Using computing Grid for research: HEP and beyond

- *Use the Grid for research:*
  What for and how?

- Who does it?

- Grid as a research tool
  vs.
  Grid as a research object

Wojciech Wiślicki, A. Sołtan Institute and ICM Univ. Warsaw

Using computing Grid for research: HEP and beyond

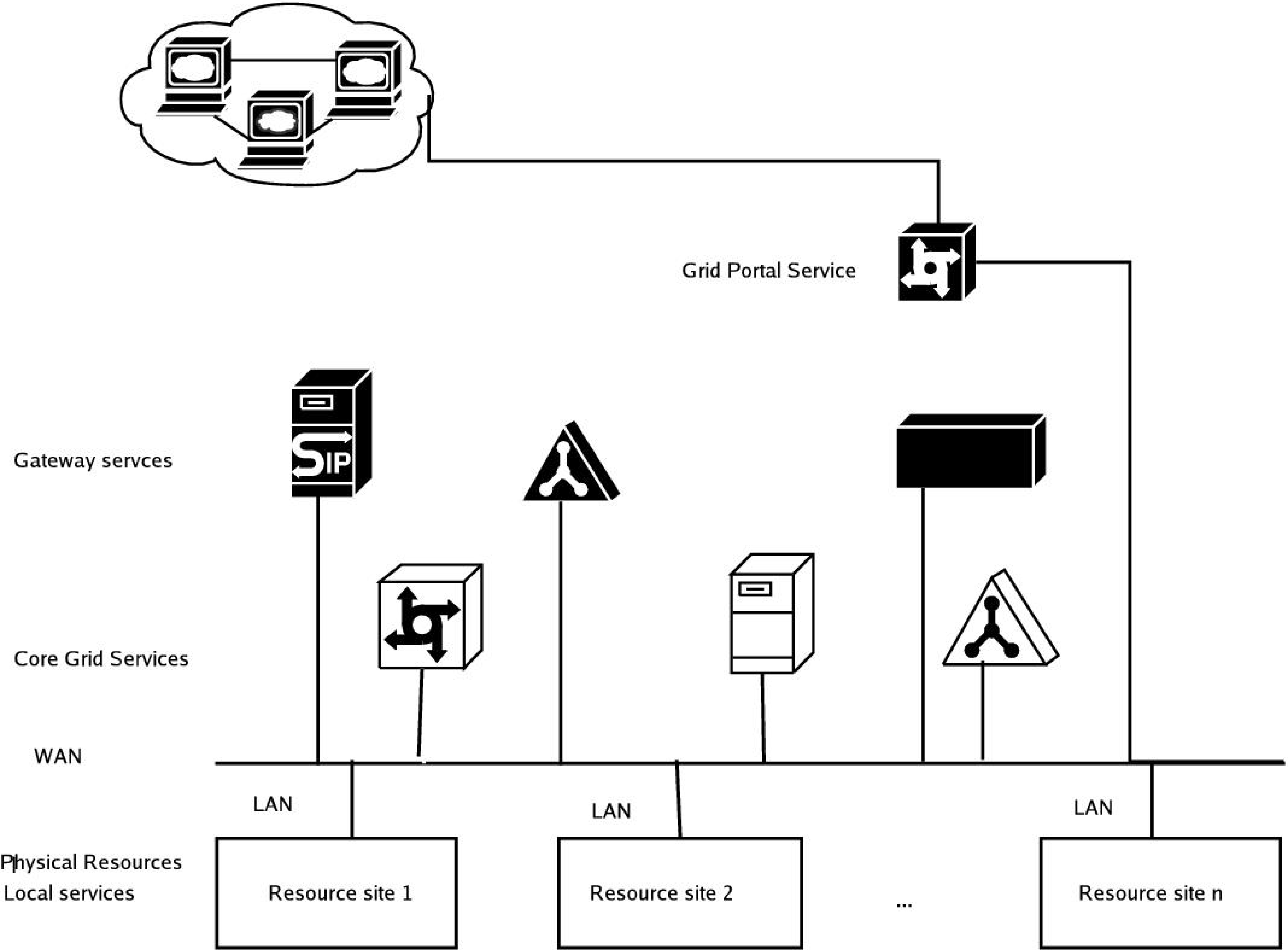
- Daily practice from both the engineer's (viz. system management, maintenance and development) and the scientist's (viz. non-commercial end-users) viewpoint

- Business perspective: security and autonomous tools

- Other issues

Usual way of doing things: off-line batch processing on the Grid

- Physical resource layer is really hidden behind core Grid services
- In some cases only, there do exist user access portals
- In HEP applications, mainly massive, asynchronous data-intensive processing of segmented data, very rare use of MPI on the Grid


Grid Portal Service
Gateway servces
SIP
Core Grid Services
WAN
LAN
LAN
LAN
Physical Resources
Local services
Resource site 1
Resource site 2
...
Resource site n

Usual way of doing things: off-line batch processing on the Grid

- Limited use of gateway services (user metadata catalogs, application resource catalogs, workflow engines), although tools do exist in many cases (e.g., LCG File Catalog, Replica Location Service,..)

- Semi-manual control of resource brokers

Usual way of doing things: off-line batch processing on the Grid

- No job migration; but no high-level checkpointing service; under development within CoreGRID NoE
- Rare, or no at all, use of workflow engines (PGRADE, CoreGRID developments)
- Grid monitoring should work for users; coupling to RBs (Ganglia-based monit./predictor developed in CrossGrid)

Desired line of development

Towards Service-Oriented Architecture (SOA), where:

- Applications are distibuted over the net and each function is a service, announcing itself on the network and accessible from everywhere
- Data are virtualized
- Worklows are dynamically designed and re-designed, according to needs

Interactive applications

- What does *interactivity* mean for most of today's applications on the Grid?
- How does it relate to real interactivity on vector machines or local farms?
- *Pros* and *contras* for interactive applications on the Grid with the system-human interaction in a loop
- Interactivity *vs*. autonomy

Typical HEP application

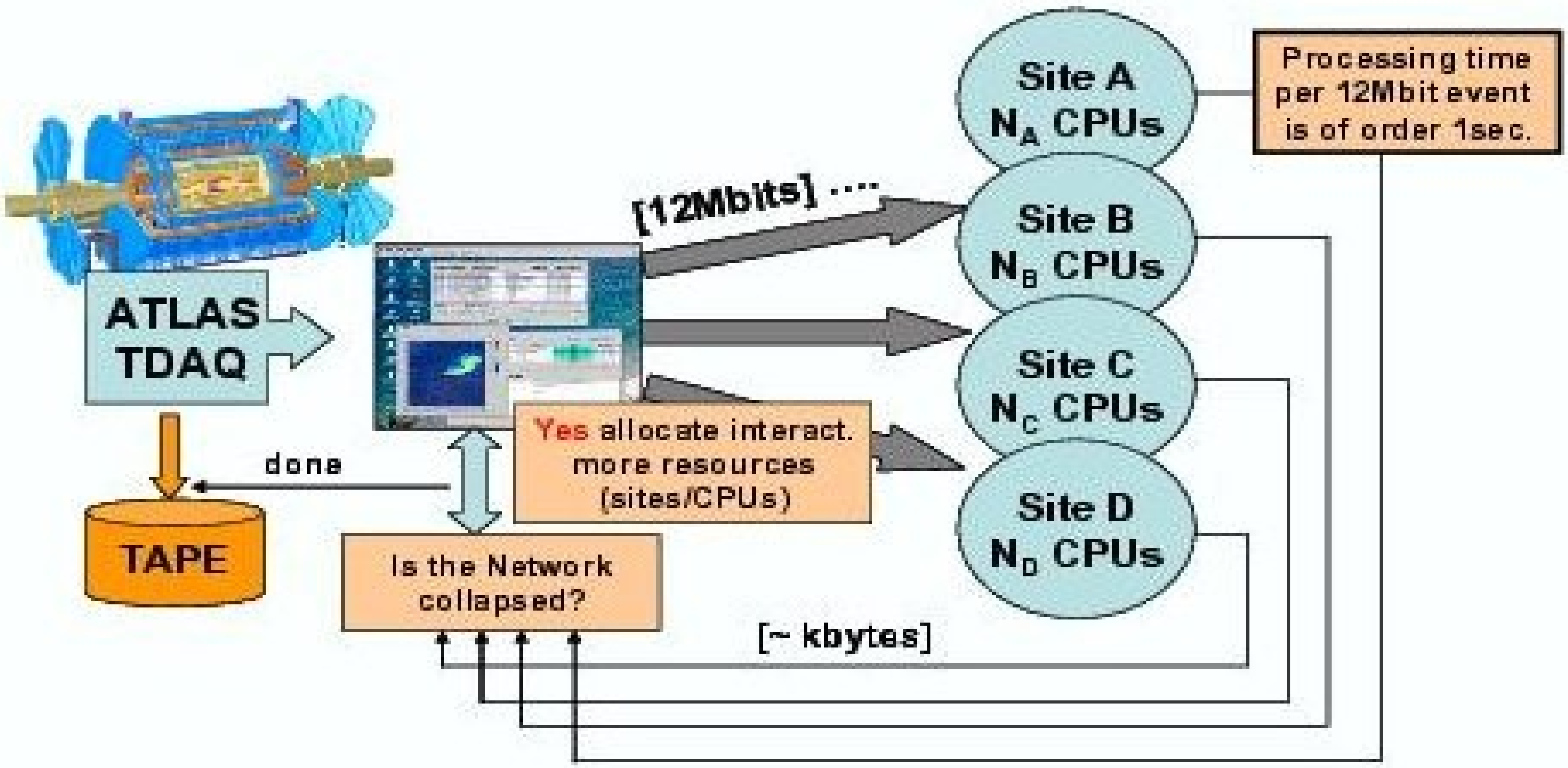
As used in off-line analysis of modern high-rate experiments in HEP:

- LHC experiments (CERN),
- BaBar (SLAC),
- COMPASS (CERN),
- CDF, D0 (FNAL),
- ZEUS (DESY),
- ...

Not much interactivity, but things are changing:

Example from Int.eu.grid (K. Korcyl, INP)

ATLAS online monitoring and calibration system



Wojciech Wiślicki, A. Sołtan Institute and ICM Univ. Warsaw

Similar developments in other fields

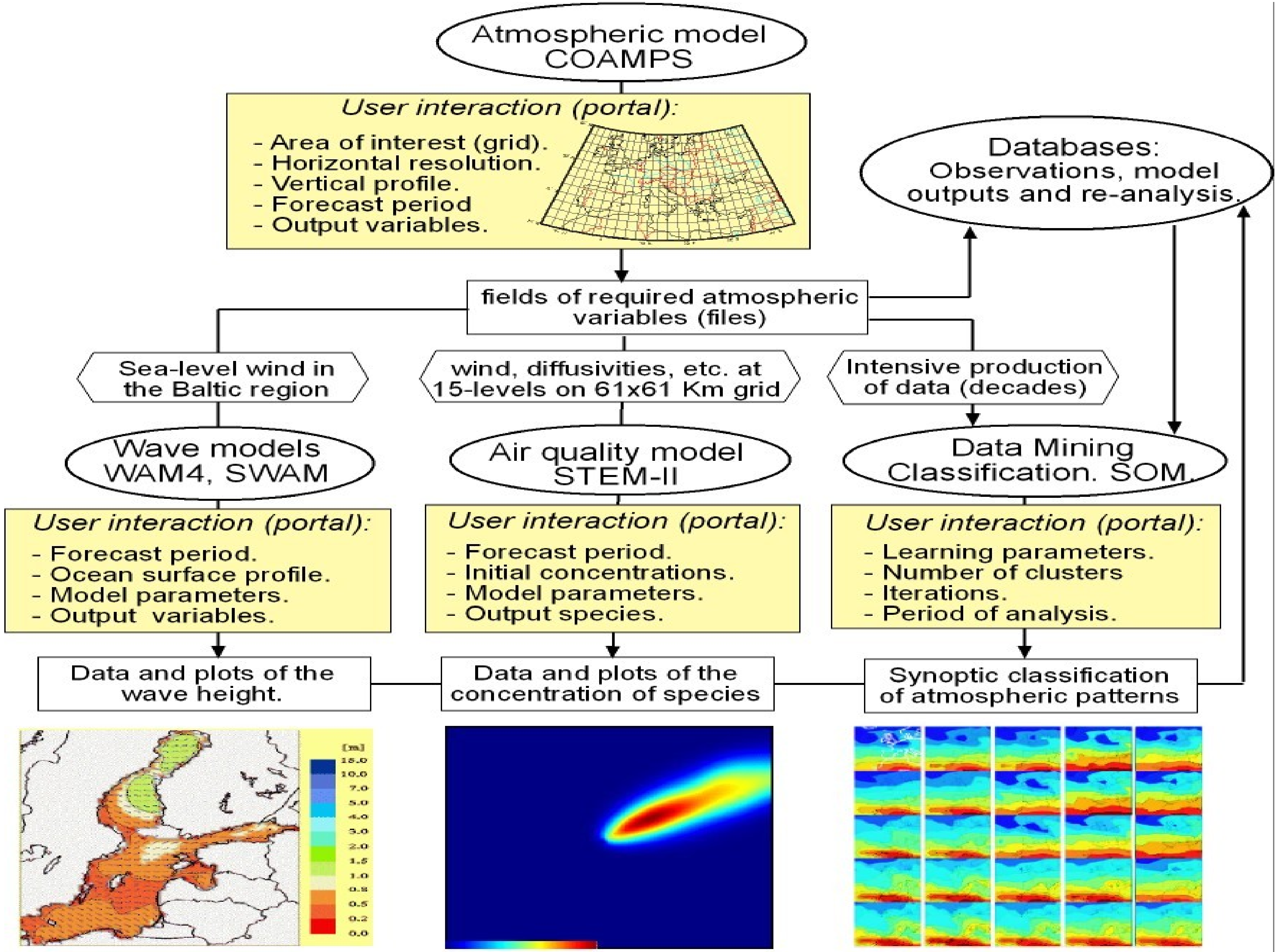
Using similar scheme and perhaps OpenMPI for parallelization:

- Vizualization of plasma in fusion reactors
- Ultrasound computer tomography for clinical purposes (20-30 mins response time with a precision 3D image)
- Environmental applications

Example of progress in environmental applications: meteo

Weather forecast coupled to models of pollution propagation and ocean waving (example CrossGrid: J.Gajewski)

- No strong advantage of its Grid implementation over a parallel machine's (e.g. CRAY) for high communication overhead for fine space-time granularity
- Need for instrument's feeback
- Need for dynamic workflow



Wojciech Wiślicki, A. Sołtan Institute and ICM Univ. Warsaw

The LEAD project

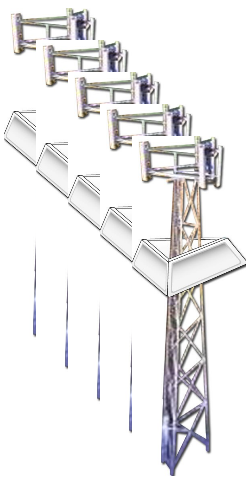
LEAD=Linked Environment for Atmospheric Discovery

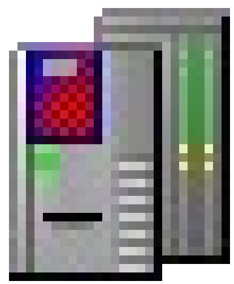
Strongly motivated by a need to foresee tropical tornados and hurricanes

Instrumentation: adaptive Doppler radars covering USA

(Presented by D. Gannon, Indiana U., last EGEE'06 Conference)

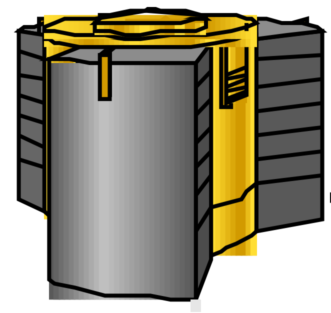
The LEAD Vision:
Adaptive Cyberinfrastructure





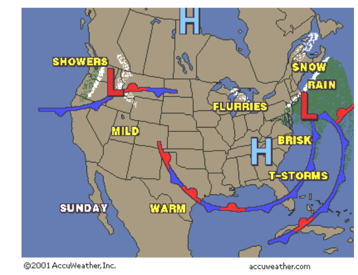
Analysis/Assimilation

Quality Control
Retrieval of Unobserved Quantities
Creation of Gridded Fields

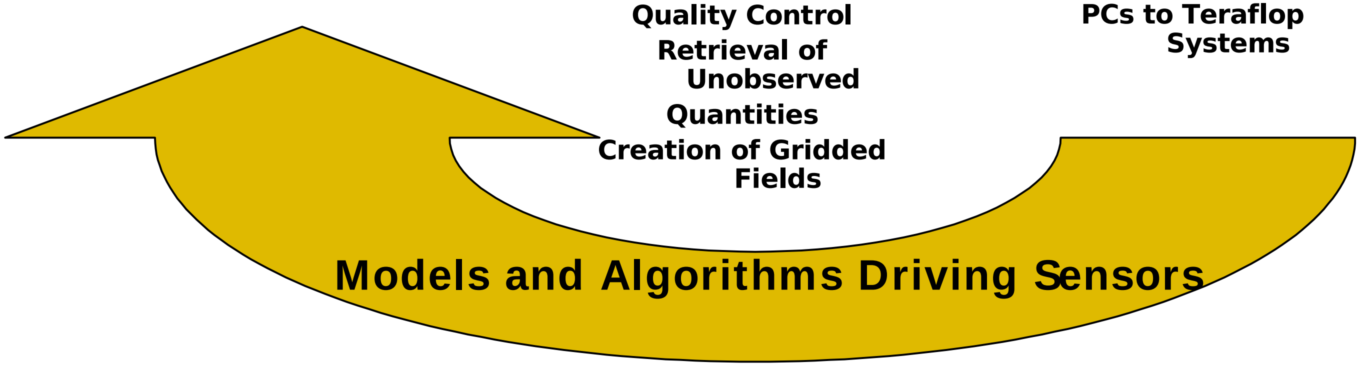


Prediction/Detection

PCs to Teraflop Systems



Product Generation, Display, Dissemination



End Users

NWS
Private Companies
Students

The CS challenge: Build cyberinfrastructure services that provide adaptability, scalability, availability, useability, and real-time response.

LEAD: example of paradigm shift

- Execute complex scenarios in response to weather events
  - Stream processing, triggers
  - Close loop with the instruments.
- Acquire computational resources on demand
  - Need supercomputer-scale resources
  - Invoked in response to weather events
- Deal with data deluge
  - User can no longer manage his/her own experiment products

Workflow as an experiment

- Workflow is a process of moving data through analysis steps to achieve a goal
- Most of analysis tools can be regarded as services
- Hence workflow is a sequence of services
- It is also a way of composing services in order to build new services
- Workflows can be composed and reshaped interactively using GUI

Avian flu example

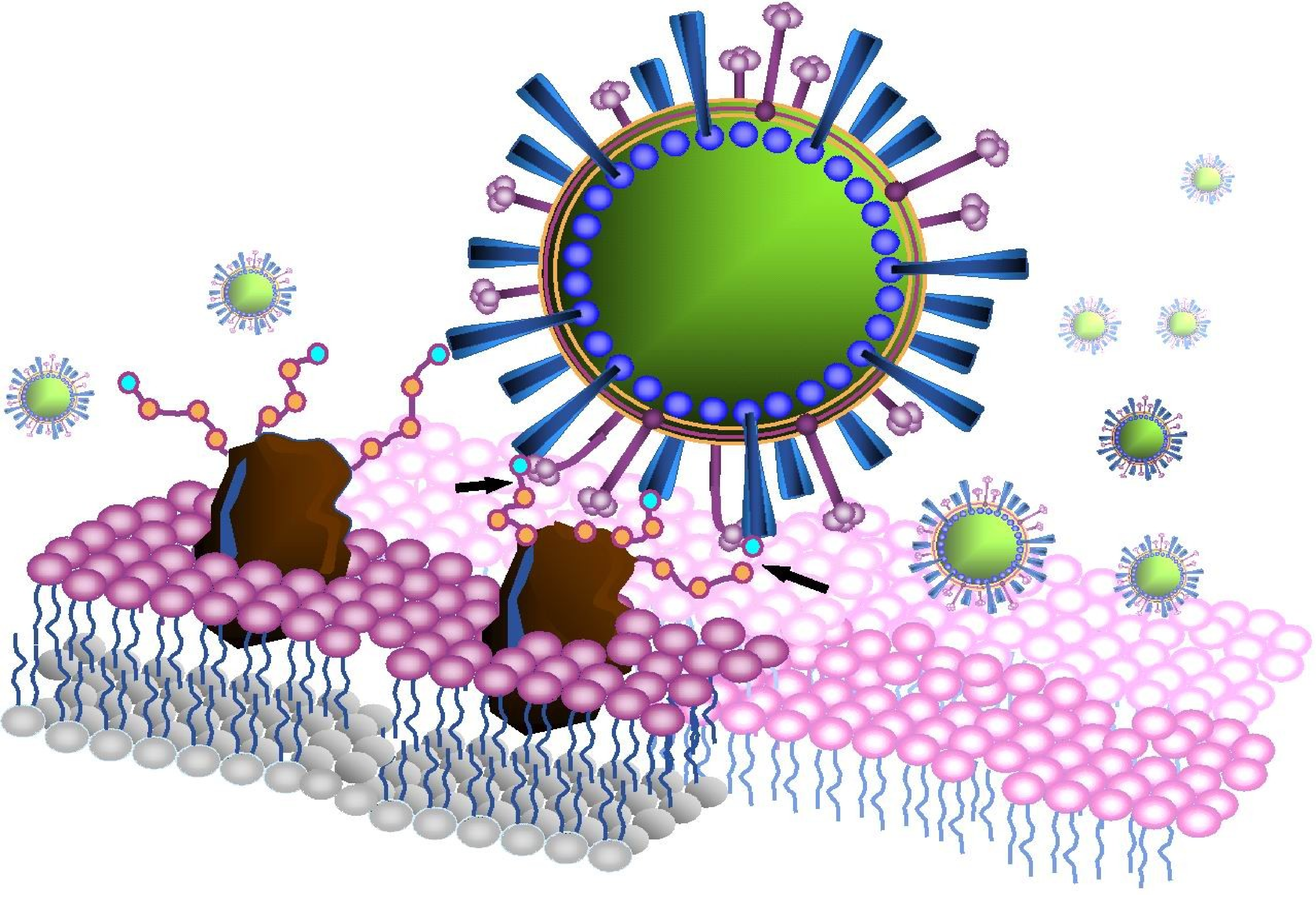
In Silico Docking on Grid infrastructures to accelerate drug design against H5N1 neuraminidases

The goal:

Block neuraminidasis activity on surface of the avian flu virus

Avian flu example, cont.

- Neuraminidasis increases avian flu virus reproduction
- Tries to be insensitive to drugs by using rapid structural mutations
- The task: find potential drugs to be docked on the virus surface and able to block neuraminidasis



Wojciech Wiślicki, A. Sołtan Institute and ICM Univ. Warsaw

Avian flu example, cont.

- The problem is computatinally hard:

  300 000 potential drugs, 8 structural mutations of neuraminidasis !

- 100 CPU-years, done in 1 month using 2000 CPUs on EGEE Grid

- Collaboration of: EGEE, Academia Sinica, Taiwan Grid, French Institutes (CNRS/IN2P3, CNR), BioInfoGrid, AUVERGRID, EMBRACE

Commercial perspective

- Pure research does not badly need robustness and security for applications, unless facilities are built
- Operations are gladly relegated to other (commercial) entities
- Business itself is only interested in research for publicity, in view of stable (long-term) income

Clearly visible commercial applications where non-trivial research is called for

- Autonomous tools for Grid monitoring, simulation and optimization for telecoms, e.g Grid simulator based on thermodynamic utility concepts (W. Wislicki)
- Interactive drug design from *first principles* (based on quantum molecular dynamics); in the framework of Bioexploratorium at ICM, but on local clusters so far (B. Lesyng)

Implications: unified way to the goal (result or product)

Novum for both research and industry:

- Applications as network services (SOA) &
- Including measurements and searchable databases into processes &
- Interactivity: feedback mechanisms built-in (either human or autonomic) &
- On-demand resource allocation &
- Visual workflows design