Next Hop: Knowledge

e-Science needs in terms of data and computing e-infrastructures
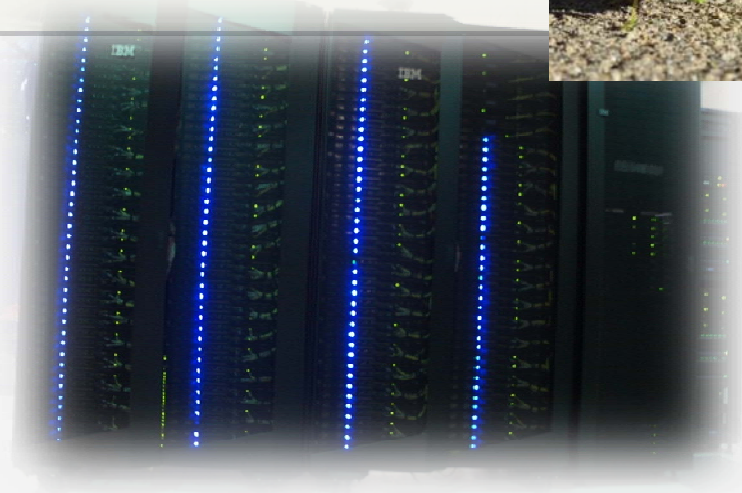
Jesús Marco de Lucas

National Research Council in SPAIN (CSIC)

Instituto de Física de Cantabria (IFCA)

marco@ifca.unican.es

Input from many sources, but all errors/mistakes are mine and only mine

My background: basic research

- **CSIC: National Research Council in Spain, > 120 Research Centers in**
  Social Sciences and Humanities, Biology, Agriculture, Natural Resources, **Physics**,
  Materials Science, Chemistry + National Labs

- **IFCA is a Research Center in Physics in Cantabria (North Spain)**
  Particle Physics (LHC), Astrophysics, Nonlinear systems, Meteorology +...

- **Research line on Advanced Computing and e-Science** *(since ~2000)*

  - **e-Infrastructure support to large projects**
    - EGI.eu / NATIONAL GRID NGI-ES & IBERGRID
    - TIER-2 CMS @ LHC
    - ALTAMIRA SUPERCOMPUTER
    - CLOUD ENABLED RESOURCES
      - FEDERATED CLOUD AT EUROPEAN LEVEL

  - **MULTI/(EVEN INTER)/DISCIPLINARY PROJECTS**
    - BIODIVERSITY (LIFEWATCH)
    - CYANOS @ WATER RESERVOIR
    - VENOME EVOLUTION
    - ECONO-SOCIO-PHYSICS
    - INTEGRAL TRACEBILITY

Jesús Marco de Lucas, e-Science evolution

My to-do list for e-Science

- 🔴 THINK AND PLAN comes FIRST
  - ■ Is it possible an e-Science forum? i.e., can we connect our brains for new ideas?
  - ■ How to setup priorities? How to take reputation on board?
- 🔴 TECHNICAL CHALLENGES SHOULD BE DRIVEN BY SCIENCE AND INNOVATION
  - ■ But, what could I do TODAY if I have access to UNLIMITED COMPUTING RESOURCES?
  - ■ Technical personnel training to bridge SCIENCE and TECHNICAL COMPONENTS
- 🔴 WE NEED A LONG-TERM (20 YEARS?) INTEGRATED FRAMEWORK:
  - ■ DISTRIBUTED, OPEN SOURCE Community with an Exploitation strategy
  - ■ EVOLVING e-Infrastructure & Middleware, LEAD BY SCIENCE IDEAS/REQUIREMENTS
  - ■ STABLE Analysis Toolboxes, with "transparent" access to resources/apps
  - ■ PRESERVATION included since the start
  - ■ FLEXIBLE TO SUPPORT DIFFERENT PARADIGMS, AREAS AND TECHNIQUES
- 🔴 BE PREPARED FOR AN UNCERTAIN FUTURE

  - ■ FOCUS ON KNOWLEDGE

  - ■ ENGAGE SOCIETY

1.- THINK & PLAN

- **ADVANTAGES OF AN e-SCIENCE FORUM** (EU level, International level)
  - Do we have a Map of Research and e-Science for EUROPE?
    - RESEARCHERS & TEAMS: EXPERTISE
    - INSTITUTIONS
    - RESOURCES (DATA SOURCES, NETWORK, COMPUTING)
    - AIMS, OBJECTIVES, INTERESTS
    - PROJECTS & RESULTS & NEW STRATEGIES
  - Do we have Tools to support an e-Science Forum (in a sustainable way!!!!)
    - MUCH MORE OPEN DISCUSSION, TRUSTED REPOSITORIES OF PRESENTATIONS, PAPERS
    - MUCH LESS TRAVELS
    - MUCH LESS REPORTS
    - MUCH LESS ABOUT MONEY/FUNDING
    - MUCH LESS FRAGMENTATION
  - **How can we integrate REPUTATION?**

VIVO -like Tool? By institution? By field?

TELECONF system

SCI-TECH oriented SLIDES Arxiv?

I don´t know! Learn from Social Networks?

2.- TECHNICAL CHALLENGES

● What could I do TODAY WITH UNLIMITED COMPUTING?

- Pattern Matching/Pattern recognition (i.e. Brain)
- Unlimited Simulation

Example (I know): can I improve the reconstruction of an LHC collision by simulating all possible final state configurations and re-matching?

Example (I talk): can I improve image recognition using a simulation of the context?
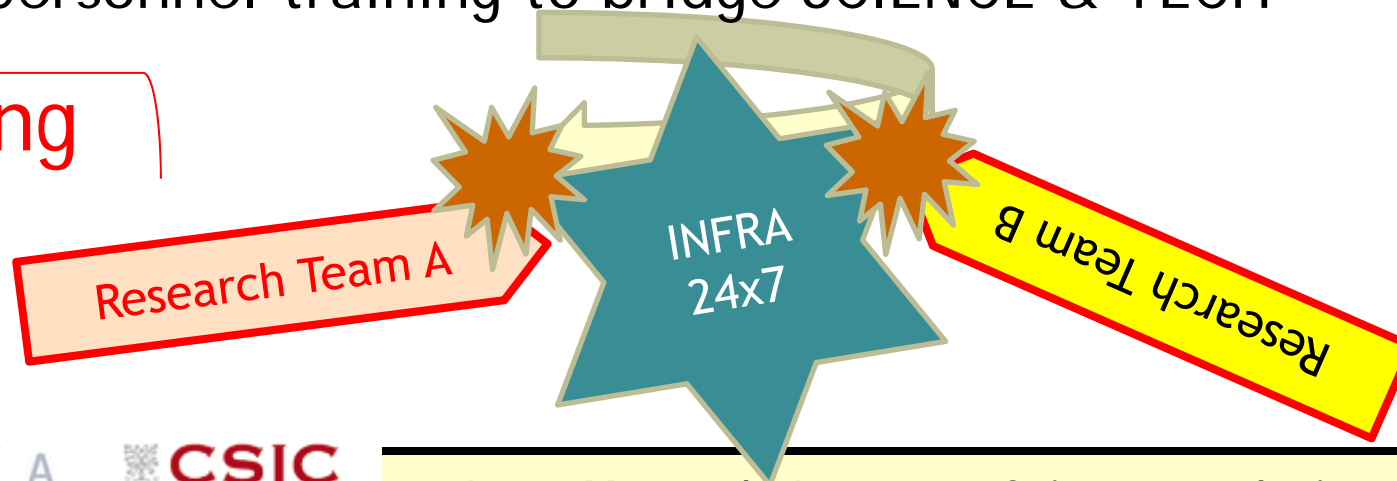
HOW TO USE UNLIMITED COMPUTING RESOURCES?

Programming (low le... "Many" Cores, GPU

● Technical personnel training to bridge SCIENCE & TECH

young

Research Team A

INFRA 24x7

Research Team B

3.- A LONG-TERM (20 YEARS?) FRAMEWORK

Who will analyze SENTINEL-1 data

- **DISTRIBUTED, OPEN SOURCE Community**

  with an Exploitation strategy

  Supported with Funds for Innovation (Consider new RIS-3 and Structural Funds)

  An stable PPP ecosystem with SME accessing Public e-INFRA???

- **EVOLVING e-Infrastructure & Middleware, LEAD BY SCIENCE IDEAS/REQUIREMENTS**

  - I hope EGI, PRACE, will be there... with stronger USER INPUT
    - Discuss at e-Science Forum global cases (and not at the e-Infra level)
    - Support both Long Tail (Capillarity) and Large Initiatives
    - Keep a healthy competitiveness, but do not aim to make "business"
  - I hope new "CLOUD" middleware will make easier the life of infrastructure and... also of final users

A Vision in Common

Davide Salomoni, INFN

- A **PaaS integration layer** able to adapt to the needs of European scientific communities and to customized user environments.

- **Technical focus** on:
  - Flexible and expandable user interfaces.
  - Integration with hybrid e-infrastructures.
  - Distributed service composition.
  - Key missing services in the area of distributed authorization, efficient (e.g. container-based) virtualization, resource allocation, standards-based storage connectors.

- **Enabling platform** for VREs and for the evolution of data/computing scientific frameworks, targeting **optimal usage of public or private resources.**

eScience Evolution: DataCloud,
a Proposal for the EINFRA-1 Call, Topics 4-5

- **DataCloud**, a proposal being prepared right now for submission to the EINFRA-1 call, addressing the overall vision shown above.
  - Large-scale virtualization of data/compute resources.
  - Development and adoption of a standards-based computing platform (with open software stack).
- Focusing on the extension of industry-standard open solutions to **deliver a PaaS platform enabling integration and interoperability across European e-Infrastructures**.
  - For HTC, HPC and Hybrid (public/private) Clouds.
- The proposed Consortium is **tightly linked to data/compute needs of diverse scientific projects and communities**.
  - E.g. ESFRI projects such as ELIXIR, LifeWatch, DARIAH, EMSO, INSTRUCT
  - Involving key European scientific communities, technology providers and developers, resource providers, industries and e-infrastructures.
- Info: davide.salomoni@cnaf.infn.it

3.bis- My (very long term) framework

- STABLE

    Analysis Toolboxes

    "transparent" access to resources/apps
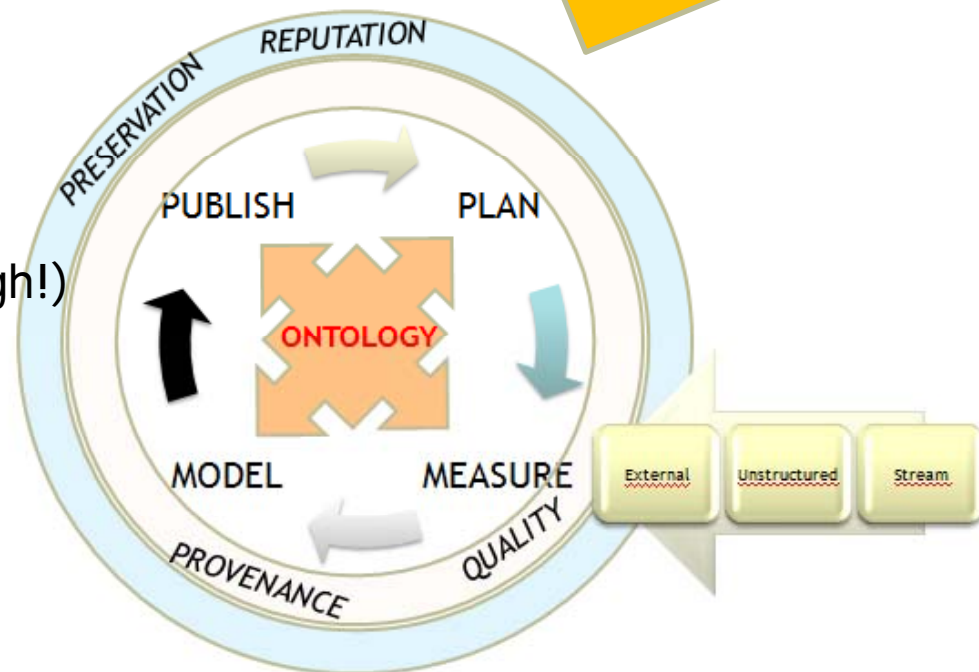
> You don't use IPython?
> in your tablet?

> My students learn R
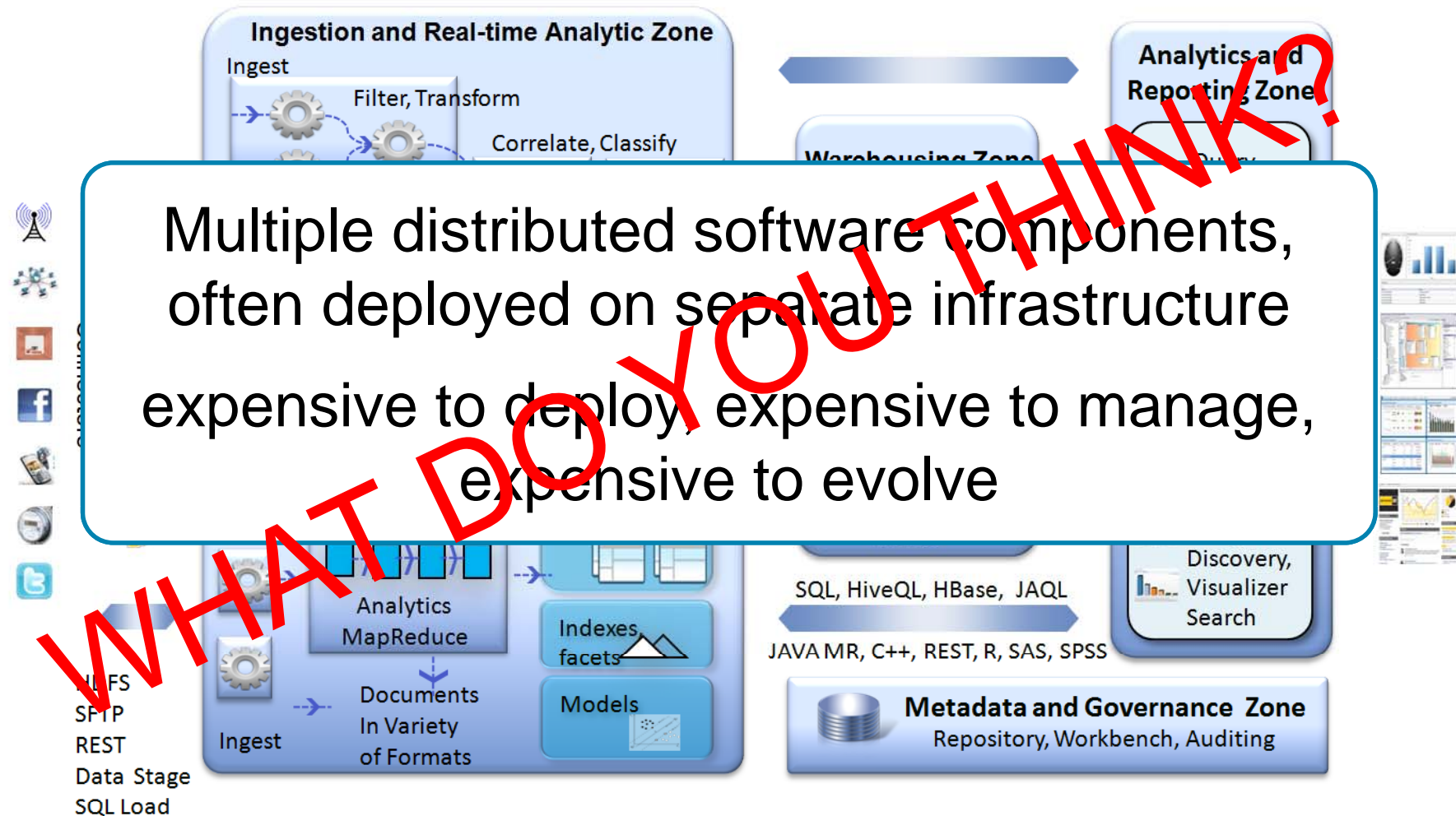> R.Guralnick, COOPEUS meeting Madrid

> Engineers use ANSYS,
> MATLAB, COMSOL...

- PRESERVATION included
  since project start

    (Data Management Plan is not enough!)

REPUTATION
PRESERVATION
PUBLISH   PLAN
ONTOLOGY
MODEL   MEASURE
PROVENANCE   QUALITY

External   Unstructured   Stream

- FLEXIBLE TO SUPPORT DIFFERENT TECHNIQUES

Complexity - A Key Customer Challenge

Multiple distributed software components, often deployed on separate infrastructure

expensive to deploy, expensive to manage, expensive to evolve

WHAT DO YOU THINK?

OK, now let's do econo-socio-physics

- Idea: Model Anthropogenic effects using Agents (in the Cloud)
  - A new ABM suite
    - existing ones like REST, are not (yet) cloud oriented
  - Ad-hoc cloud/HPC/Grid messaging for agents (?)
    - mqueue? MPI in the cloud?
  - Model any activity
    - Example: all components of urban life, all rural impact sources
  - Interface to models and to sensors (validation?)
- Technical Idea:
  - ABM + integrate sensors + GIS (for "environment")
  - + non-SQL DB (tracking activities) + analytics
- What for?
  - Analyze and Optimize "smart" ideas
    - Example: design of sensors in cities (Friendly cities)
    - Example: design of exploitation activities in rural environment

4.- AN UNCERTAIN FUTURE

- **FOCUS ON KNOWLEDGE (btw, aren't we simply Knowledge?)**
  - By 2025 computers will be able to handle most human tasks

    > **Will we be able to "ask" to computers about anything? Cf. WATSON**

  - So, computers will "preserve" knowledge, but...

    will computers generate "new" knowledge? will we know that?
  - What about us? Will we be preserved?   **I have no idea at all!**

- **ENGAGING SOCIETY**
  - Yes, YouTube, movies and TV series are very ok but... PASSIVE!
  - We need URGENTLY to engage ACTIVELY young (15-20 y) students
  - Why not combine Citizen Science, Open Data Access, Edu Resources

  > **DATA ORIENTED COURSES where students start by hand on a REAL problem, learn, contribute, using EDUCATION resources into e-INFRA. Ex: Image Recognition**

Am I alone?

Is my BSE test ok?