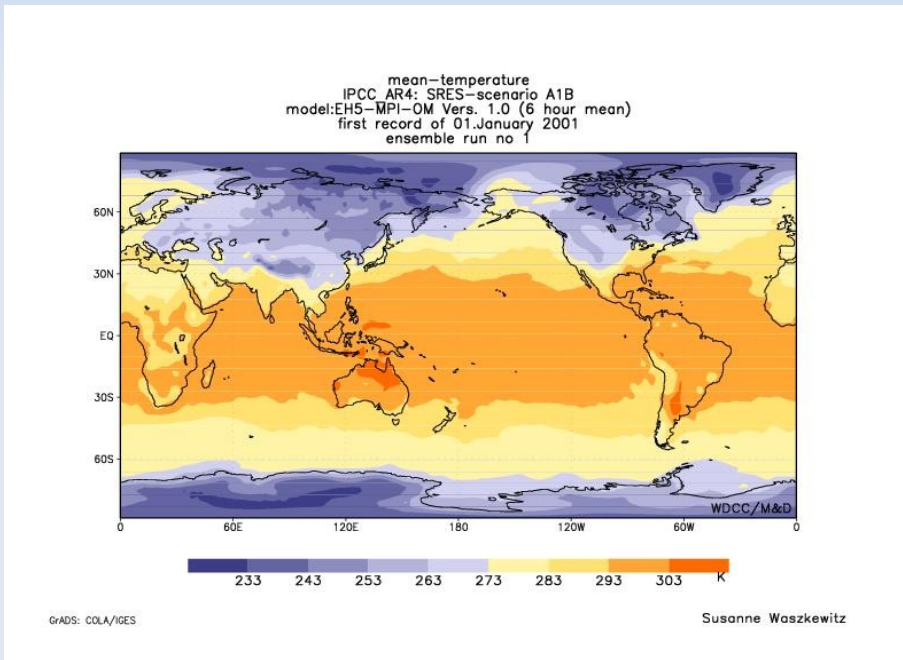
Prof. Dr.
Andreas V.
Hense



Meteorological Data Sources

Climate Simulations

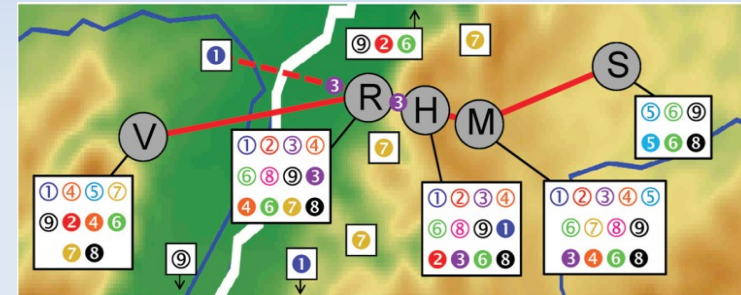
Computed Data from Models: grid-oriented, 2-3 spatial- & 1 time-dimensions



Large amount, but simple structure

Weather Experiments

Empirical Data: various structures in time and space

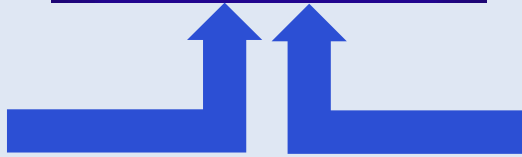
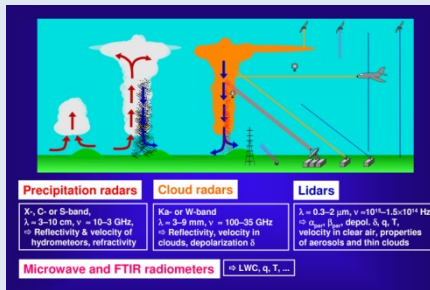
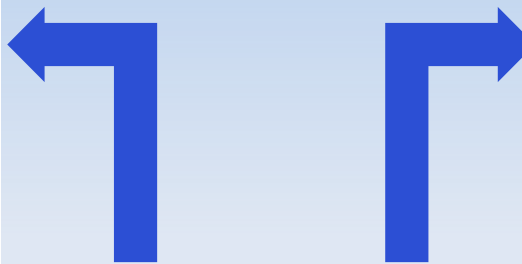
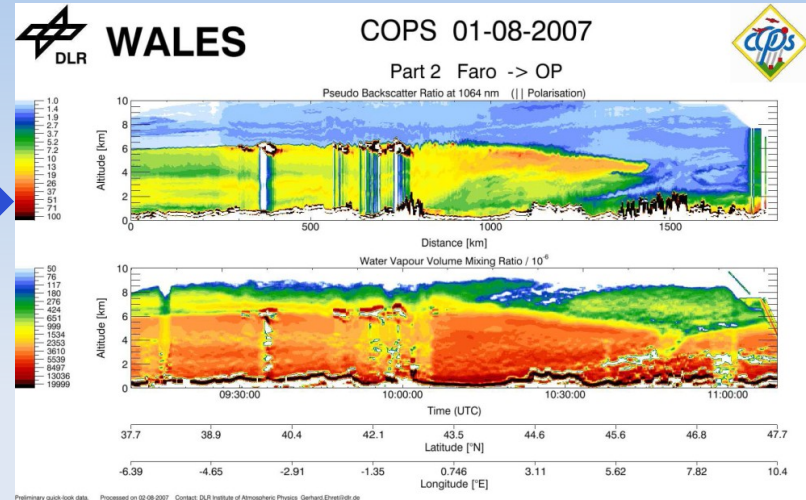
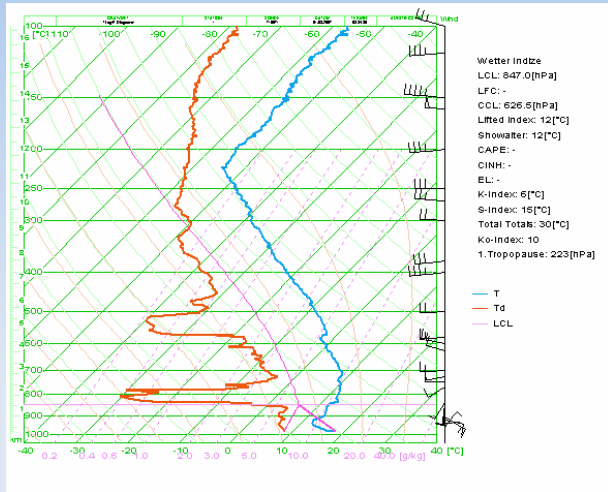


Airborne lidar platforms: DLR Falcon (①③) and SAFIRE Falcon (①)
2 mobile Doppler-On-Wheels (②②)

Legend:	
① Water Vapor Lidar	④ C-Band Polarization Radar
② Temperature Lidar	⑤ Precipitation Radar (other)
③ Wind Lidar	⑥ Micro-Rain-Radar
④ Aerosol Raman Lidar	⑦ Wind Profiler
⑤ Ceilometer	⑧ Wind-Temperature-Radar
⑥ Microwave Radiometer	⑨ Energy balance station
⑦ FTIR Radiometer	⑩ Sodar
⑧ Cloud Radar	⑪ GPS receiver
⑨ Radiosonde station	

Not so big amount but more complex

Weather Experiments



Life cycle of a scientific experiment





Long-term archival

- The WDCC in Hamburg, Germany operates large databases (60 PB) for the long-term archival of data from climate simulation and weather experiments.
- WDCC is controlled by Max-Planck-Society and "Deutsches Klimarechenzentrum" (German climate data processing center)
- Data production: 50 PB/year
- Limit for mass storage archive: 10 PB/year
 - Data with expiration date
- Limit for long-term data archive: 1 PB/year
 - Data without expiration date
- Currently only a very small amount of data is published (approx. 1,5 TB), this is expected to grow significantly.

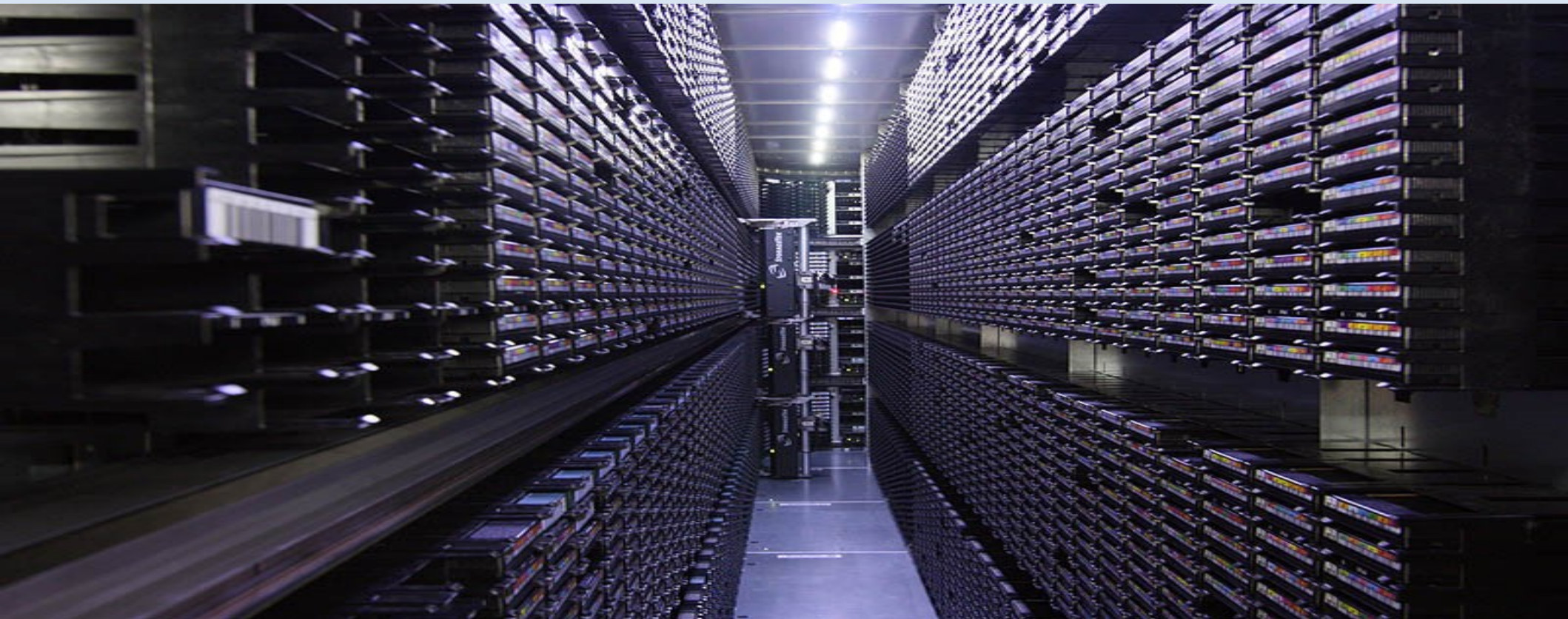
HLRE2 – an IBM Power6 System



- HPC Cluster ("blizzard")
 - IBM p575 "Power6" cluster
 - water cooled, 16 dual core CPUs per node, total: 264 nodes, 8448 cores
 - Total system peak performance: 158 TeraFlops/s
 - Top500: Rank 27 in 06/09
 - 20 TeraByte memory
 - 3 PetaByte GPFS file system (additional 3 PetaByte in 2011)

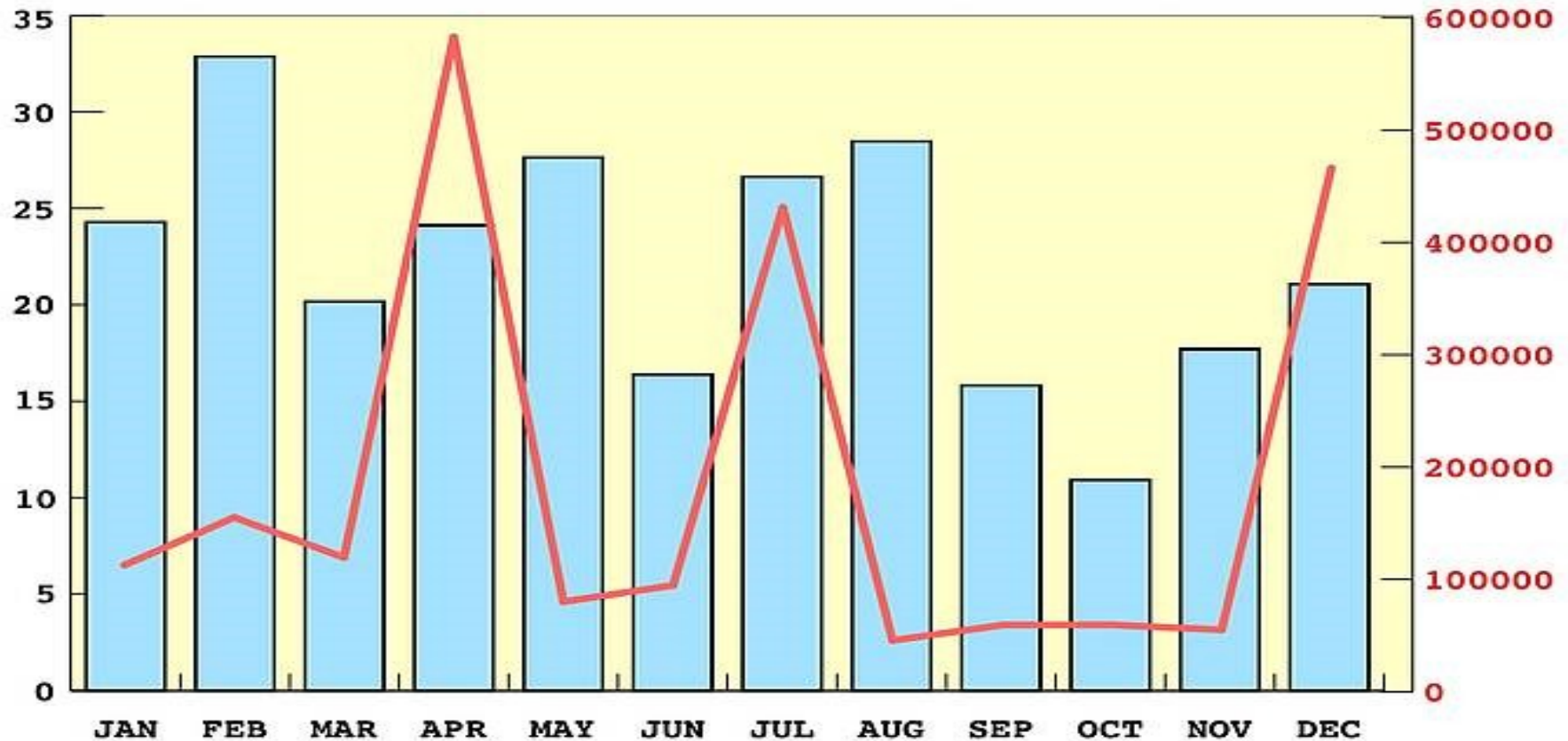
HLRE2 Data Archive: HPSS

- 6 Sun StorageTek SL8500 tape libraries
 - 10 000 media slots per library, 8 robots per library, 73 tape drives
 - total capacity: 60 PetaByte.
- projected fill rate: 10 PetaByte/year



Strong demand for scientific data

Cera Download 2008 in terabyte (left, blue) and counts (right, red) Metadata requests are not counted





Publication via DOI and URN

- Experiments of particular importance will additionally be published with a DOI and a URN.
- The decision making will take part at the WDCC.
- The editing and resolving of a DOI or URN is done by "Technische Informationsbibliothek Hannover" (Technical library at Hannover, Germany).
- Data has to be double-checked before publication.
- That means scientific and technical quality-assurance.
- Most important is a complete and correct meta-data record.

The Publication Process



The Publication Process

Process detail:

<http://academic.signavio.com/p/model/2ec9ea9b426f4f03907b0c609b649203/png?inline&authkey=6feed23d97c47894cffe5177df7649e12df0e2a6197b42e99b7ed9865946>



New STD-DOI publication

1. Entity | 2. General | 3. Authors | 4. Contributors | 5. Relations | 6. Coverage | 7. Details | 8. DOI contact | 9. Submission

Please check the location reference and the spatial coverage and enter/change if necessary:

Location

North America
Northern Black Forest
 Northern Black Forest, Germany
 Northern Mariana Islands
 Northern Hemisphere

Latitude 48.5993 - 48.7109

Longitude 8.2015 - 8.5198

Experiment spatial coverage

Latitude

Longitude

Altitude

Min

Min

Min

m



Max

Max

Max

m



Temporal coverage

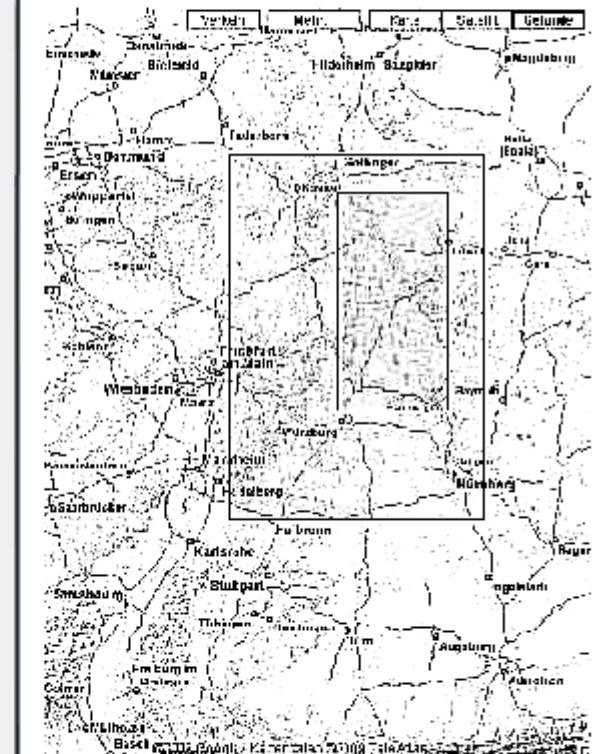
Day Month Year

Day Month Year

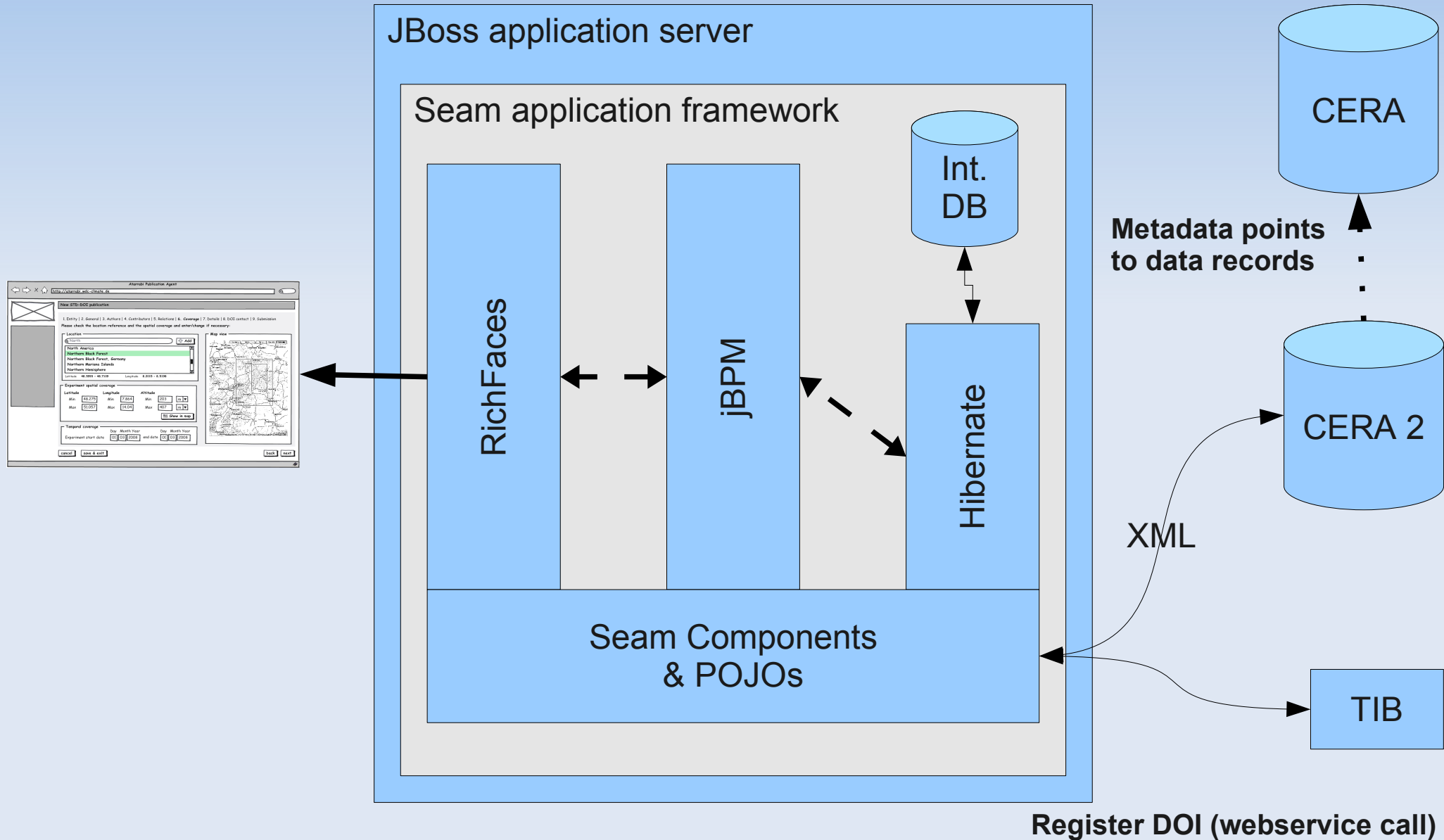
Experiment start date

end date

Map view



System Architecture



Wrap Up

- Development of a standard procedure for publication of weather experiment data.
- Development of a web-based workflow-system to automate this process.
- Integrating the workflow-system into WDCC.
- Preparation for subsequent use of the system in other data-publication tasks in environmental sciences.