# 10 years of Data Management In the Grid World

Mario David

(david@lip.pt)

LIP Lisbon

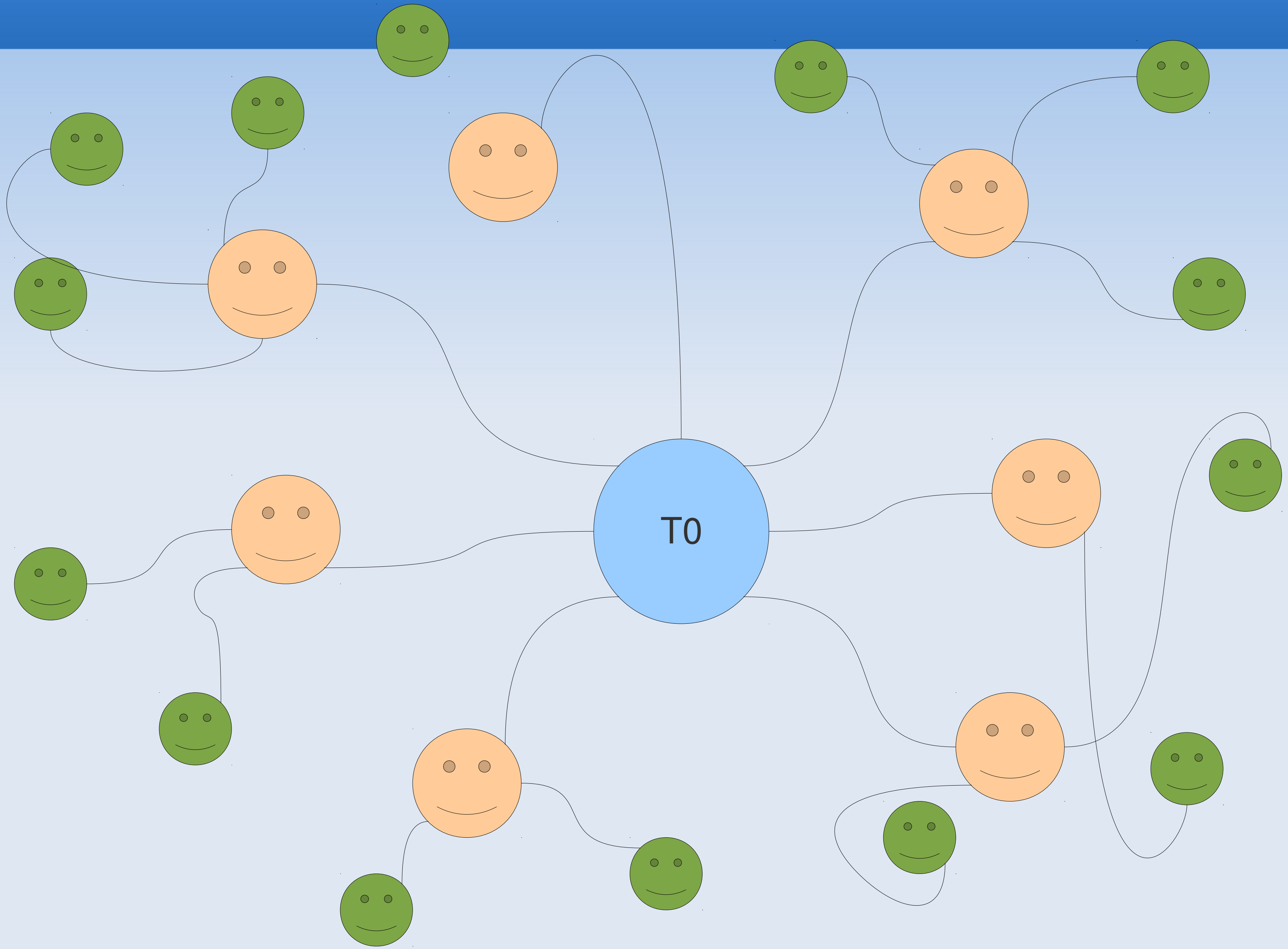
# A Disclaimer

- What follows is the opinion and views of the presenter. It does not, reflect the views of the organizations or projects for which the presenter is collaborating.

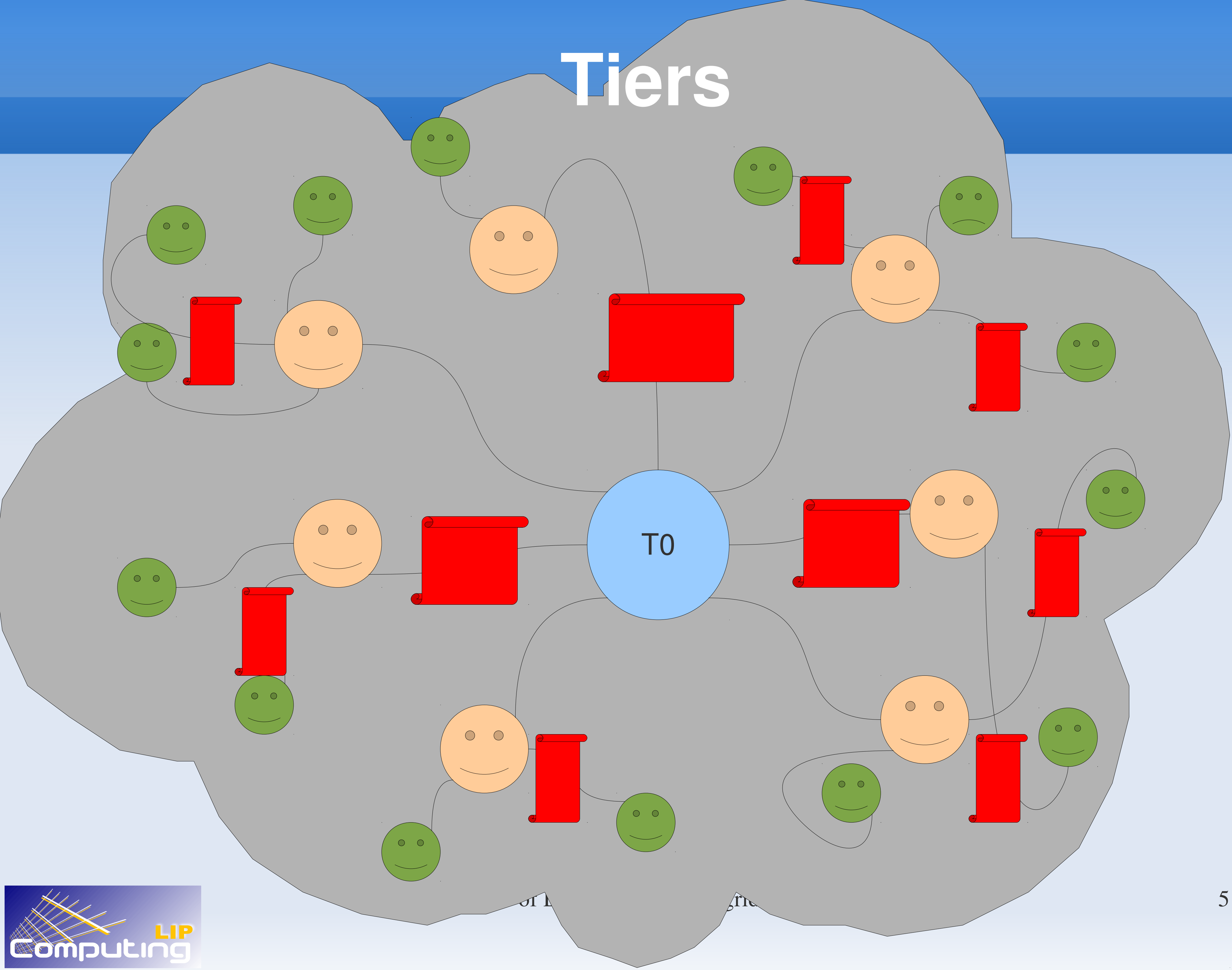
# Pre-history

- Once upon a time...
- There was ”**Monarch**” which lived in the late 90's.
  - Among other things, it was studying the **future**, I mean the technological trends, networking, storage, data management, … for the computing in the LHC era.
  - And... The Tier structure was born

# Tiers



# Tiers



# And by the way

Don't know if they remember about the existence of users (**a**lso **k**nown **a**s **scientists**) who would want to do data analysis

# The “Must Have” slide

- LHC experiments are expected to produce 10 Petabytes of experimental data annually
- $10^9$ collisions/second (1 GHz) – 100Hz (after filtering)
- 1 collision = 1MB of data
- 100 MB/s : **10 PB year = 20 Km CD stack**
- **OK, only a few Km if you use DVDs**

Balloon 30 Km
CD stack with 1 year LHC data ≈ 20 Km
Concorde 15 Km
Mt Blanc 4.8 Km

# The story begins:1999

- There were a couple of Americans that “said”:

  **The Grid**: Blueprint for a New Computing Infrastructure

- “Great, this is perfect, just what we need”: quoting LHC guys.
- Yep those americans are great, but we need something more complex, since the LHC is the most complex machine we ever built, so...

# The story begins:1999

- There were a couple of Americans that “said”:

  **The Grid**: Blueprint for a New Computing Infrastructure

- “Great, this is perfect, just what we need”: quoting LHC guys.
- Yep those americans are great, but we need something more complex, since the LHC is the most complex machine we ever built, so...

The birth of Globus

# EDG: 2001 - 2004

- It stands for “European **Data** Grid”
  - Notice “Data” in the middle of the name.
  - We will produce a lot of data, so we will have to move it around, take care of knowing where each piece is located, and have sufficient room for all of them, so …
  - It has to be shared among collaborators, and it has to be secure from external guys.

# What is Data Management (DM)

- Catalogs: where is it located?
- Storage: room for everybody!
- Movement... travelling a lot between different cities and hotels and rooms
  - At the time there was a thing called the “Replica Location Service” or RLS for friends.
  - There was the “Classical Storage Element” or Classical SE
  - There was the GridFTP protocol

# The “File Catalog”

- Whatever name this guy had in the past, or will have in the future, a “File Catalog” is:
  - There are a lot of sites, each with it's SE holding also lot of files, so...
  - I don't want to know the strange details of it's location and name.
  - I will gather in a **nice database** a “map” between a human understandable logical name and it's strange physical location and name.

# mini-HOWTO “File Catalogs”

- In the EDG times, the “File Catalog” was not so bad, at least in terms of design.
- There was some hierarchy, I mean each site had it's own catalog, meaning it could control and manage it, and would publish it's mappings at a higher level.
- But... at those times this was Research and Development.

# LCG: 2003 -... (until the end of time)

- OK, I tell you (if you don't already know): **L**HC **C**omputing **G**rid.
- Yeah, the EDG guys have done a lot of work, we take some of it, but... we need to re-design some of this stuff.
- And so, it was in this spirit that, among other things, the LFC and the SRM were born.

# A small digression through LFC

- It's the “LCG File Catalog”, as you can see, the name says it all.
- Each LHC experiment said, we will control everything, or at least as much as possible.
- Remember me saying a few slides back:
  - “I will gather in a **nice database** a “map”....”
- That is what was done with the design of the LFC, it's a “**single**” database, no hierarchy, no means of failover or replication, no site control or management.

# LFC: cont.

- The LFC is a “single” Unix like namespace.
- Holds the information of every file in every site of every user of a given “Virtual Organization”.
- Of course some **minor** questions can be asked:
  - What if a site (SE) is not available for some reason, well it's a possibility with hundred's of them around?
  - What if some files were lost at a site, for whatever reason?
  - And what if a site wants to migrate the files to another “machine”??

# LFC: cont. II

- The only way is for the site admin to send, through “old tech.”, an email to the administrators of that LFC saying:
  - Hey guys we lost some files here at our SE, it will take some time to see if we can recover them from backup.
  - Or better yet: We changed our SE, could you please update the LFC DB to reflect that for the files registered from our site?
- Does anybody has any idea of the coherency of the mappings in an LFC of a big experiment?

# LFC: cont. III

- And by the way: as of today, from the 4 LHC experiments only 2 are using the LFC in their production workflows, Atlas and LHCb.
- And even them had to do some complex work on top of it...

# Another “goodfellow”: the SRM

- It is a “Standard” recognized in the Open Grid Forum as: **Storage Resource Managers**.
- It was thought in the first place to be the standard Grid interface to access any Storage Element regardless of it's technology.
- There are 5 known players in the whole world:
  - Bestman   Castor   dCache   DPM   StoRM

# What do you need from it?

- You would think: I need all that I can normally do in a filesystem:
  - cp , ls, mkdir , rm(dir) chmod , acl's, stat , plus options
- In the first incarnation of it, only some of those were possible.
- In the last incarnation there are so many things you can do, that it became incredibly complex.

# Why did it get so complicated?

- Two simple words answer a large part of this question:
  - Space Tokens: it's a way to partition and limit the total space dedicated to a given purpose. (**I confess that I like this one**).
  - Nearline storage: just access Hierarchical Mass Storage systems “almost” as if it was a disc based FS.
- OK just remembered about: advanced reservation, lifetime of files and spaces.
- The list goes on...

# I will not rumble about the SE's

- I will not go on with discussing personal favourites of which storage element is best or worse.
- You have 5 to choose from, so...

# GridFTP

- It's still the preferred protocol to move files around, no much say in this area.
  - It's performant
  - It uses the security infrastructure
  - Is able to use close to 100% of the raw bandwidth
  - ...

# EGEE: 2004 – 06 – 08 – 10

- It just finished in the last day of last month.
- Enabling Grids for E-sciencE
- Quoting some of the political and dissemination fireflies:
    - “It's the largest **multidisciplinary** Grid infrastructure in the World”
    - As of today you might have around 200 VO's there, from Particle Physics to Finance and Economics
        - Why not simulate the stock market in the grid, or in the same line the sub-prime or the oil price.

# EGEE: cont.

- Though an infrastructure project, they invented the “gLite” middleware. OK, at least they invented the name, and truth be told a few new services.
- One of such new services is the “File Transfer Service” or FTS. It's a “thing” to schedule the movement of large amount of files as needed for example for the LHC data.

# My tale is almost at an END

- We now have the European Grid Initiative which you heard about yesterday, and the European Middleware Initiative (which the Kick of Meeting is occurring today).
- So I do not talk more about it.

# Predicting the future

- My tale does not have what it could be considered an end. That only the future can tell.
- But I do not resist in telling you a bit about the future, which in fact starts at the 16 of June in Amsterdam (I know that at the moment Amsterdam seems to be the center of the world but...).
- That day marks the start of the discussion about the future, LCG is longing for...

# Continue to predict the future

- Global filesystems
- Storage clouds
- **There are** users wanting and needing to do data analysis
- The SRM was after all a big monster.
- … and a few more things.

# To conclude

- I have to say that I continue to have fun with:
  - Learning and testing new pieces of Grid middleware
  - Giving feedback to developers, and saying to them this is good or this not so good, or suggesting you could do it this way or that.
  - There are pieces they could have designed them much better, I mean more reliable/robust/scalable/easier to use.
  - There are pieces which I find quite nice to have around.

# Questions/Comments are welcome

And the Higgs will
be found, or whatever
breaks symmetry
Or binds all interactions