

The EU DataGrid Project

Three years of research and development in
Grid technologies



Erwin.Laure@cern.ch
DataGrid Technical Coordinator

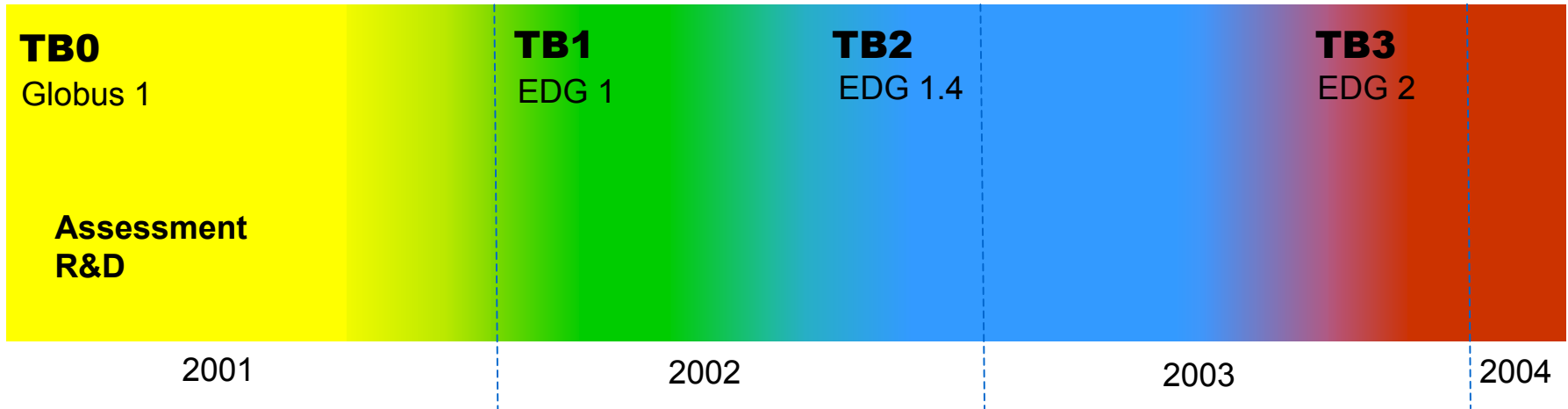
Outline

- ◆ DataGrid at a glance
- ◆ A chronological overview
- ◆ DataGrid assets
- ◆ Lessons learned
- ◆ Summary

Chronological overview

Jan 2001

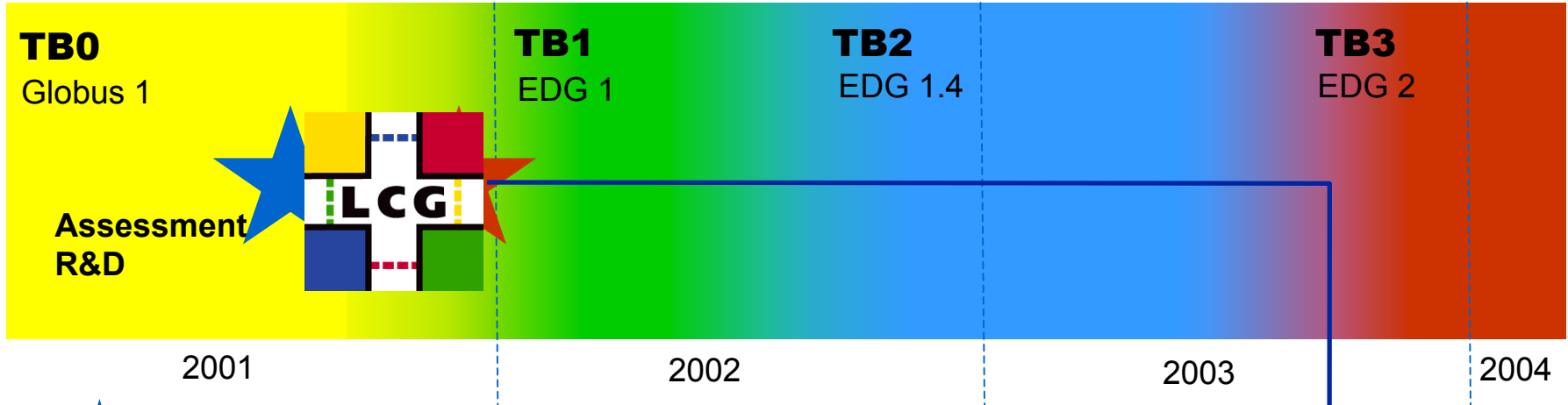
Mar 2004



- ◆ Project started on Jan 1st 2001
- ◆ Early distributed testbed based on Globus 1
- ◆ CA infrastructure established
- ◆ Development of higher level Grid middleware started
 - Workload management (“Broker”)
 - Data management (GDMP, edg-replica-manager, SE)
 - Information Services (R-GMA)
 - Fabric management (adopt LCFG)

Chronological overview

Jan 2001



★ Decided to base development on GT 2

- Delayed rollout of TB1 (EDG v1.0)

★ TB1 deployed on 5 sites

- CERN, NIKHEF, RAL, IN2P3, CNAF

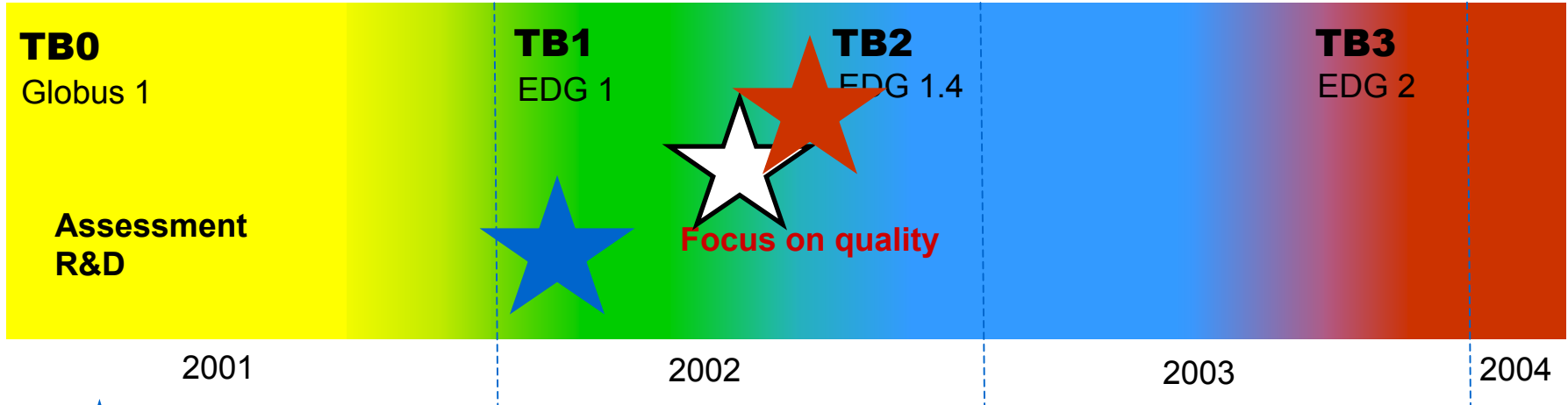
◆ Application evaluation started

- **1st HEP job run on TB1 on December 11th, 2001**

**CERN launched
LCG project in
September 2001**

Chronological overview

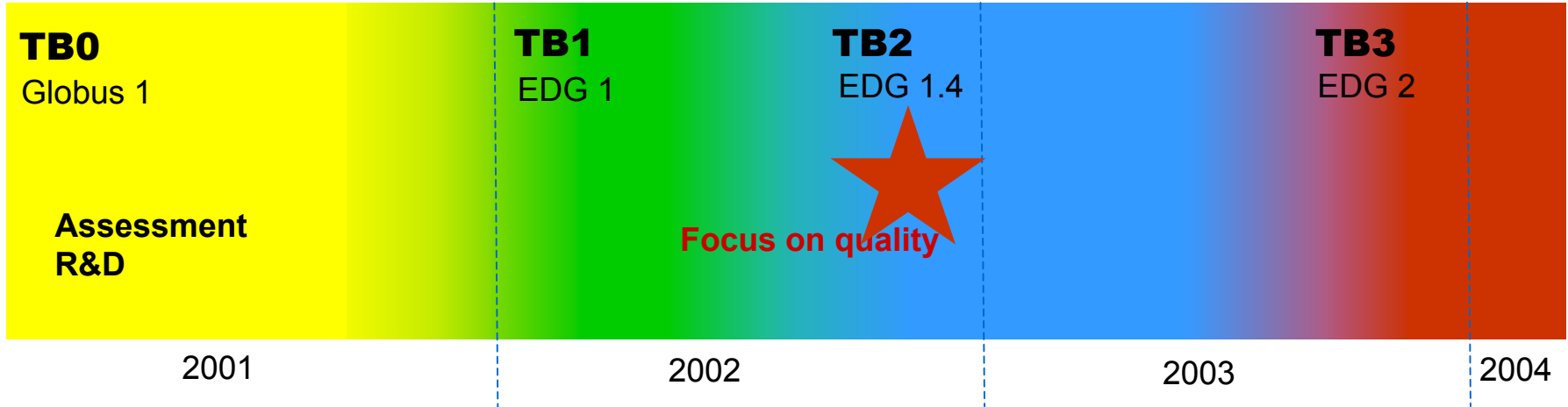
Jan 2001



- ★ 1st EU review successfully passed on March 1st 2002
- ★ Evaluation by end users revealed the need to **focus on stability** rather than new functionality
 - **Project retreat in August resulted in re-focus on quality**
- ★ **Open Source license** established in June 2002
 - Served as model for globus and CrossGrid license
- ★ Start of **tutorial program** in July 2002 (GGF5)
 - Developed into a road-show with hands-on sessions; more than 600 people trained in over 25 events

Chronological overview

Jan 2001



- ◆ EDG technologies widely recognized:
 - Many sites joined testbed (up to 20)
 - Software used and evaluated by other projects (e.g. CrossGrid, LCG)
 - Collaboration with sister projects demonstrated at **IST** and **SC**

★ Testbed 2 (End 2002, **release 1.4.x**)

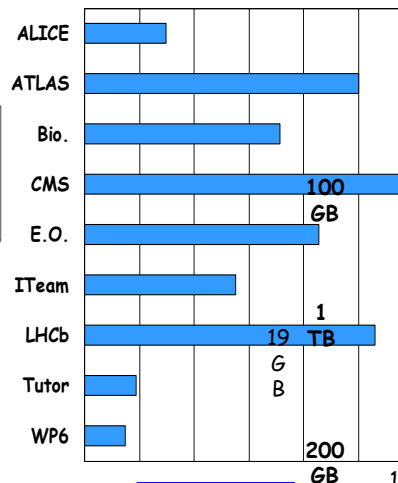
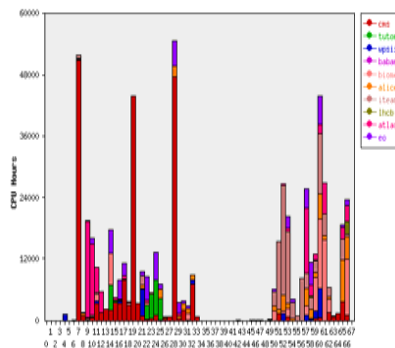
- One of the largest Grid testbeds worldwide
- Allowed first production tests by applications:
 - HEP monte-carlo simulation
 - EO grid portal developed
 - Many bio informatics applications

Evaluation of Release 1.4 (Dec 02/Jan 03)



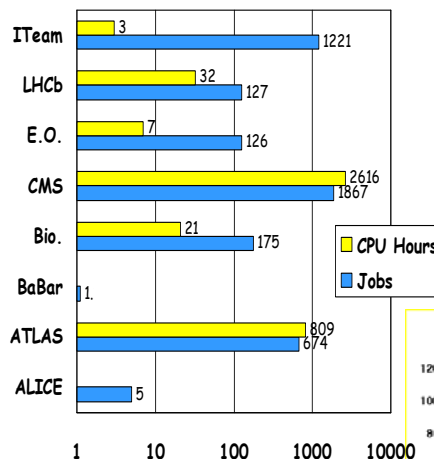
- ◆ Large increase in users
- ◆ Many sites interested in joining
- ◆ Pushing real jobs through system
- ◆ Stability and scalability not yet satisfactory
- ◆ Release 2.0 addresses the problems revealed

CPU Usage

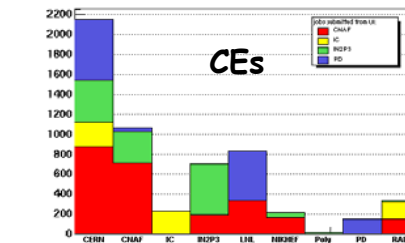
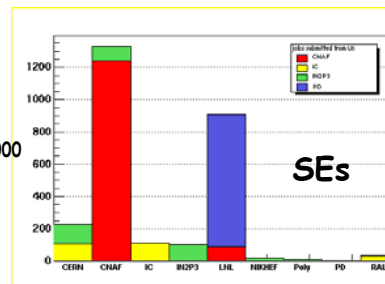
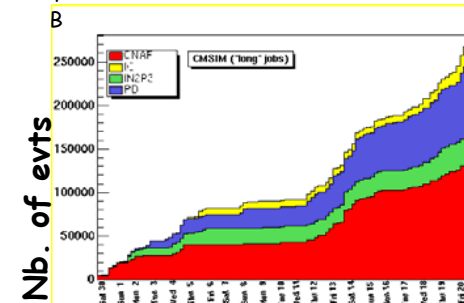


TOTAL: >1.5 TB

Disk Usage (CERN)



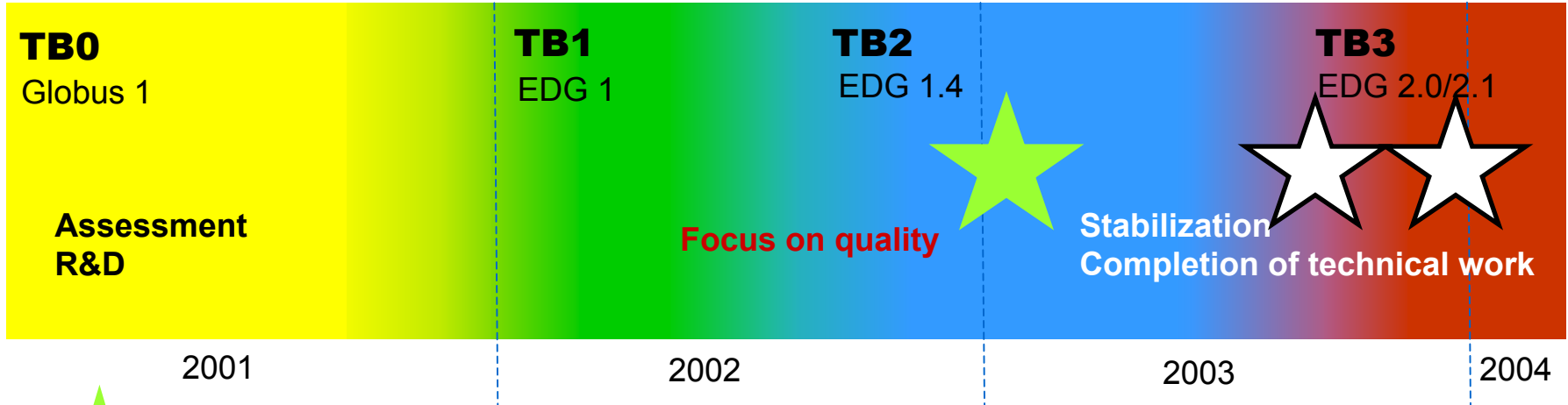
HEP Simulation



Chronological overview

Jan 2001

Mar 2004



★ Successfully passed 2nd annual EU review on February 4-5

◆ Shortcomings identified in application tests addressed:

- WMS re-factored
- RLS introduced
- Data management re-factored
- R-GMA introduced
- Storage Element (SE) introduced
- VOMS based security
- Fabric monitoring
- Upgrade underlying software (move to VDT managed releases of Globus and CondorG)

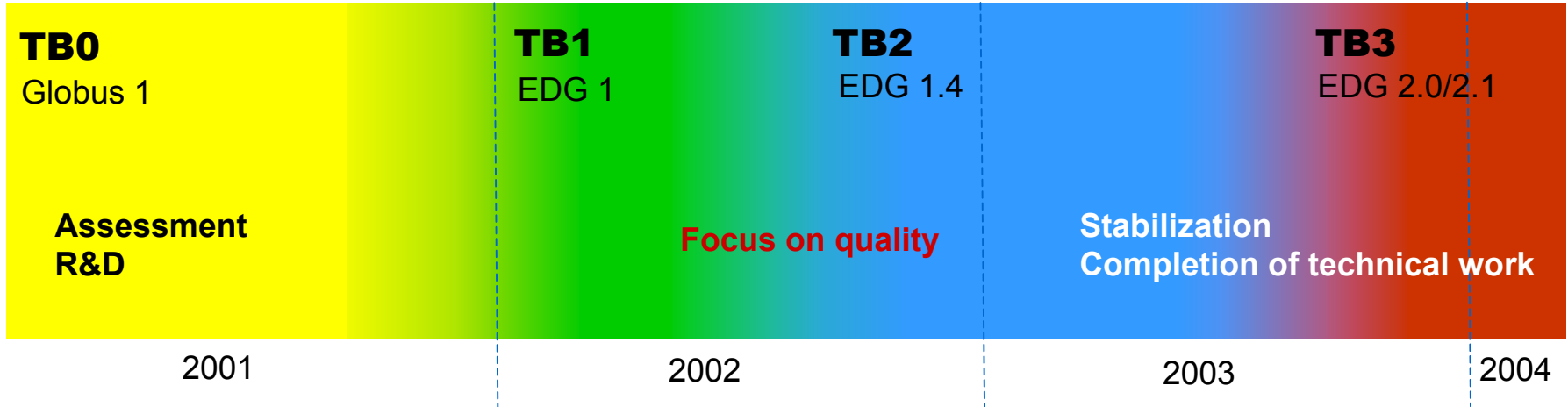
★ Testbed 3 (**release 2.x**)

- Advanced functionality, better scalability and reliability
- **2.0 released end of August**
- **2.1 released in November**

Chronological overview

Jan 2001

Mar 2004



- ◆ LCG deployed many components of EDG 2.0 in their LCG-1 service (started summer 2003) and subsequently EDG 2.1 components for LCG-2 (early 2004)
- ◆ Many other Grid projects started to use EDG software in 2003:
 - **Grace, grid.it, DutchGrid, UK e-Science programme, CERN's openlab, etc.**

DataGrid assets



- ◆ **Large scale testbed** continuously available throughout the project duration
 - Have gone further than any other project in providing a continuous, large-scale grid facility
- ◆ **CA Infrastructure** (21 CAs worldwide)
- ◆ **Innovative middleware**
 - Resource Broker
 - Replica Location Service and layered data management tools (Replica Manager & Optimizer)
 - R-GMA Information and Monitoring System
 - Automated configuration and installation tools
 - Access to diverse mass storage systems (StorageElement)
 - VOMS security model
- ◆ Software released under **open source license**
 - Facilitates wide-spread usage and encourages feedback that improves quality
- ◆ **Distributed team of people** across Europe that can work together effectively to produce concrete results
- ◆ **Application groups** are an integral part of the project contributing to all aspects of the work

Main lessons learned

- ◆ **Applications** need to be **involved** in all phases of the project
 - Grid middleware is relatively new and, despite all efforts, not yet “shrink-wrap” quality – requires skilled people to be used efficiently
 - Middleware prototypes need to be available for application testing early
 - **Caveat:** prototypes tend to stay longer than expected – more advanced software might be delayed.
- ◆ **Cross-WP activities** are essential and need to be coordinated
 - Application working group, architecture task force, integration team, security group, tutorial team, quality group.
- ◆ A sequence of (distributed) **testbeds** is needed
 - Developers need their own distributed testbed to test bleeding edge software
 - Managed integration/certification/application testbeds – eventually production infrastructure
- ◆ **Site certification and validation** needs to be automated and run regularly
 - Misconfigured sites may cause many failures
- ◆ **Security** needs to be an integrated part from the very beginning
 - Adding security to existing systems is hard
- ◆ Prompt hiring and continuity of **Personnel** is critical

Summary



◆ DataGrid as Grid Technology Innovator

- High level middleware developed in many areas (workload and data mgmt, information services, fabric mgmt)

◆ DataGrid as Technology Provider

- Software taken up by many other Grid projects (LCG, Grace, CrossGrid, grid.it, DutchGrid, UK e-science, openlab, ...)
- Extensive training in more than 25 tutorials held in US, Europe, AP
- Substantial contributions to standardization bodies like GGF

◆ DataGrid as Demonstrator

- Successful evaluation of Grid technologies as production platform by High Energy Physics, Earth Observation, and Bioinformatics applications. This paved the way towards

◆ Grid as next generation production infrastructure ⇒

