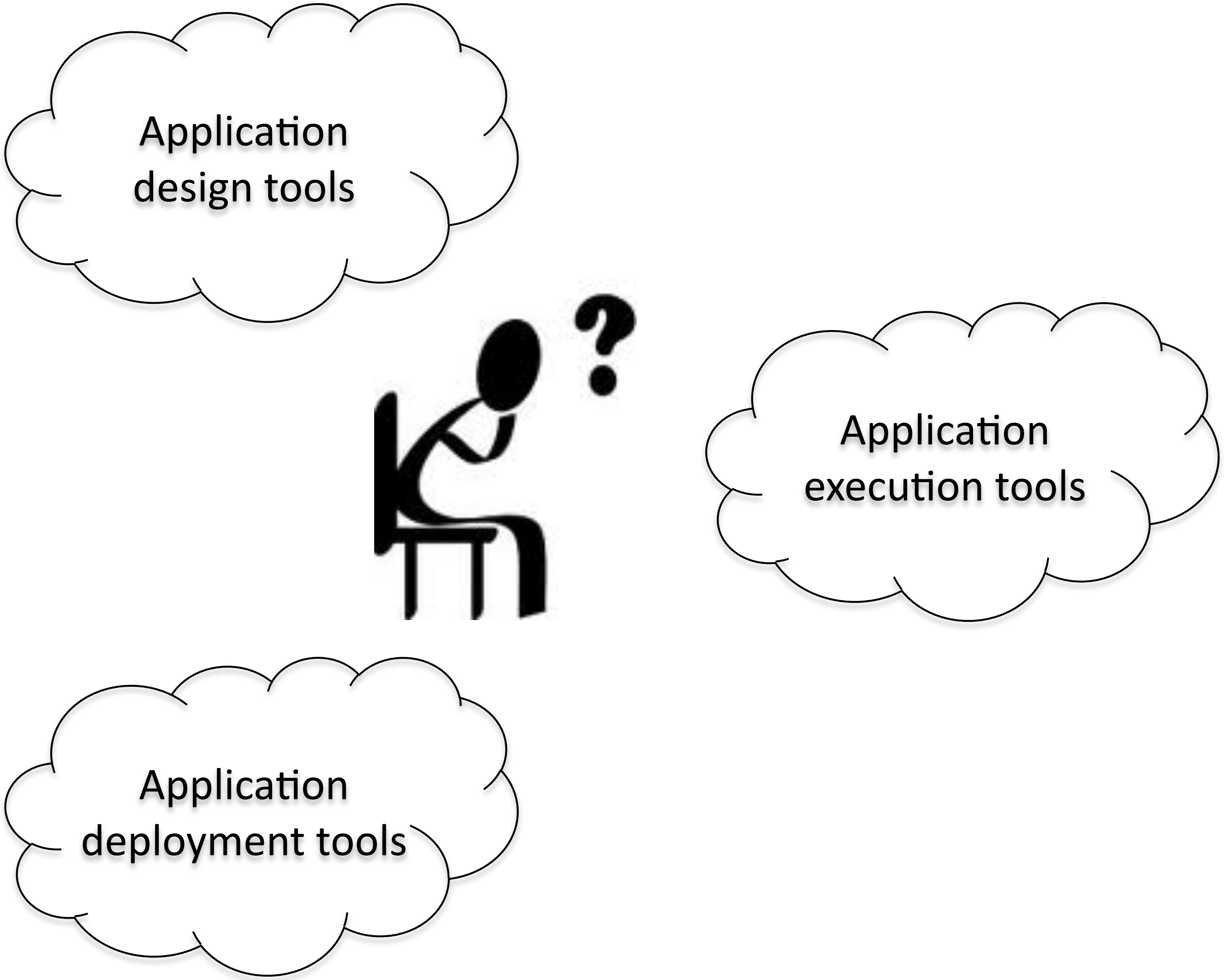
A Comprehensive Framework for Scientific Applications Execution on Distributed Computing Infrastructures

Javier Rojas Balderrama, ***Tram Truong Huu***, *Johan Montagnat*

I3S/ CNRS – Sophia Antipolis

End-users' concerns

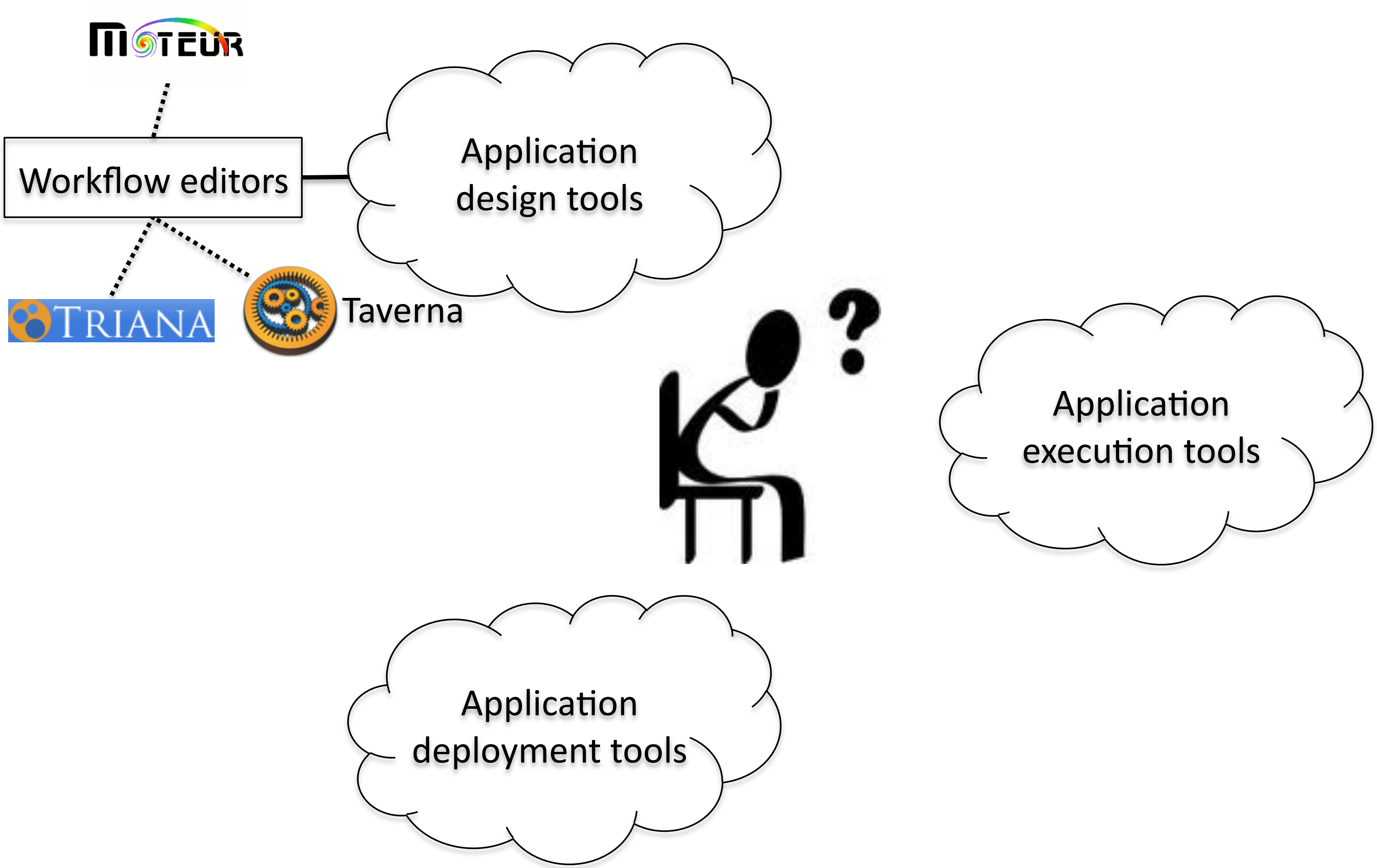


Application design tools

Application execution tools

Application deployment tools

End-users' concerns



Workflow editors

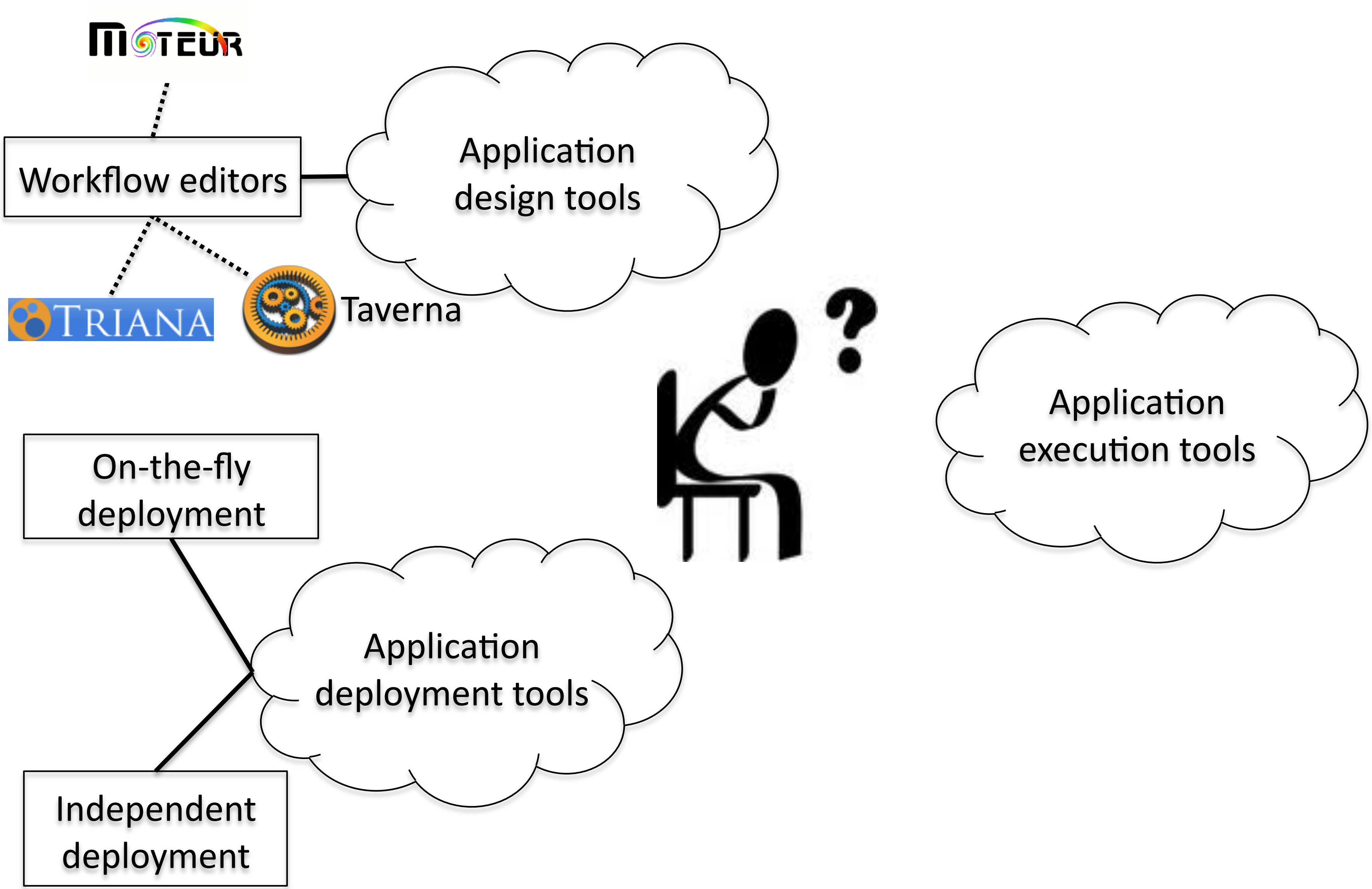
Application design tools

Taverna

Application execution tools

Application deployment tools

End-users' concerns



Workflow editors

Taverna

Application design tools

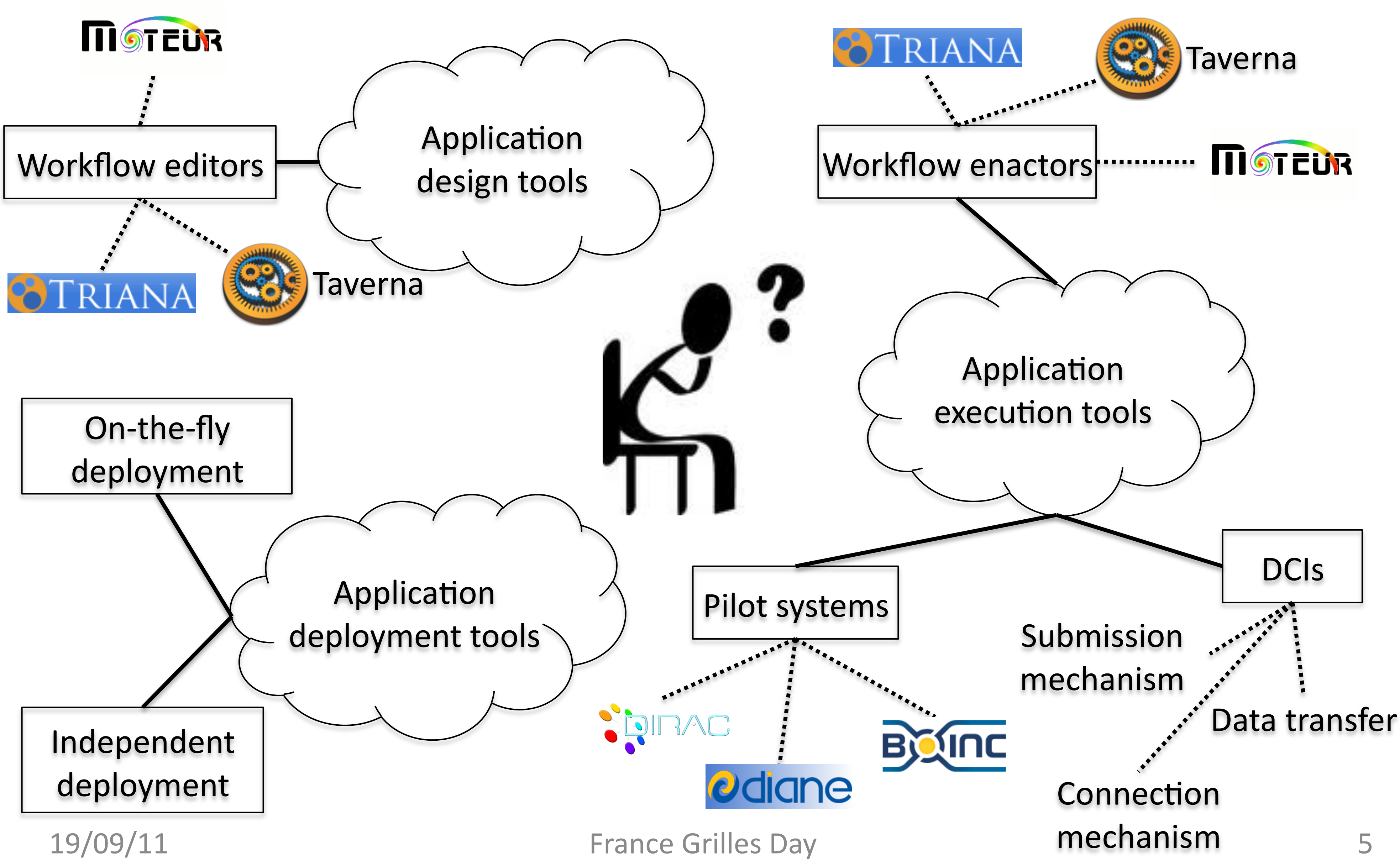
Application execution tools

On-the-fly deployment

Independent deployment

Application deployment tools

End-users' concerns



MOTEUR

Workflow editors

Application design tools

TRIANA

Taverna

TRIANA

Taverna

Workflow enactors

MOTEUR

Application execution tools

On-the-fly deployment

Application deployment tools

Independent deployment

Pilot systems

DIRAC

diane

BOINC

DCIs

Submission mechanism

Data transfer

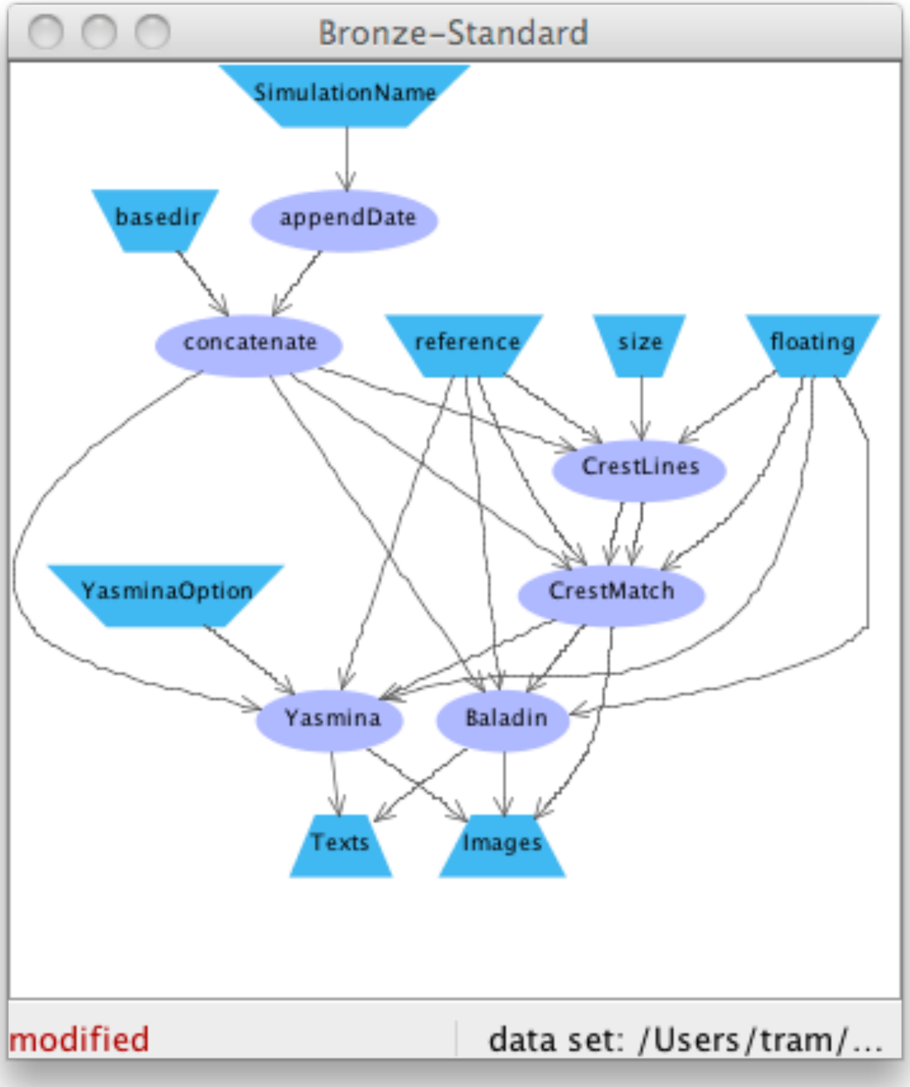
Connection mechanism

Work's objectives

- Main questions:
  - How end-users can design and deploy their application(s)?
  - How end-users can execute their application(s) on DCIs?
- Propositions:
  - Collection of tools used in design, deployment and execution
  - Integration of these tools in a comprehensive framework

MOTEUR workflow engine

- Moteur client
  - Allows users to design applications composition at abstract level
  - Connects users with Moteur server and the rest of framework to manage execution
  - Supports GUI and CLI

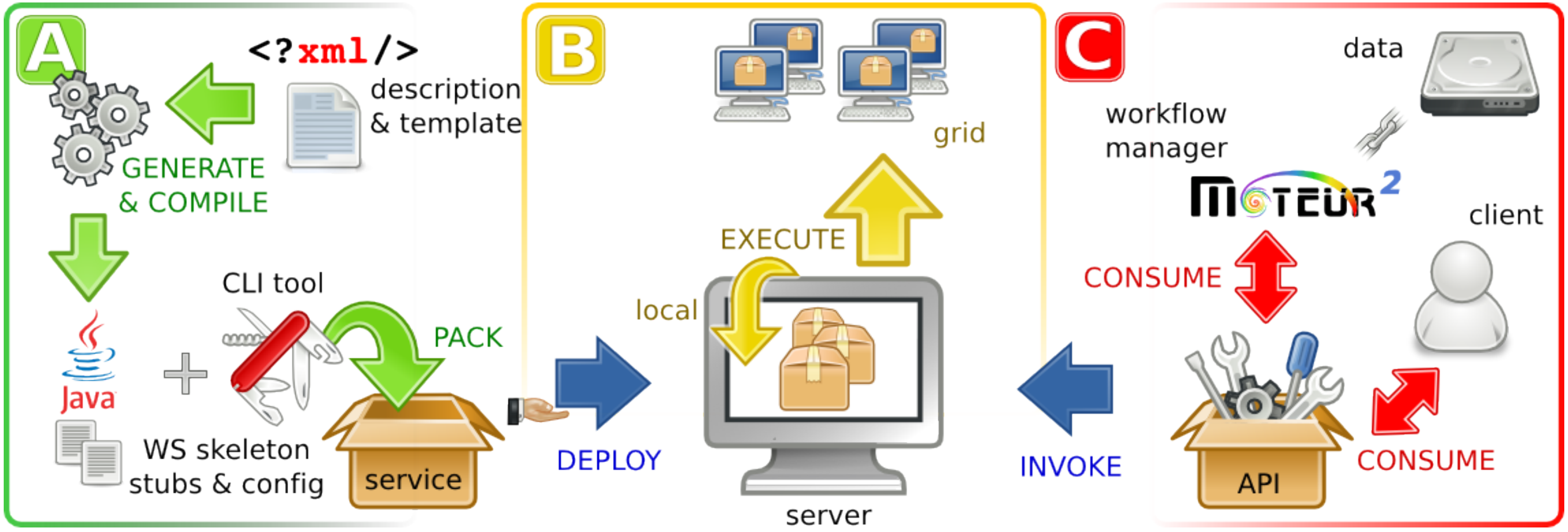


MOTEUR workflow engine

- Moteur server
  - Invokes the execution of workflow services locally or remotely through a generic interface.
  - Enables parameter sweep on input dataset

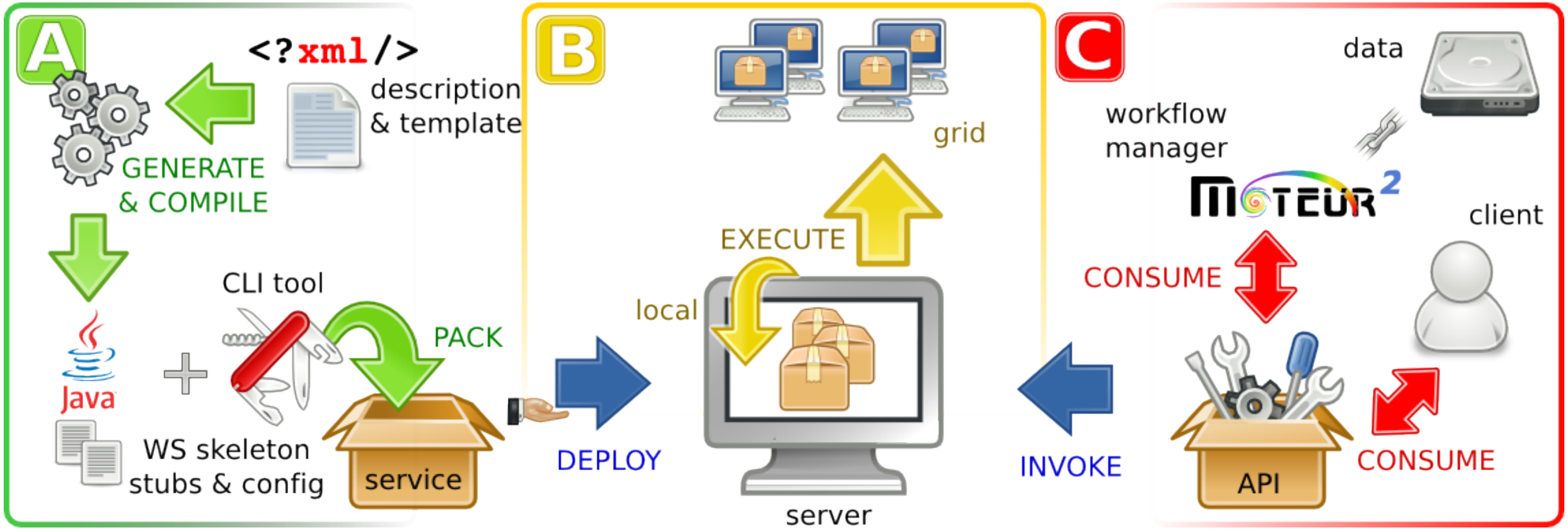
Jigsaw for CLIs

- A: Creation of services
  - Mapping of complex I/O structures
  - Strong data types
  - Dynamic generation & hot-deployment



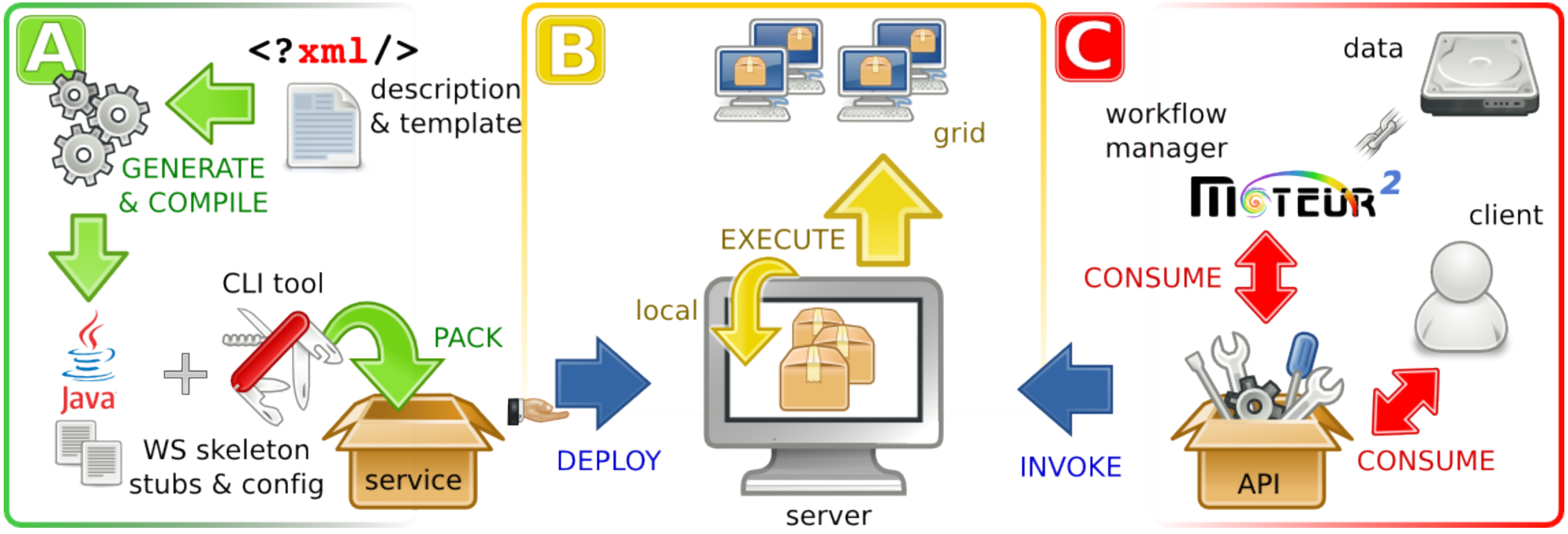
Jigsaw for CLIs

- B: Libraries for transfer and execution
  - Support of local and DCIs executions
  - Applications staging and data transfers
  - Integration of non-functional concerns



Jigsaw for CLIs

- C: Application Programming Interface (API) for direct execution or external integration



GASW – DCIs execution adaptor

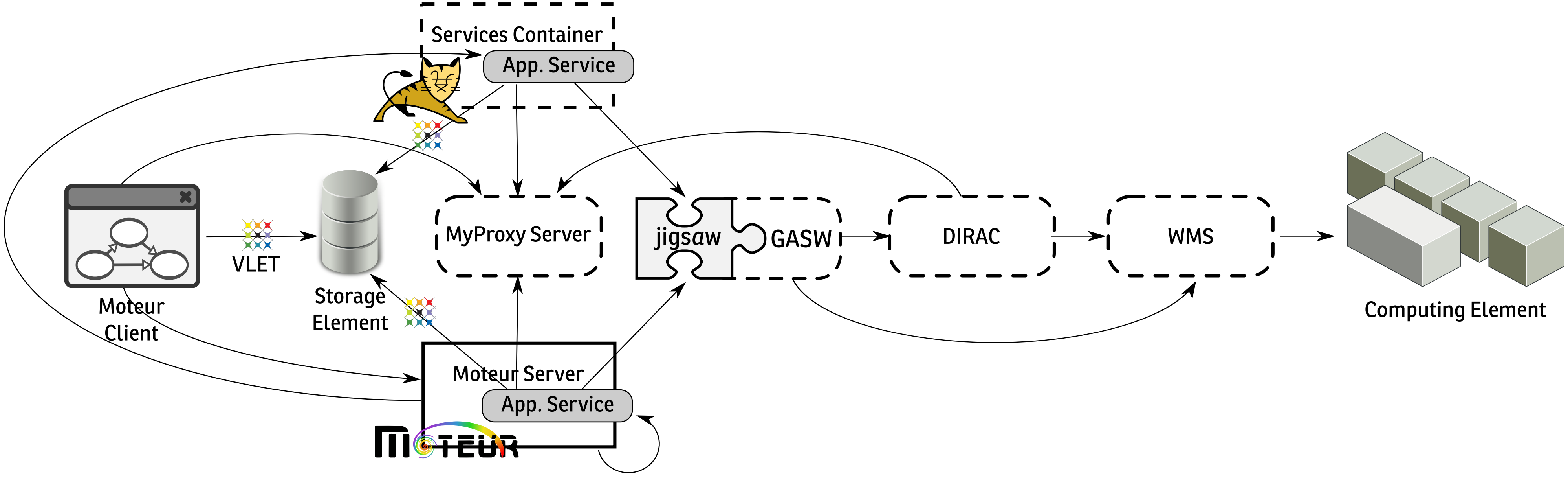
- Dispatches execution of workflow services on multiple DCIs directly or through pilot systems
  - EGI
  - GRID'5000
  - PBS
- Monitors the execution of services
  - Pooling the execution status from DCIs
  - Storing execution information for statistics

Integration of other tools

- DIRAC framework speeds up the execution by using pilot jobs for resources reservation
- Vlet Toolkit provides unified view of heterogeneous file systems such as HTTP, FTP, LFN, local FILEs
- MyProxy Server provides fresh proxy for authentication to any DCI/service.

Architecture of provided framework

- Integration at two levels:
  - Front-end level provides interface for gridifying and invoking scientific applications
  - Back-end level gives a transparent access to multiple DCIs



Conclusions and future work

- Conclusions
  - Provides a comprehensive framework for application lifecycle management
  - Enables performing experiments on different DCIs or local servers
  - Gives a transparent access to those DCIs
- Future work
  - Support workflow interoperability with other scientific workflow languages
  - Support additional DCIs