

Selecting High Performance Computing and High Throughput Computing Capabilities for Hydro Meteo Research e-Infrastructures

*A. Clematis, D. D' Agostino, A. Galizia,
A. Quarati, IMATI-CNR;*

A. Parodi, N. Rebora, M. Morando, CIMA Foundation;

G. C. Craig, Deutsches Zentrum für Luft- und Raumfahrt;

*D. Kranzlmüller, M. Schiffers, Ludwig Maximilian Universität
and Leibniz Supercomputing Center*



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

Goals

- The efficient and effective use of ICT technologies and infrastructures in Hydro Meteo Research (HMR) activities
- Adequate selection of suitable methods, tools and technologies, considering final user requirements are the key points
- Avoid a discouraging and sceptical approach to new ICT capabilities perceived as non productive and costly optional tools
- This is the scenario that we try to address in the DRIHMS framework



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

Outline

- A general overview of the problem
- How to combine HMR and HPC
- The DRIHMS approach
- An analysis of ICT requirements and perceptions of HM community
- Next steps and conclusions



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

Workshop on HPCCE 2010, 20
September, Crete

A general overview of the problem

- HM science has made strong progress over the last decade:
 - new modelling tools, post processing methodologies, and observational data are available
- Huge amount of data and information is available for the European region: size is against ease of use



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

A general overview of the problem

- An adequate and dedicated infrastructure of computation to connect HM models and observations is often missed
- European efforts in developing e-science Infrastructures provide an ideal basis to share complex HM data sets and tools
 - EGEE (Enabling Grids for E-scienceE), SEE-GRID-SCI (South East Europe -GRID e-Infrastructure for regional e-Science), German C3-Grid



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

A general overview of the problem

- An unprecedented portfolio of resources and services enabling a more thorough HMR
 - Grid & Cloud Computing combined with cluster, and multicore
- Technologies provide a wide set of tools and capabilities:
 - High performance and high throughput computing – HPC & HTC, Data sharing, Interoperability, Workflow management, Visualization



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

A general overview of the problem

- But in some case it is not so clear how to exploit them and what is actually required by HMR

DRIHMS - Distributed Research Infrastructure For Hydro-Meteorology Study

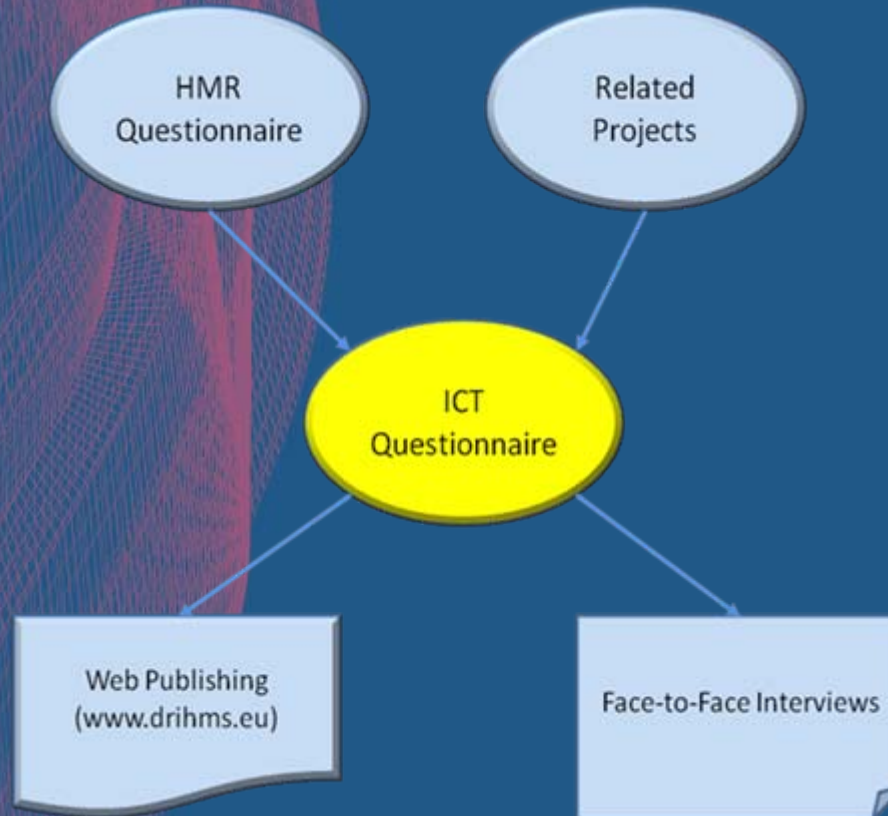
- Aim: define requirements, identify & evaluate gaps in ICT supporting the needs and issues of HMR
- The specific focus of the talk is on HPC aspects



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

The DRIHMS approach



Base Methodology following a Delphi-like approach

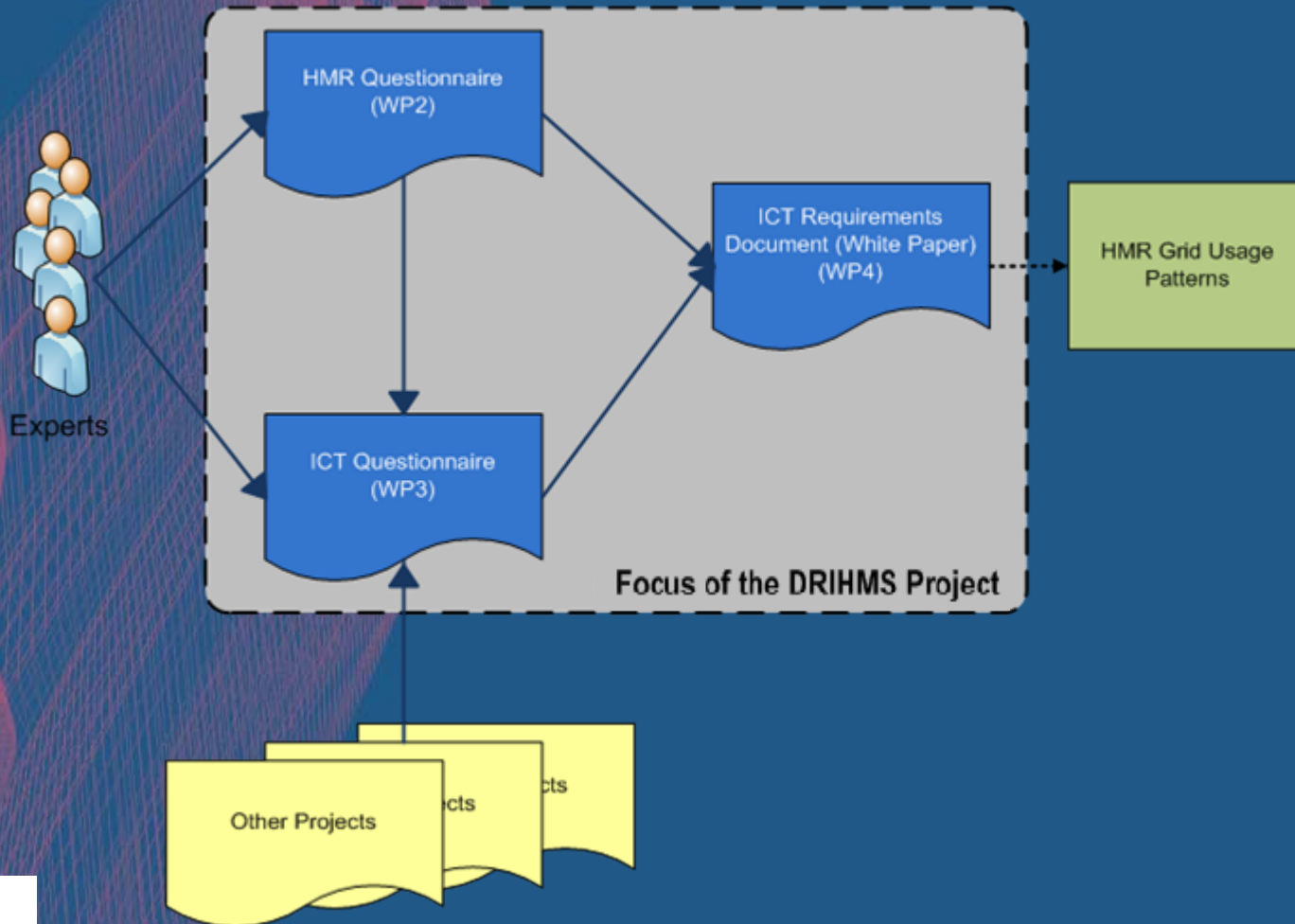


DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

Workshop on HPCCE 2010, 20
September, Crete

The DRIHMS approach



Consultation Process

Where we are

- HMR questionnaire → currently around 200 answers
- ICT questionnaire → currently more than 60 answers
- The process is still going on (ICT) but significant data are already available



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

The HMR Questionnaire

1. Ranking of HMR topics
2. For each topic marking of ICT challenges
3. Real time issues
4. Open questions and free text
 - a) Bottlenecks and problems in current practice
 - b) Possible improvements due to better ICT support
5. Info for adjusting the answers to previous sections to the field of activity in order to get a balanced coverage of different expertise



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

HMR Topics

- Data merging / fusion
 - Probabilistic forecasting
 - Model verification metrics
 - Precipitation downscaling
 - Integration of large data sets for HM prediction
- Related to ICT technical issues



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

Results of HMR questionnaire: rank

- Probabilistic forecasting: top topic
- The model verification metrics the second
- The data merging / fusion and precipitation downscaling lower scores
 - in contrast, with the major limitations outlined

Rank	Full audience
1	Probabilistic forecasting
2	Model verification metrics
3	Data merging / fusion
4	Precipitation downscaling



Results of HMR questionnaire (the free text)

- Data sharing is the main explicit problem
 - Regulation and access, difference between EU & USA
 - Technology related: format, metadata, storage, assimilation, transfer, bandwidth, ...
- HPC is very important but often is a hidden requirement
 - Science related issues >>> Performance related issues



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

Results of HMR questionnaire

- Few existing experiences of
 - systematic use of High performance computing such as Open-MI (cluster)
 - use of Grid resources
- Few indications of possible improvements deriving from the use of new technologies such as GP-GPU



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

The ICT Questionnaire

- Twofold purpose
 - ICT implications of the HMR topics
 - ICT possibilities for their HMR applicability
- To get a better understanding, we asked
 - the *importance* of technological advances,
 - the *maturity* of the same technologies
- Advances in ICT could be HMR benefits



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

The ICT Questionnaire

1. Assessing hot and relevant topics
2. Assessing ICT challenges within each topic special emphasis on cross-cutting issues
3. Computational oriented perspective on a suitable e-science infrastructure for HMR activities
4. Free format feedback starting from current practice experiences
5. Info for adjusting the answers to previous sections to the field of activity in order to get a balanced coverage of different expertise



ICT Topics

- Data management
- High performance computing provisioning
- Workflow management and specification
- Virtual Organization Management
- Portal and user interfaces
- Other



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

Workshop on HPCCE 2010, 20
September, Crete

ICT Topic evaluation

- Hot topics evaluation
 - Data management: important without significant progress in the next years
 - discrepancy again with the limitation
 - HPC: important with significant progress
 - Workflow management: important and significant progress in short term
 - Portals and user interfaces: important and mature, but some progress is expected
 - Virtual organisation management: mature



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

ICT Topic evaluation - HPC

- High performance computing provisioning
 - Significant developments in interoperability for both HPC infrastructures, HTC infrastructures and the combined HPC/HTC infrastructures
 - Adequate match making support services still immature
 - Rudimentary scheduling services and QoS-aware services, no significant improvements is perceived



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

ICT Topic evaluation - HPC

- High performance computing provisioning
 - mentioned the following other topics:
 - development and deployment of general and specific problem solving environments (PSE)
 - support tools to enable the reuse / customization / adaptation by scientists and application developers
 - robustness and fault-tolerance
 - interoperability in heterogeneous Grids (which is different from HPC/HTC interoperability)
 - Cloud computing



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

Results of ITC questionnaire (the free text)

- Following issues have been reported
 - Security issues, as certificate authentication
 - Standardization and interoperability
 - High-quality compilers and programming tools to cope with hardware development – multi-core, FPGA, GP-GPU, ...
 - ...developers seems to develop what they want, and not what is needed...
 - ...there'll be for sure an improvement because it can't be worst than now...



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

Next Steps & Conclusions

- DRIHMS open meeting will be held in October 14th in Genoa
- Analysis of data
- Web consultation is planned
- White paper + final event

→ Participate to DRIHMS activities

→ www.drihms.eu



DRIHMS

DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY STUDY

THE DRIHMS TEAM:



CIMA Research Foundation
www.cimafoundation.org



Institute for Atmospheric Physics (DLR)
<http://www.dlr.de>



Ludwig-Maximilians-Universität (LMU) München
<http://www.mnm-team.org>



Institute of Applied Mathematics and Information Technology (IMATI)
<http://www.ge.imati.cnr.it>