GaiaGrid

S.G. Ansari, U. Lammers, M. ter Linden

ADASS XIV

Pasadena October 25th 2004

The Gaia Mission

Objectives

- Galaxy origin and formation;
- physics of stars and their evolution
- Galactic dynamics and distance scale
- Solar System census: minor planets, Kuiper Belt objects
- large-scale detection of all cla
  of astrophysical objects in
  brown dwarfs, white d
  planetary systems
- fundamental p

Launch Date: 2011

Capabilities

- catalogue: illion stars:
  0.34 x 0 mag
  26 = 15 mag
  V = 18 mag
  10^6 to V = 20 mag
  pleteness to about 20 mag
  sky density:
  mean density ~ 25 000 stars deg^{-2}
  max density ~ 3 x 10^6 stars deg^{-2}
- accuracies: median parallax errors:
  4 µas at 10 mag
  11 µas at 15 mag
  160 µas at 20 mag
- Radial velocity accuracies: 1 – 10 km s^{-1} to V=16-17 depending on Spectral Type
- Photometry: V=20 in 4 borad and 11 medium bands

Number Crunching

Total Performance

Core processing

[CCD calibration to fluxes, GIS, IDT] : **40 x 10^{18}** FLOPS

Shell processing

[variability, double star, RVS, etc.]: **90 x 10^{18}** FLOPS

Science Telemetry: **120 Terabytes** (over 5 years)

Estimated Archive: **1 Petabyte**

What will I cover?

Requirements

Architecture

The Community

Background

Gaia Data Processing has two basic components:

Core Tasks:

- Initial Data
- Glob olution
- C elations

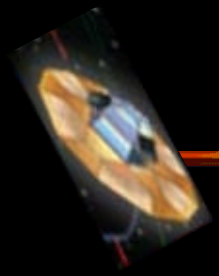
Centralised

Acts upon the totality of the data

Shell Tasks:

- Data Analysis
- Classification
- Photometric properties
- Spectroscopic Properties

Core Processing



Telemetry

Centroiding
Object Matching

Ingestion

Gaia Database

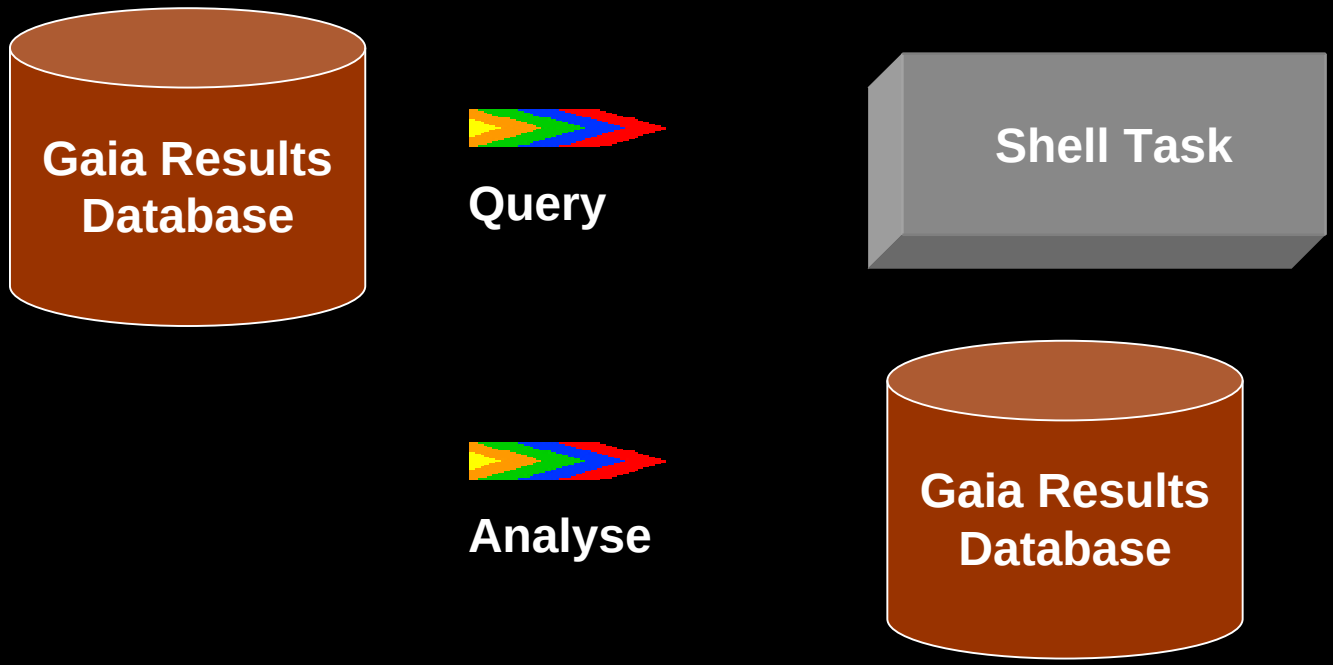
Process

Global Iterative Solution
Attitude
Calibration
Source
Global

Ingestion

Gaia Results Database

Shell Tasks


Gaia Results Database
Query
Shell Task
Analyse
Gaia Results Database

Shell Task Facts

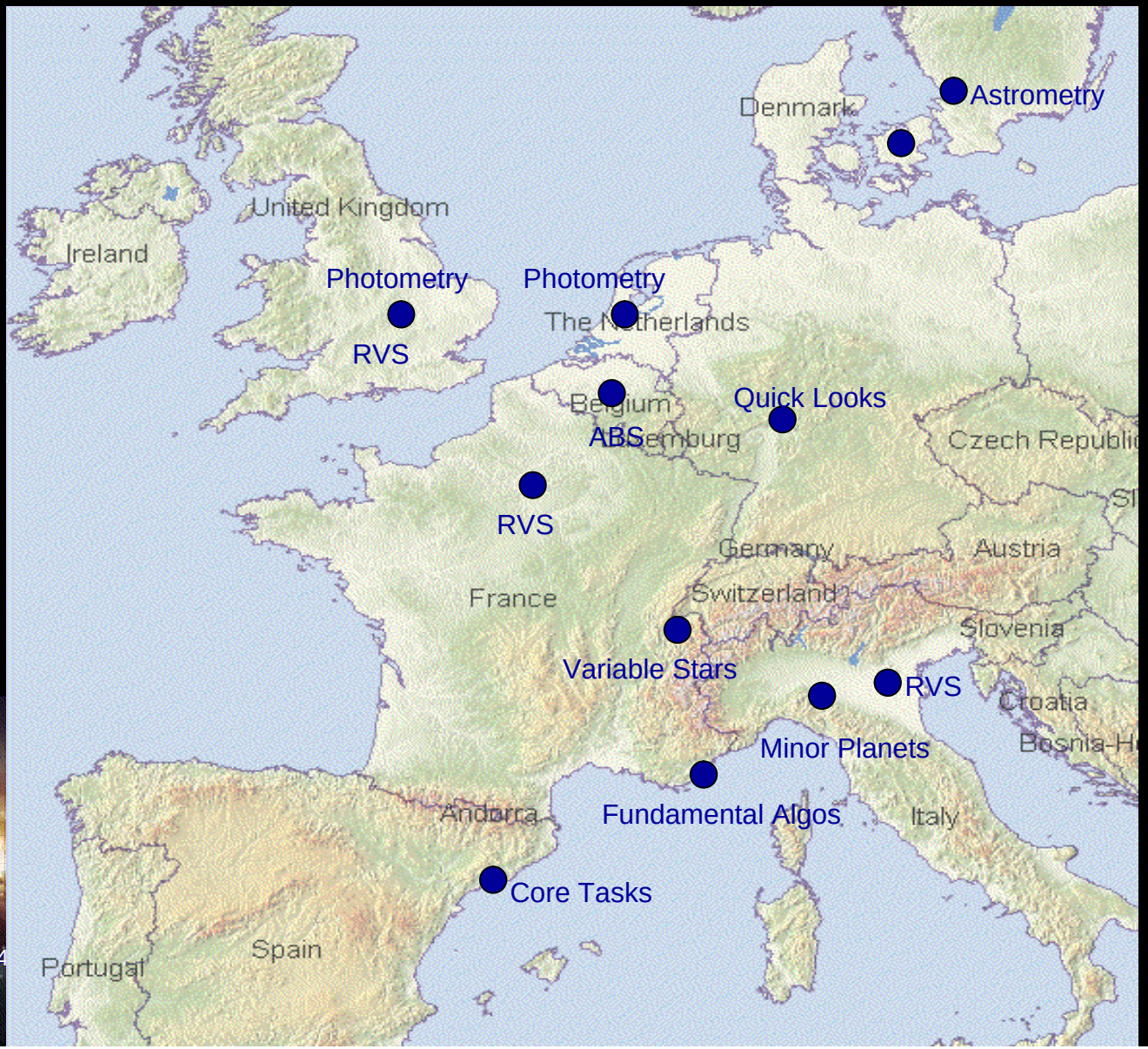
Scientific:

- Shell Tasks involve the Gaia community
- Shell Tasks may be developed by autonomous groups, independent of a core team
- Shell Tasks deliver “derived” data
- Shell Tasks can be collaborative tools
- Shell Tasks are building blocks for data analysis. They may be combined to address more complex processing tasks

Technical:

- Shell Tasks can be modular
- Shell Tasks access the Gaia Database to work on a subset of data
- Shell Task results can be independently validated. Less interaction with the core data.
- Shell Tasks could be developed in multiple set programming languages

Gaia Virtual Organisation



The Grid

The Grid is:

- A resource sharing concept
- Used to augment comp[illegible]al resources whenever and whe[illegible]eeded
- Ideal to build a [illegible]borative environment, where users ca[illegible] algorithms and analyse data

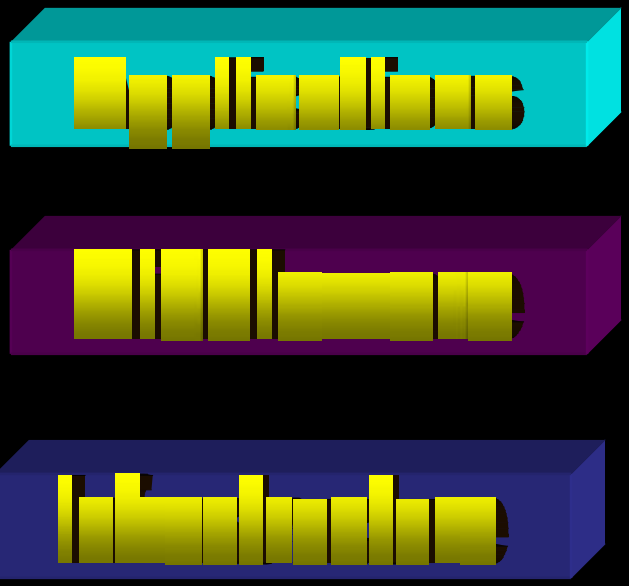
It is ideal for the Shell Tasks

The Grid Paradigm within the concept of Gaia

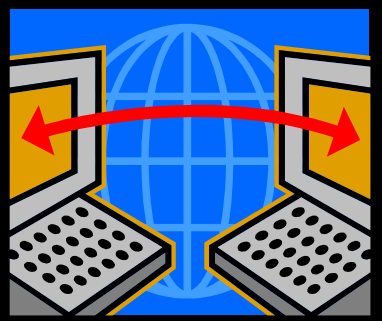
Three basic objectives

- Access to the data → Virtual Database
- Collaborative Computation → Virtual Organisation
- Computational Performance → Virtual Supercomputer

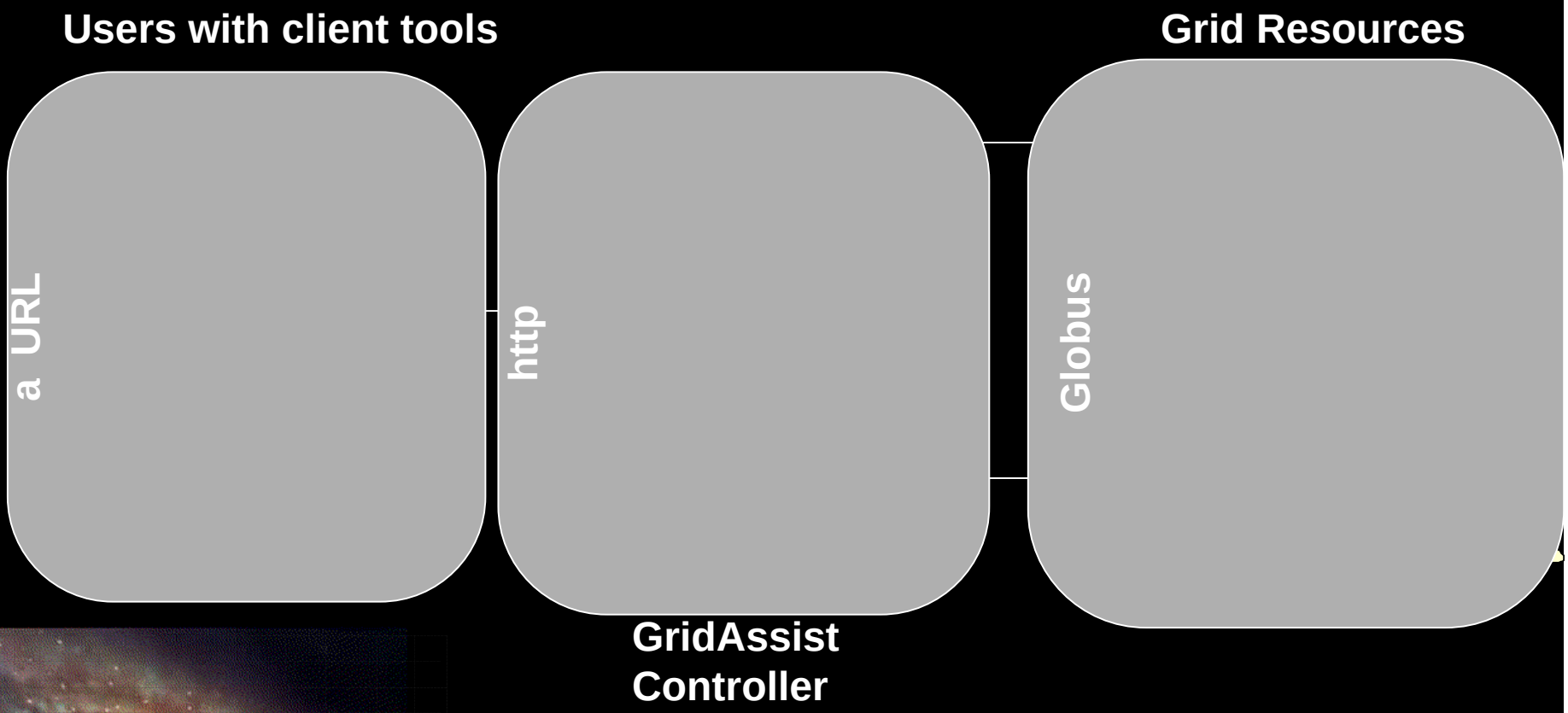
The Grid Architecture



Globus
http

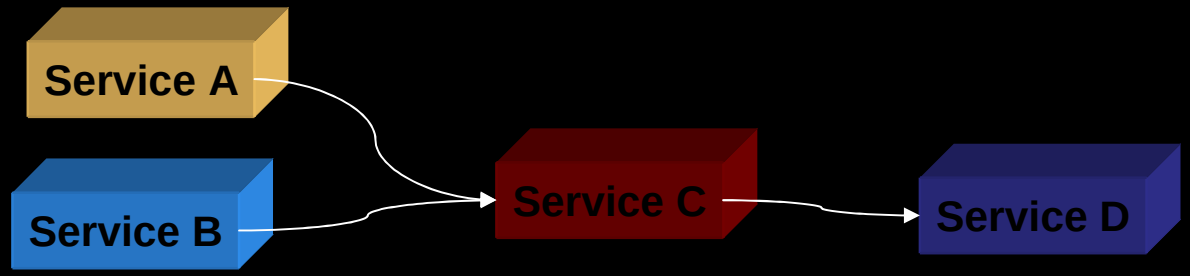


The GridAssist Infrastructure



Workflow

Workflow: Collection of services with dependencies



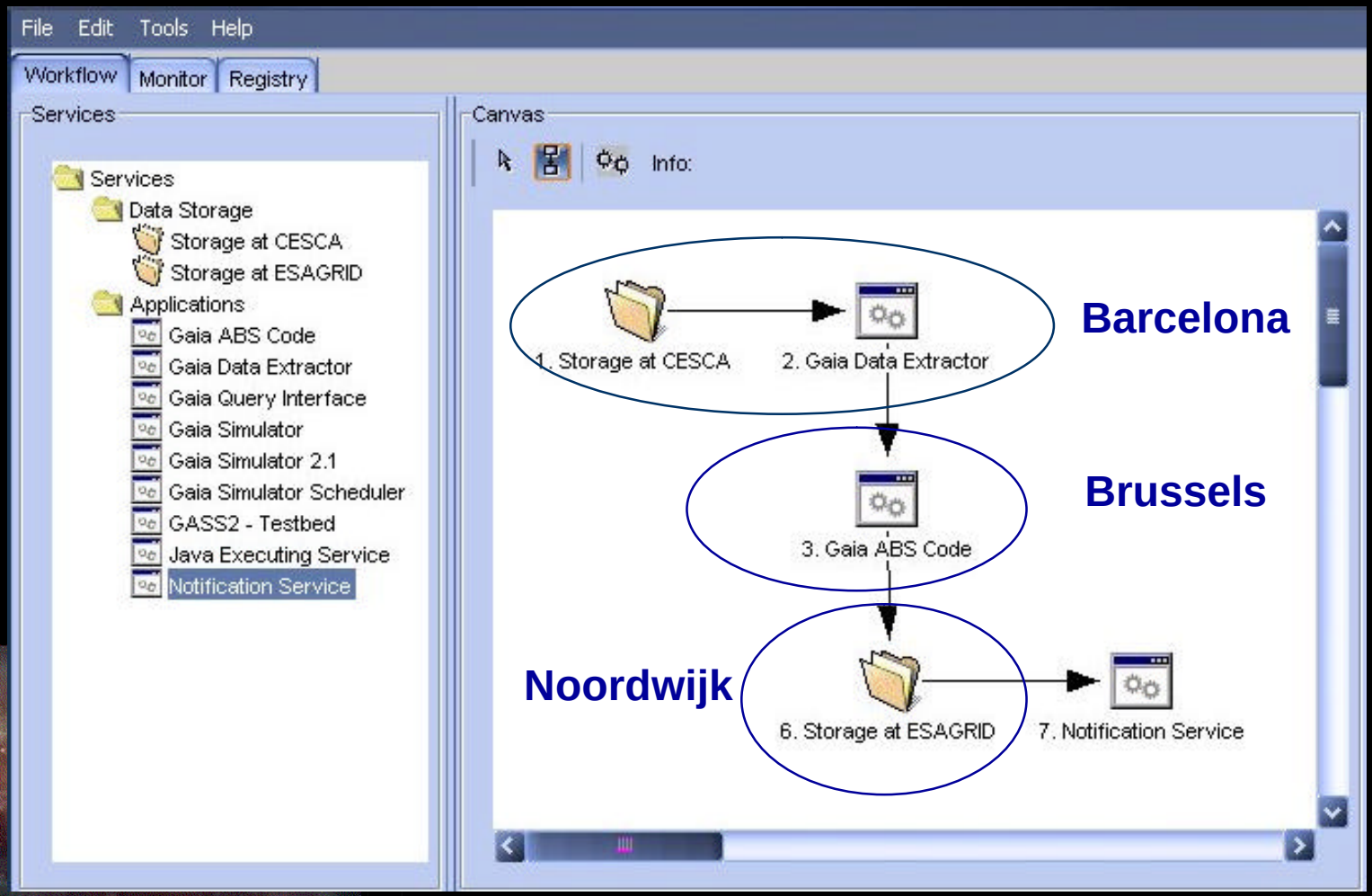
Examples of services:

- Retrieval of data from a database
- Execution of an Algorithm
- Storage of data

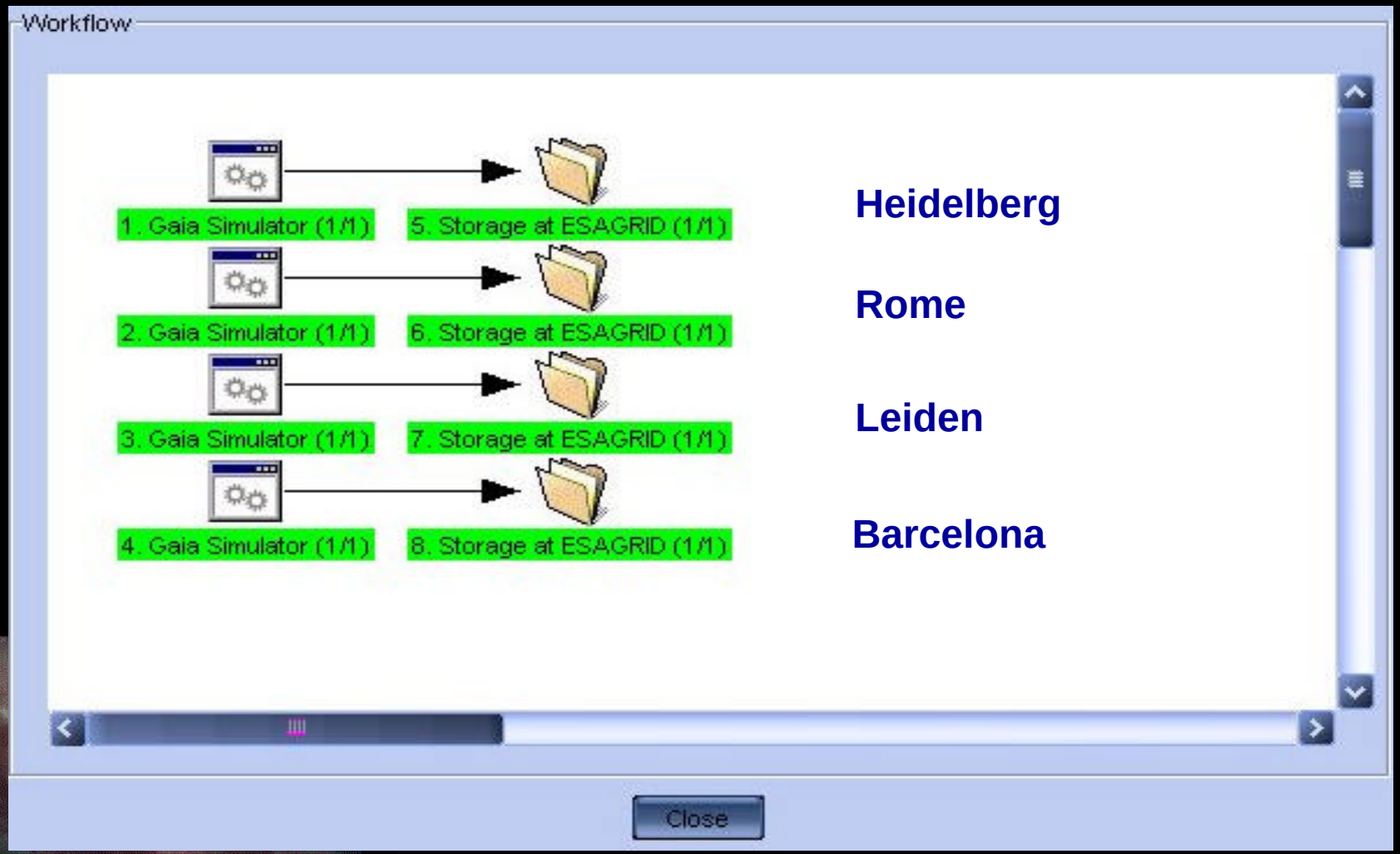
Examples of dependencies

- Data flows

The GridAssist Client
Distributed Grid Computation



The GridAssist Client
Performance Grid Computation



Heidelberg

Rome

Leiden

Barcelona

GridAssist Features

Function

- File Transfer
- Multiple Protocol
- Job Monitoring
- User Access Control
- Task Scheduling

Current Status

- Identified two potential Shell Algorithms for initial tests
- Connected a few sites: UB, ARI, ULB, ESTEC, ESRIN
- Ran medium-scale Gaia Simulations on the Grid.

Future Plans

- Stabilise the environment and deploy more Grid nodes
- Establish a testbed for all shell tasks
- Converge on an operational environment for Gaia