ALFRED-WEGENER-INSTITUT
HELMHOLTZ-ZENTRUM FÜR POLAR- UND MEERESFORSCHUNG

Workflow Treatment in C3Grid

Bernadette Fritzsch
Computing Center
Alfred Wegener Institut for Polar and Marine Research

&

C3-Team

C3Grid Team

Collaborative Climate Community Data and Processing Grid – C3Grid

DKRZ
DEUTSCHES KLIMARECHENZENTRUM

AWI
ALFRED-WEGENER-INSTITUT
HELMHOLTZ-ZENTRUM FÜR POLAR- UND MEERESFORSCHUNG

Freie Universität Berlin

WDC MARE

WDC CLIMATE

Universität zu Köln

rrzk
Regionales Rechenzentrum

CSC
Climate Service Center
Germany
Geesthacht

IGM - Bereich Meteorologie Universität zu Köln
Institut für Geophysik und Meteorologie

P I K
POTSDAM INSTITUTE FOR CLIMATE IMPACT RESEARCH

technische universität dortmund

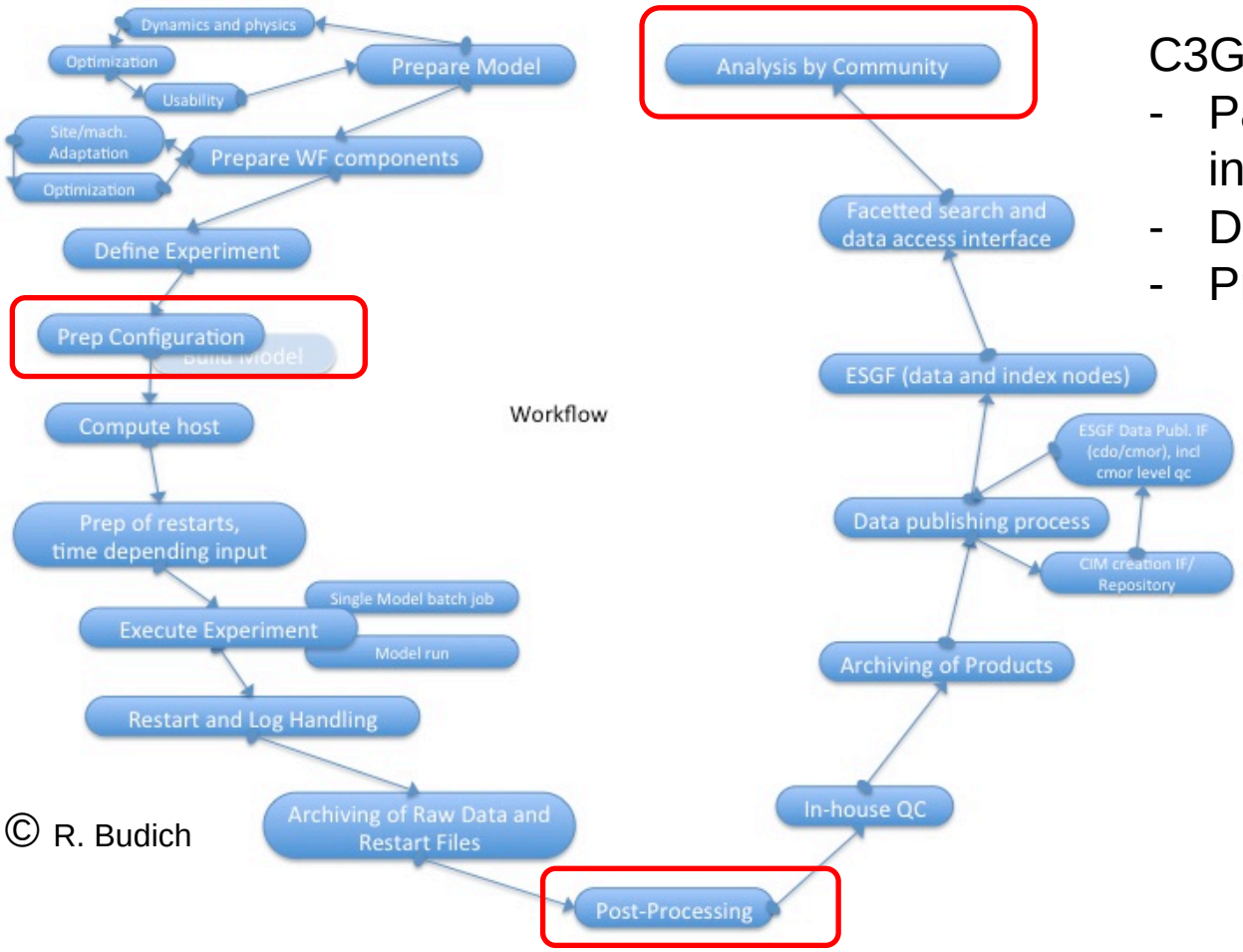
inf
Institut für Roboterforschung

ZIB

DLR

The World Data Center for Remote Sensing of the Atmosphere

WFs in C3Grid

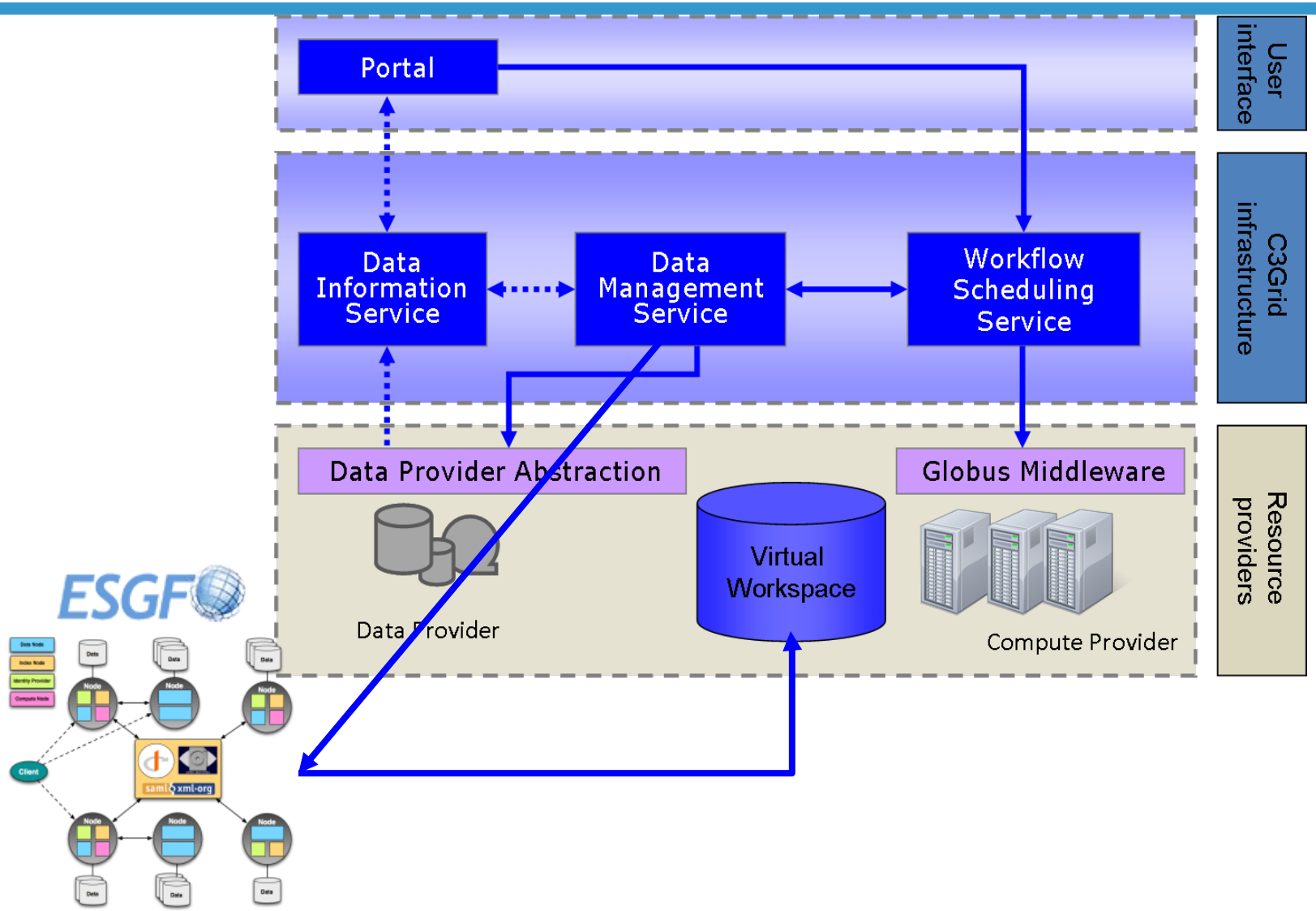


© R. Budich

C3Grid-WFs

- Part of general WF in ESM
- Diagnostics
- Preprocessing

C3Grid Architecture



Workflow execution in C3Grid

Select DP, CP

Portal

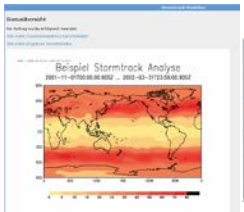
WSS

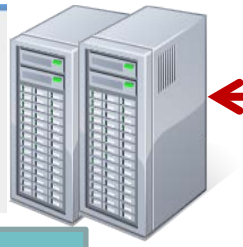
DMS

WSL

User

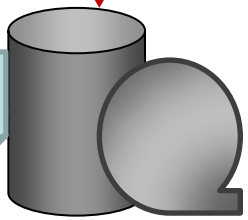
- WF
- Datasets
- Control parameter
- Region/time





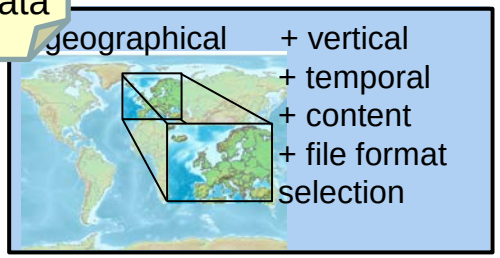
Data

Meta Data



Data

Meta Data



WSL

- Based on Job Specification Language (OGF) with some extensions for
  - Data extraction and ESGF Data staging
  - Data publishing
  - Conditional data transfers
- Description of WF
  - Tasks
  - Dependencies
- Scheduler selects in co-scheduling with DMS optimal CP and DP
  - data transfers (replica)
  - Time to solution
  - Inserts file transfers automatically between choosen resources

WF development

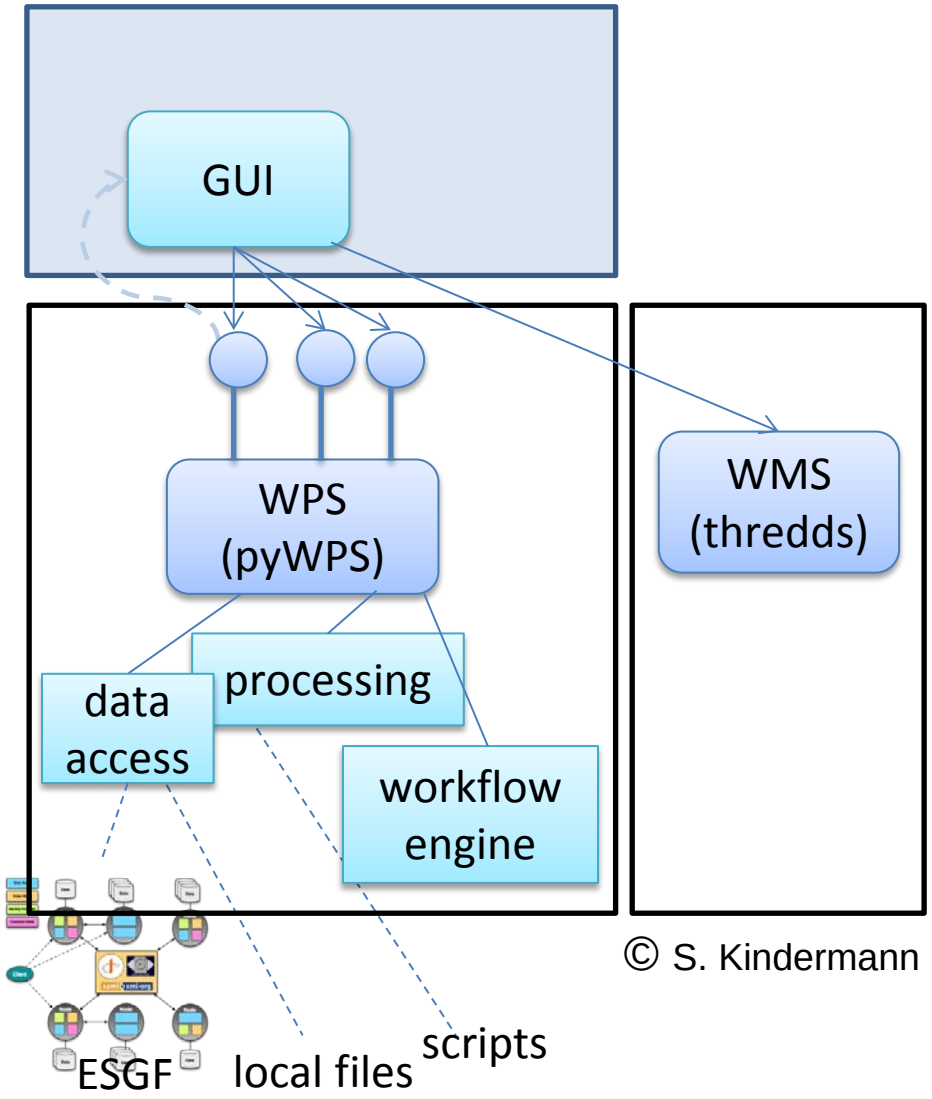
- Local development of a prototype
  - scripts, Fortran programms, …
  - local datasets
- Rollout - Installation of WF modules on CPs
- WF description in prototype WSL
  - preselected datasets in distributed data federation
  - prototypic control parameter
  - can be executed by WSL upload in portal
  - tests
- WF publication in C3Grid portal
  - generalization of prototype WSL → WSL generator
  - mask for input parameter
  - criteria for data search

Workflows

Name	Function	Name	Function
EADY	Eady Growth Rate	Stormtrack	Stormtrack Diagnosis
TroughIdent	Identification of PV	PVTroughs	Tracking of PV Troughs
QFLux	Humidity Flux	CT	Cyclone tracking
CWT	Circulation Weather Types	CAPE	Convective available potential energy
GRET	Grass reference evapotranspiration	MMME	Multi Model Multi Ensemble
SEGFLO	Abundance of Segetal Flora	Model Verification	Interpolation, Statistics
LBC	Preparation of IC/BC for regional model CLM	GIS	Data preparation for GIS applications
Low/high-Vis	Visualisation		

- diagnostic workflows
- data pre/postprocessing in modelling
- visualisation

Alternative path: WPS



© S. Kindermann

- ClimDaPS - Technical details → see Stephan
- Uses OGC conforming standards and interfaces
- C3Grid, LSDMA, ExArch projects
- Example WFs from CSC
  - Population dynamics of Anopheles Gambiae
  - Number of Segetal flora species
  - Species distribution modell
  - Evapotranspiration
  - Calculation of climate indices

Lessons learned

„native“ C3Grid WF implementation.

+ High potential optimizations (data traffic, time, replica management ..)
− Special WSL
− complicated middleware
− Long time for integration

− WFs in portal fixed
− WF as „Black Box“
+ WF as „Black Box“
+ Many checks in portal (input)
➔ Education
➔ Knowledge discovery

OGC-WPS WF implementation

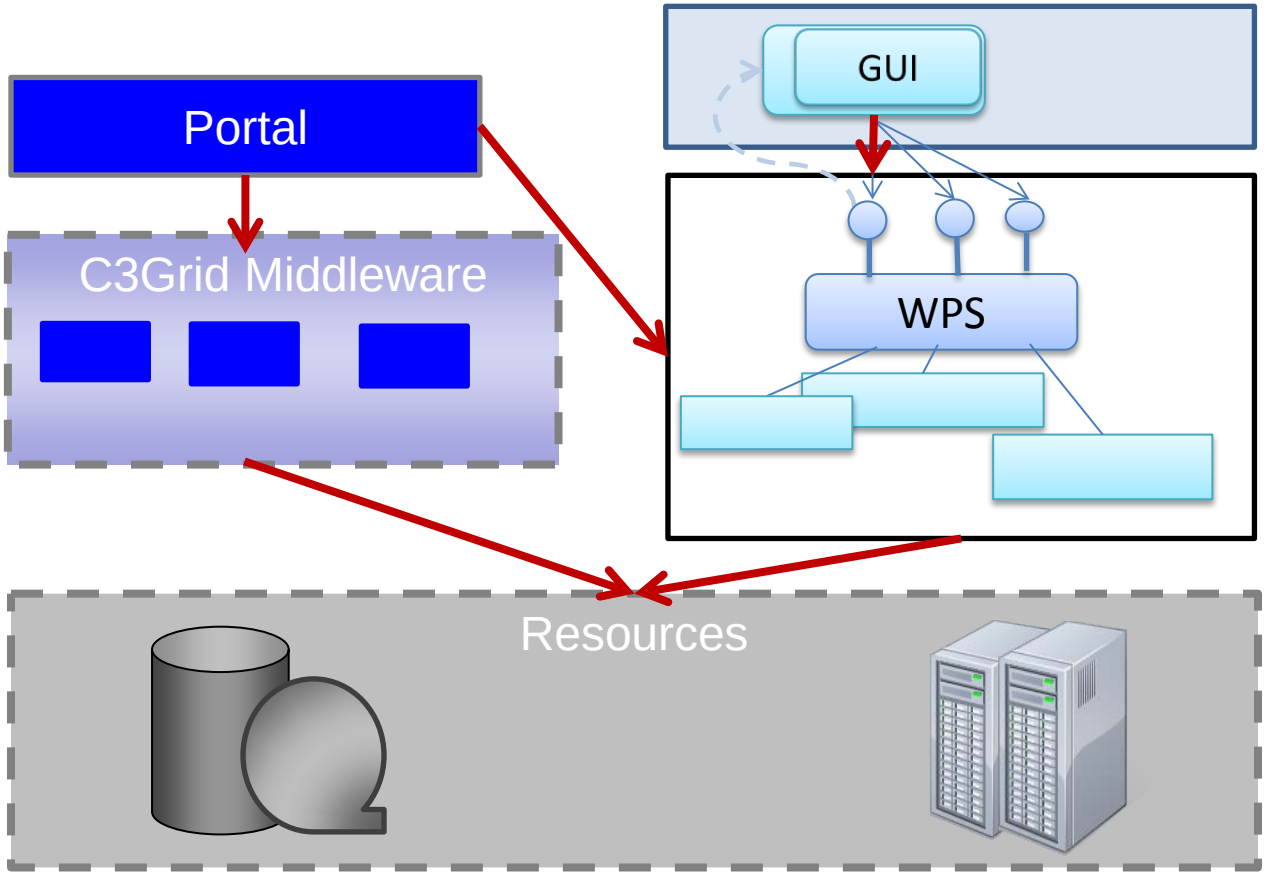
+ Lower entrance barrier for WF developer
− Security

+ More flexibel in WF: Scientist can use his own „flavour“ of a WF
− Needs more knowledge in WF use

➔ For sophisticated users

Perspectives

- New WFs
- Integration of WPS-WF into C3Grid portal



Thank you for attention!