From the workflow point of view: lessons learned from

HD(CP)²

High definition clouds and precipitation
for advancing climate prediction

Kerstin Fieg

Deutsches Klimarechenzentrum (DKRZ)

What is HD(CP)2 ?

High **D**efinition **C**louds and **P**recipitation for Advancing **C**limate **P**rediction

Goal: to advance the understanding of cloud and precipitation processes

- By modelling processes, which are usually parameterized due to limited available compute power

BMBF funded project

- 2 phases: 1st phase started in 2012, 2nd phase started in 2016.
  - 21 partners all over Germany
  - More than 150 scientists, scientific programmers and technical staff involved

Model: ICON−LEM simulates 3 different regions: DE, TA, NA

- 150 vertical levels 3 nested grids (625m, 312 m, 156m)
- 32 mio grid points per level
- 60 TB output, needs 60.000 node*h compute time / simulation day
  - To date ~ 40 days simulated, ~2.5 PB data available

The “workflow task” in HD(CP)2

From the proposal:

Develop an environment, that enables

① the simple and defined execution of experiments

② a common mechanism that takes care of simulation details including compatibilities and software versions

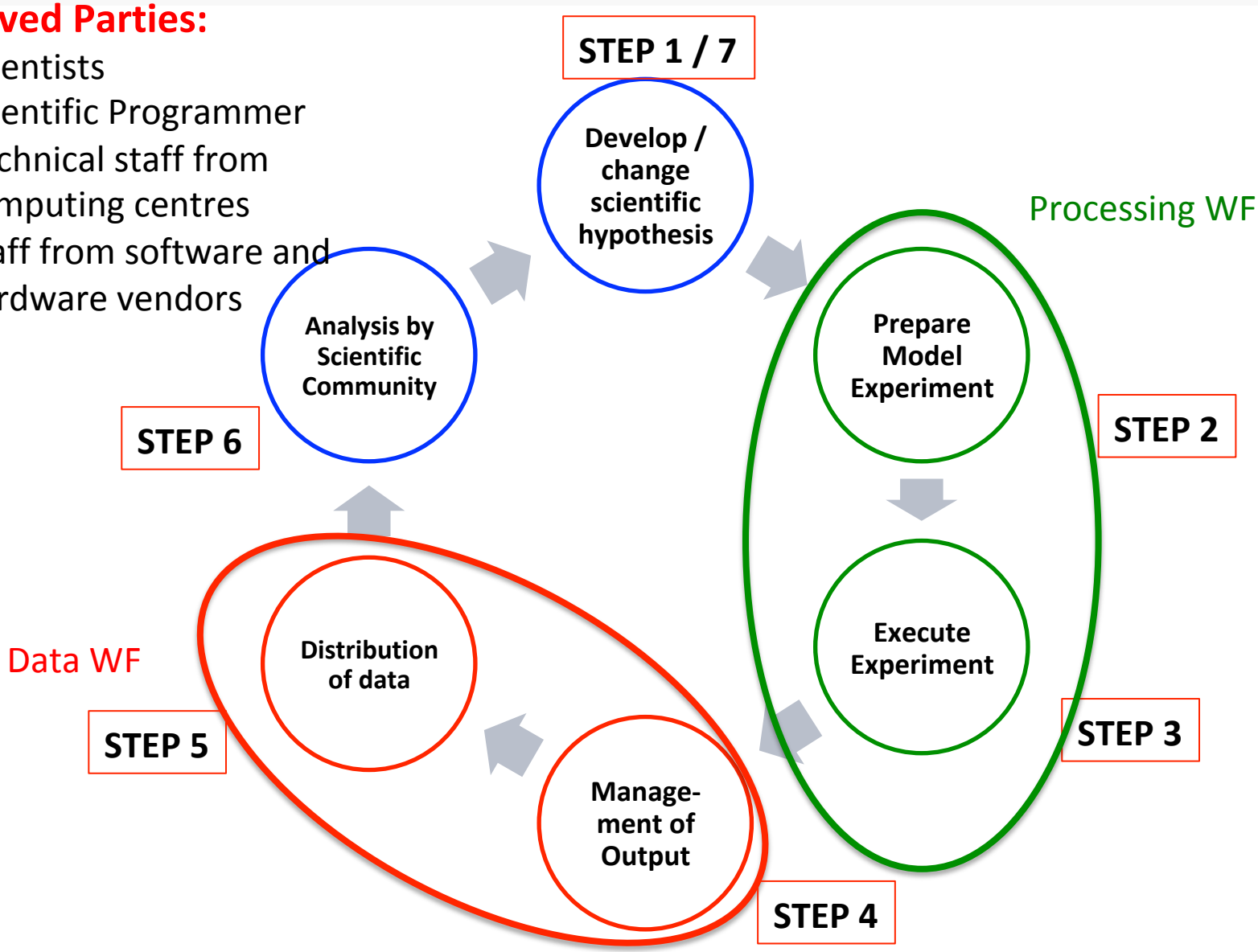
③ a standardized framework allowing description, query and reuse of the results of the performed simulations

That sounds like a adequate, clearly defined goal, that can be broken down to a number issues and challenges.

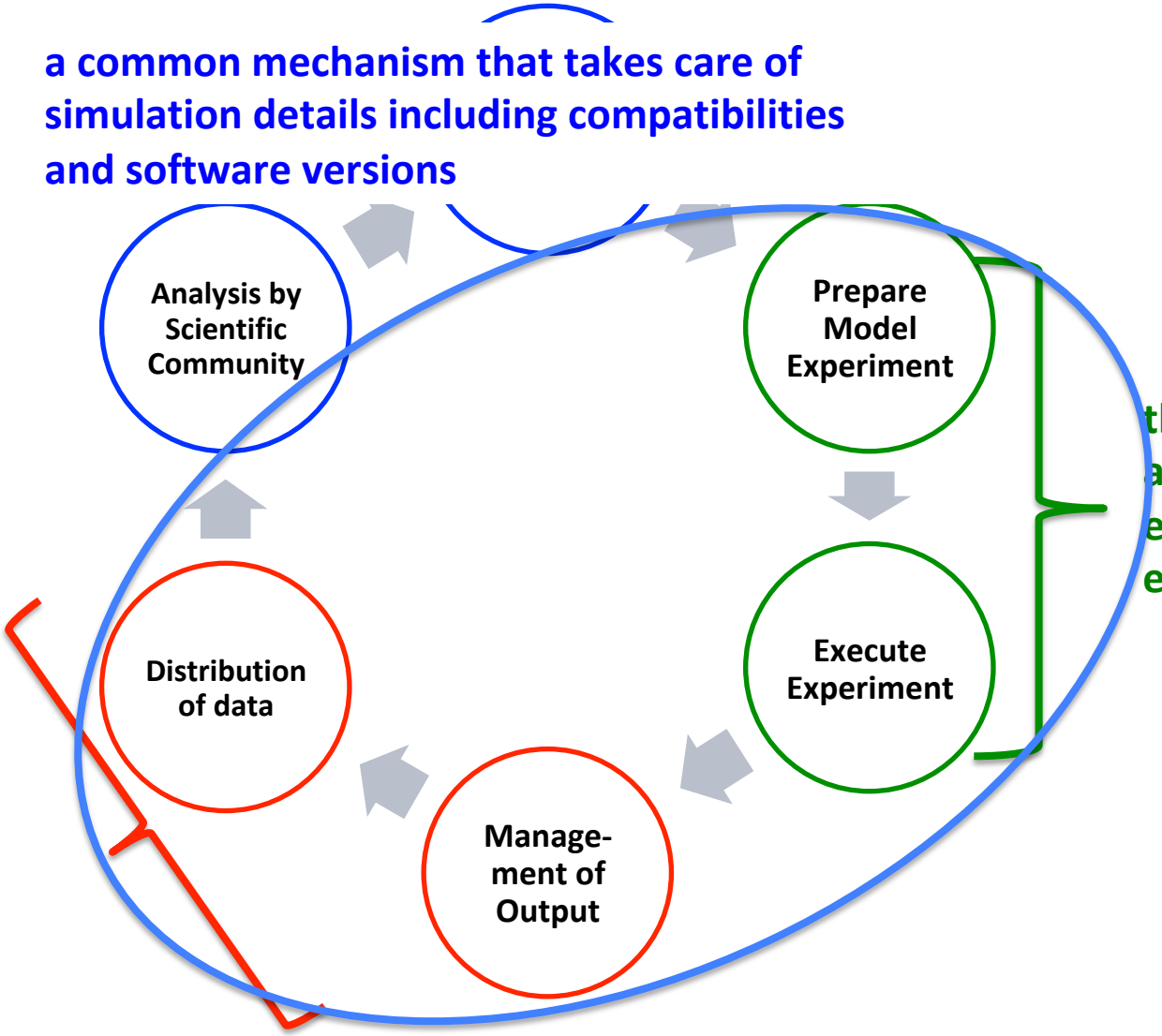
In general: our traditional workflow life cycle

Involved Parties:

- Scientists
- Scientific Programmer
- Technical staff from computing centres
- Staff from software and hardware vendors



Where are the HD(CP)2 tasks located?



In principle the life cycle is valid for research projects (like HD(CP)2) and more operational projects like CMIPs or weather forecasts.

Thus: our idea to deal with the HD(CP)2 workflow:

- Use **Cylc** to build a workflow management system to drive the compete HD(CP)2 workflow
- But it failed.
- Why?

Why?

Technical reasons:

- ICON Model had to be adapted for HD(CP)2 (= SW development, not processing)
  - optimising (DKRZ) and physical adaptation (MPI)
    - Manual interaction and interactive work for 3 – 5 years (still going on)
- HD(CP)2 experiments have huge requirements regarding computing time and data output but performs only a small number of experiments of a similar type
  - Moreover, most of the experiments differ so manual adaption is still needed

Psychological reasons:

- There was no “top down” decision / support for a tool like cylc
  - HD(CP)2 is a research project -> scientists should have freedom to create own research environment
- There was no willingness to use Cylc tool on a voluntary basis
  - Every scientist liked to stick to the own historically grown work structures

No obvious benefit: it was difficult to define a stable repeatable process flow - it was hard to convince (key-)users to invest time in learning a new tool

-> we had to realise, that to make a project run smoothly, workflow has to deal on technical AND interpersonal level.

Nevertheless: challenges HD(CP)2 workflow had to solve

"Technical" challenges:

1. Adapt the code to run on the selected hardware and software
2. Run the agreed set of experiments
3. Restructuring, aggregating and (simple) post processing the output
4. Quality checking, storing, make searchable and downloadable ~ 5 PB data for the community

- The processing WF tasks (1 &2) were finally solved using traditional methods (shell programming)
- The data WF tasks (3 & 4) where handled with semi automatic python scripts.

But because the **technical** issues and tools were broadly discussed in a number of previous talks, I would like to leave it as it is and focus on **interpersonal** issues…

”Interpersonal” challenges to keep the technical part of the workflow going:

- Manage **information flow** between different locations with
  - of 150 students, scientists, technical staff, scientific programmers, vendor staff ….
- Manage different **professional interests** and individual project goals of project members
- Manage different **technical backgrounds**
  - incl. different styles of communication (technical and interpersonal)

It is important to talk about **interpersonal** issues because

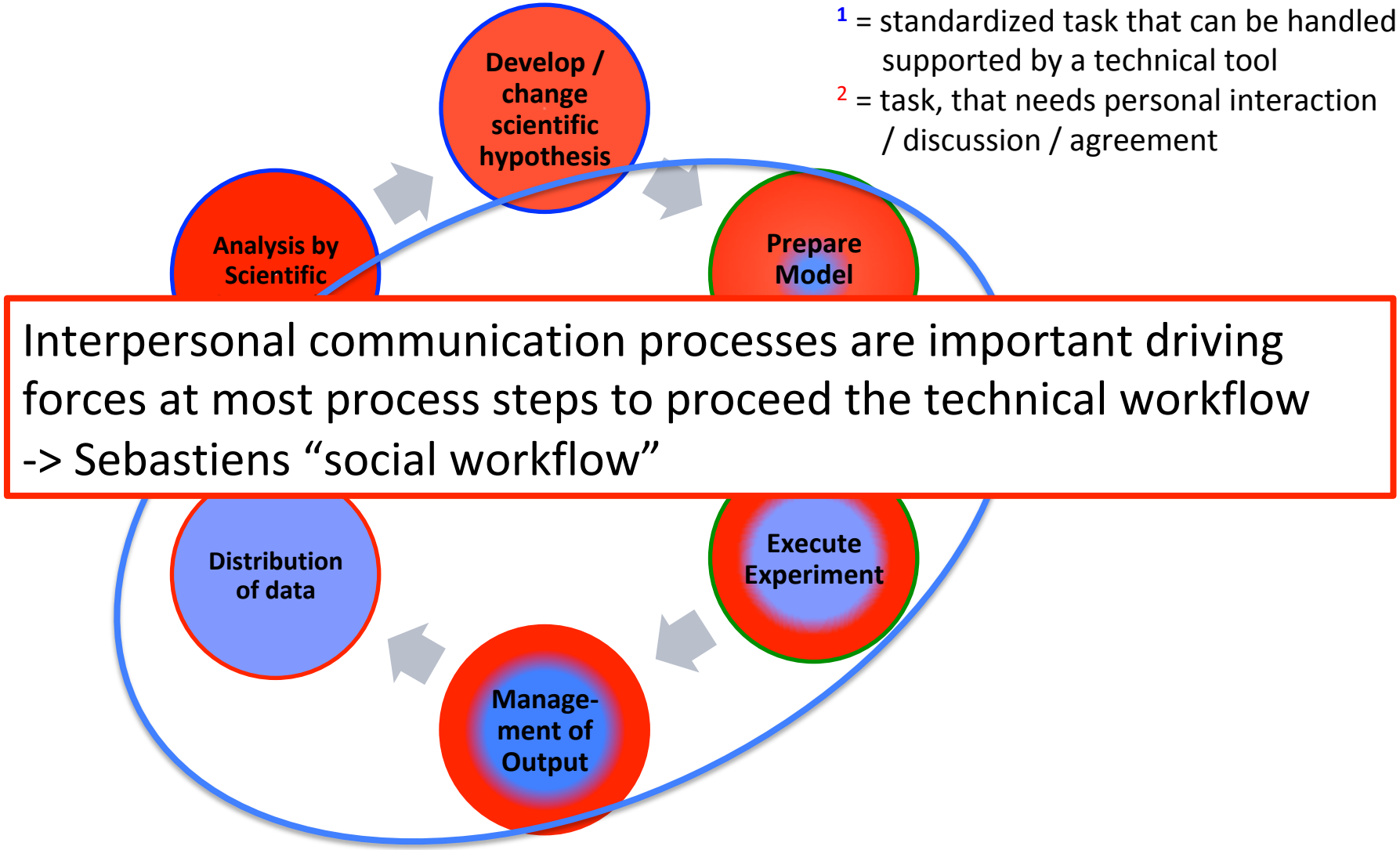
- Usually only minor attention is paid on **interpersonal** issues.
  - project members as well as coordinators often underestimate importance of soft skills (and favour technical / tool knowledge)
- Advantages, disadvantages, challenges and risks at failure are usually not discussed.
- Only interaction between project members with different qualifications, goals and personalities make the technical workflow run at all.
  - Or make it fail!
- Interpersonal communication processes have to be understood as equally important integrative parts of “the workflow” like technical processes

Where comes interpersonal communication into play?

My personal point of view: Technical issue[1] vs. interpersonal issue[2]

[1] = standardized task that can be handled supported by a technical tool

[2] = task, that needs personal interaction / discussion / agreement



Interpersonal communication processes are important driving forces at most process steps to proceed the technical workflow

-> Sebastiens “social workflow”

The methods available (not very surprising and not very sexy)**:**
workshops, assemblies, face – 2 – face – meetings, mail exchange, TC, VC, wikis, redmine ….

Goals of the methods:

- Enabling transparent decision processes
- Sharing information, communicate the issue most efficient, with minimal chance to produce errors
- Guarantee documentation of decisions and processes
- Should consider different scientific cultures, goals and working styles
- Should define clear interfaces to reduce influence of personal disturbances

3 Lessons learned from Phase 1 and Phase 2 of HD(CP)2

Lesson 1

Everyone has a private workflow to process the own part of the work. Nevertheless, certain methods and standards agreed by the community **have to be binding to be used** from all project members.

Means:

- The project management should support the workflow by enforcing the use of methods- if this is not done on a voluntary basis.
- Even if this might be unpleasant for individuals, common methods can be of benefit for the project
  - they define clear interfaces that everyone can rely on

What may happen when agreed structures / methods are not used?

Communication structures

- Unclear, whom to ask, where to put, find and look up information.
- Might lead to internal decision processes between some team members, that finally may influence the work of many.

Technical structures

- Unclear technical interfaces, unclear, how to interact with other team members.

3 Lessons learned from Phase 1 and Phase 2 of HD(CP)2

Lesson 2

Projects can fail because people are not able to communicate the issues in an appropriate way (speaking and / or understanding).

- Implementation of appropriate communication structures **can be key factors** to make a project a failure or a success

What does it mean when communication fails?

People may not communicate at all.

People may decide on topics they are not responsible for.

People may decide on topics and do not communicate the decision to the responsible partner / team member.

People think they talked about the same topics but they did not.

People talked, agreed what to do and finally do something different (….because of any − maybe good? − reason) and do not feedback that something changed.

People discussed, agreed but misunderstood.

3 Lessons learned from Phase 1 and Phase 2 of HD(CP)

Lesson 3

Only together the project can be realized!

The project management **should support the workflow** by

- appreciating all members of the project **as equally important** (scientific programmers, technical staff, staff from vendors as well as scientists). Because everyone brings his / her unique knowledge and different view on the issues.

What may happen when groups feel unequal (- inferior / + superior) ?

- No “team spirit” (+ /-)
- Frustration (-)
- Loss of motivation to engage for the project (-)
- Might reduce commitment to reach community project goals (+ /-)
  - Might lead to focus more on personal project goals (thesis, paper) (+)
- Might hamper the free communication flow between the groups (+/-)

Conclusion

- To solve the technical issues of the workflow, tools are available and may be suitable to solve the problems.
- The discussion on tools and methods to support interpersonal communication issues (soft skills) and information flow processes are usually not in the project focus.

But it is obvious, that workflow cannot be cut down to optimal use of technical tools. So, why do we often neglect the interpersonal aspect?

A broader awareness on the importance of interpersonal communication issues at team member and management level might help to ease the project workflow. (?).