



Enabling Grids for E-science

Introduction to Grid Application Development

Christos Filippidis

(filippidis@inp.demokritos.gr)

Application Support Team

NCSR “Demokritos”, Institute of Nuclear Physics

www.eu-egee.org



Information Society



- Enabling a whole-system approach
- Collaborative research / engineering / public service ...

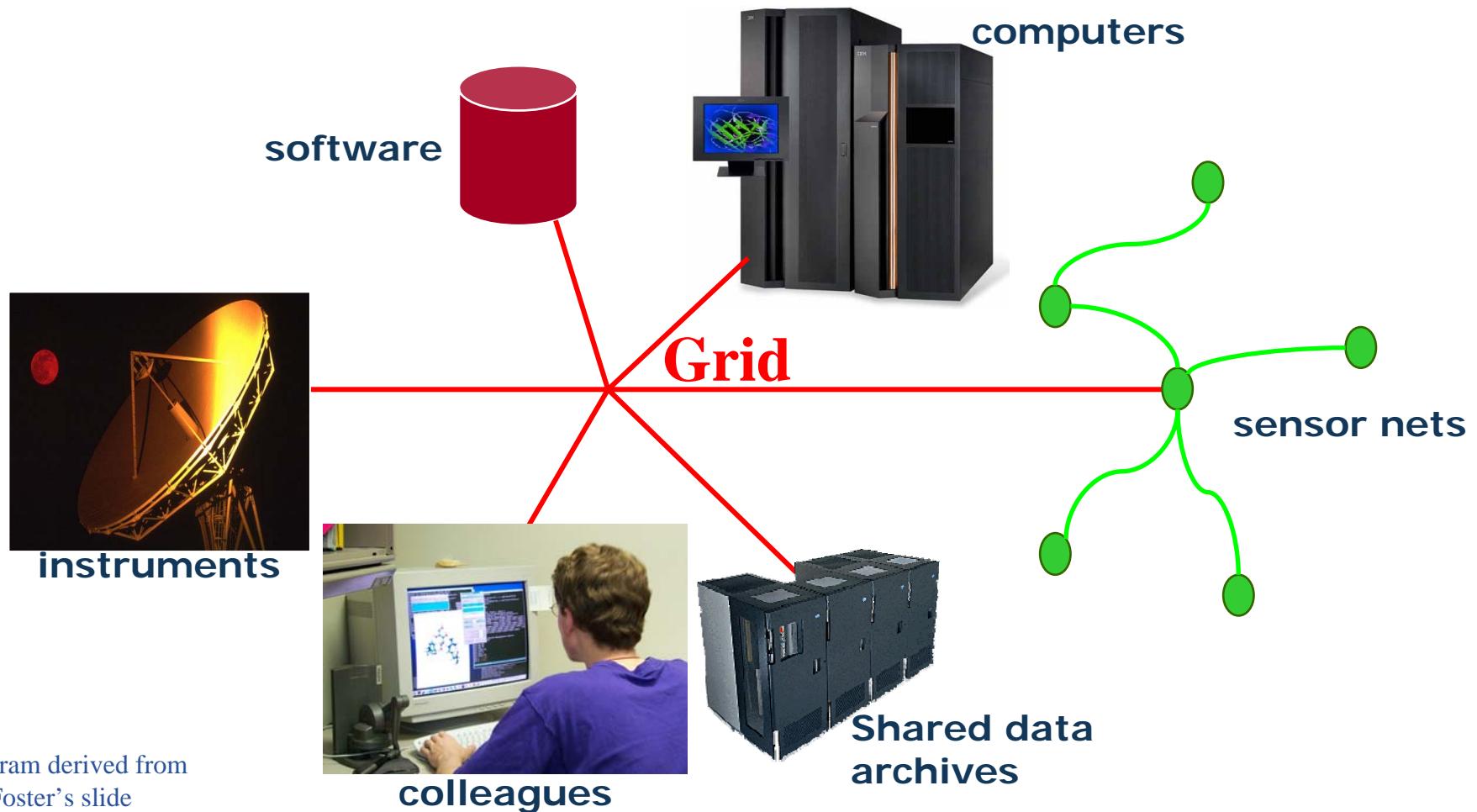


Diagram derived from
Ian Foster's slide

- **I need resources for my research**
 - I need richer functionality
 - MPI, parametric sweeps,...
 - Data and compute services together...

- **I provide an application for (y)our research**
 - How!?
 - Pre-install executables ?
 - Hosting environment?
 - Share data
 - Use it via portal?

- **We provide applications for (y)our research**
 - Also need:
 - Coordination of development
 - Standards
 - ...



Engineering challenges increasing

- **Team work!**
- **Engaged in world-wide initiatives – reuse, don't make your own! Cross disciplines for solutions.**
- **From research to production software: ~5 times the effort.**
 - “80% of the time for last 10% of the functionality & reliability”
- **Standardisation is key**
 - For re-use, for dynamic configuration of services,..
 - Both for middleware and domain specific (e.g. GEON)
- **Need to follow a deliberate development process**
 - Waterfall? Rapid prototyping?
 - Requirements engineering, design, implementation, validation, deployment
 - Engaged with the user community

- Application definition
- Environment configuration
- Information discovery
- Data management
- Program definition
- Program execution

- 1. Simple jobs – submitted to WMS to run in batch mode**
- 2. Job invokes grid services**
 - To read & write files on SE
 - Monitoring
 - For outbound connectivity (interactive jobs)
 - To manage metadata
 - ...
- 3. Complex jobs**
 - An environment controls multiple jobs on users' behalf
 - ⑩ High-level services
 - ⑩ Portals with workflow
 - ⑩ Software written for the VO (or by the user)
 - ⑩ ...

- **Direct Acyclic Graph (DAG)** is a set of jobs where the input, output, or execution of one or more jobs depends on one or more other jobs
- **A Collection** is a group of jobs with no dependencies
 - basically a collection of JDL's
- **A Parametric job** is a job having one or more attributes in the JDL that vary their values according to parameters
- **Using compound jobs** it is possible to have one shot submission of a (possibly very large, up to thousands) group of jobs
 - Submission time reduction
 - Single call to WMPProxy server
 - Single Authentication and Authorization process
 - Sharing of files between jobs
 - Availability of both a single Job Id to manage the group as a whole and an Id for each single job in the group

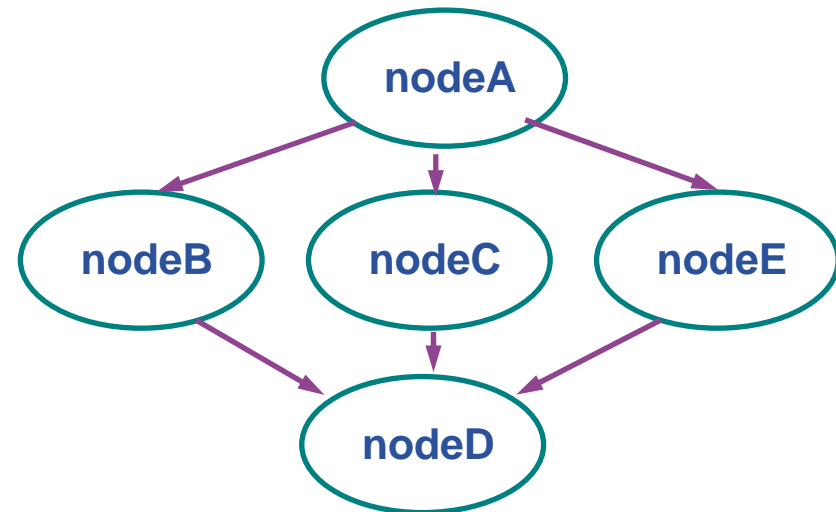
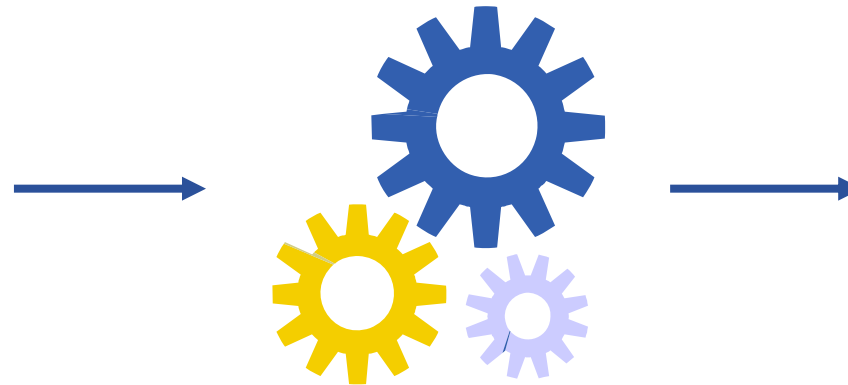


Image Compression



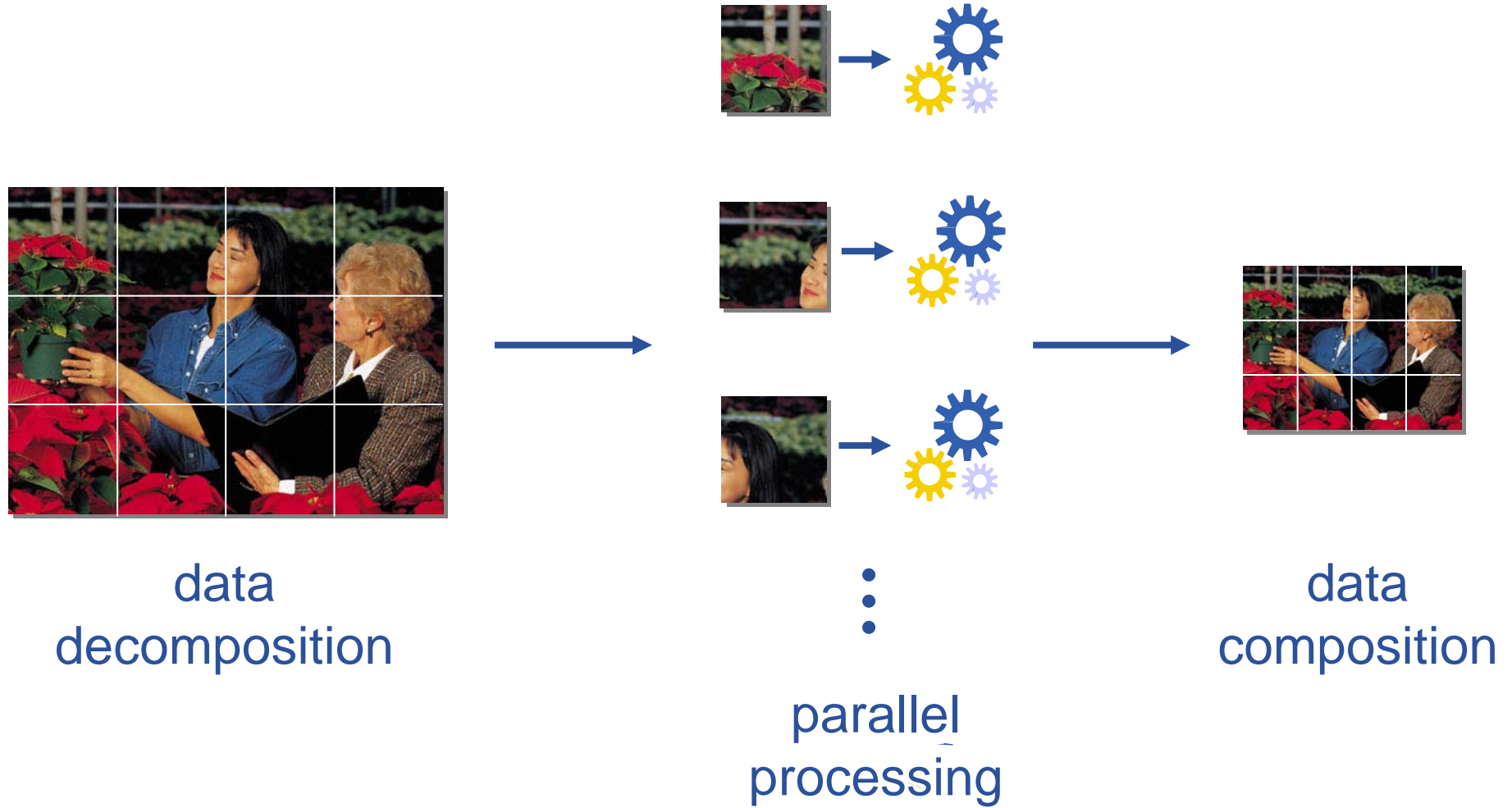
original
image



compression
algorithm



compressed
image



- **Security (VOMS)**
- **Metadata (AMGA)**
- **Monitoring (R-GMA)**
- ...



Enabling Grids for E-science

Questions?

Christos Filippidis

(filippidis@inp.demokritos.gr)

Application Support Team

NCSR “Demokritos”, Institute of Nuclear Physics

www.eu-egee.org

