



# Summary and steps ahead

Federico Carminati  
HEP Software Collaboration meeting  
April 3-4, 2014



# Message from Bertolucci



- Not much time to build a new software that scales with Moore's law
  - Already in trouble from Run 3
  - Solve not only LHC problem but also other HEP and other communities.
- Who are owners (LHC experiments?)
- Where are resources? (software is not making career of young physicists)
- Change of culture needed. Require organisation to bring new resources.
- Stability needed by running experiments while we develop new software
- Keep existing code (Geant4 & Root) stable
- Agree with experiments on version to be maintained
- Concentrate development effort on new software to be rewritten.
- Governance will play key role. Should indicate how process is evaluated and tracked (which bodies/committees)



# Summary – The vision

## HEP Software Challenges – LHC Focus

SLAC

- ◆ LHC has an increasingly data-intensive future
- ◆ Technology evolution will not meet our needs (at least not without a lot of work)
- ◆ We could ‘survive’ by tightening our focus and our triggers → ensure that we ignore the truly unexpected
- ◆ Much better to evolve our software process to exploit the complex shifting landscape of computing in the next 20 years (massive parallelism; changing cost balance between computation, memory, storage and networks)



# Summary – The recent experience



## Summary

---

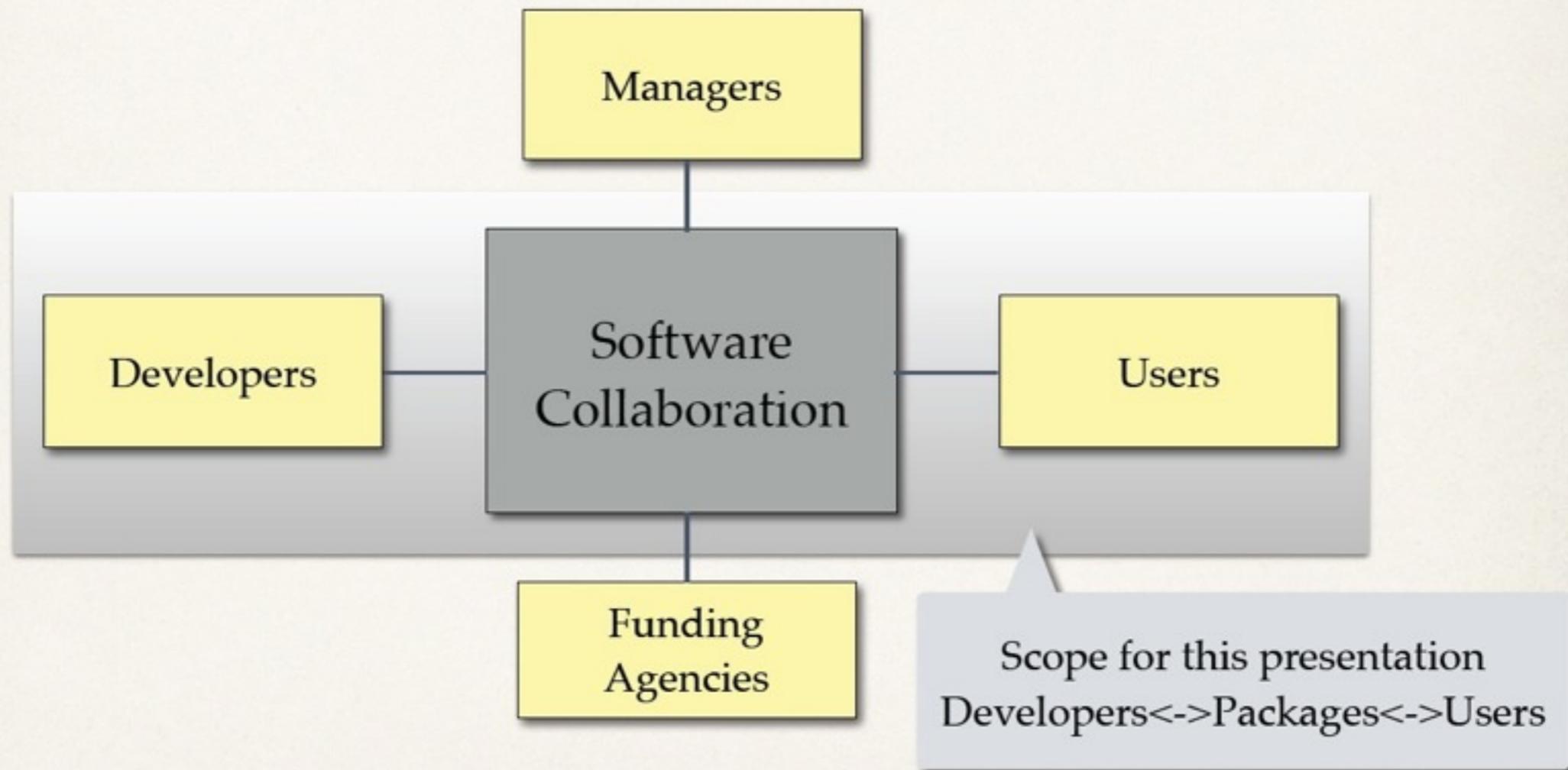
- HEP has a lot to offer the world at large, and not just for "distributed computing". To the extent that this is expected of us, we can probably deliver if we learn how.
- The rest of the world also has a lot offer us, we should embrace that, too...
- This "bug hunt" surely missed a lot of things: if you know of other interesting software packages, fields with interesting development models or have ideas as to where collaborations as above might be beneficial, I would be interested to hear them...



# Summary – Software development

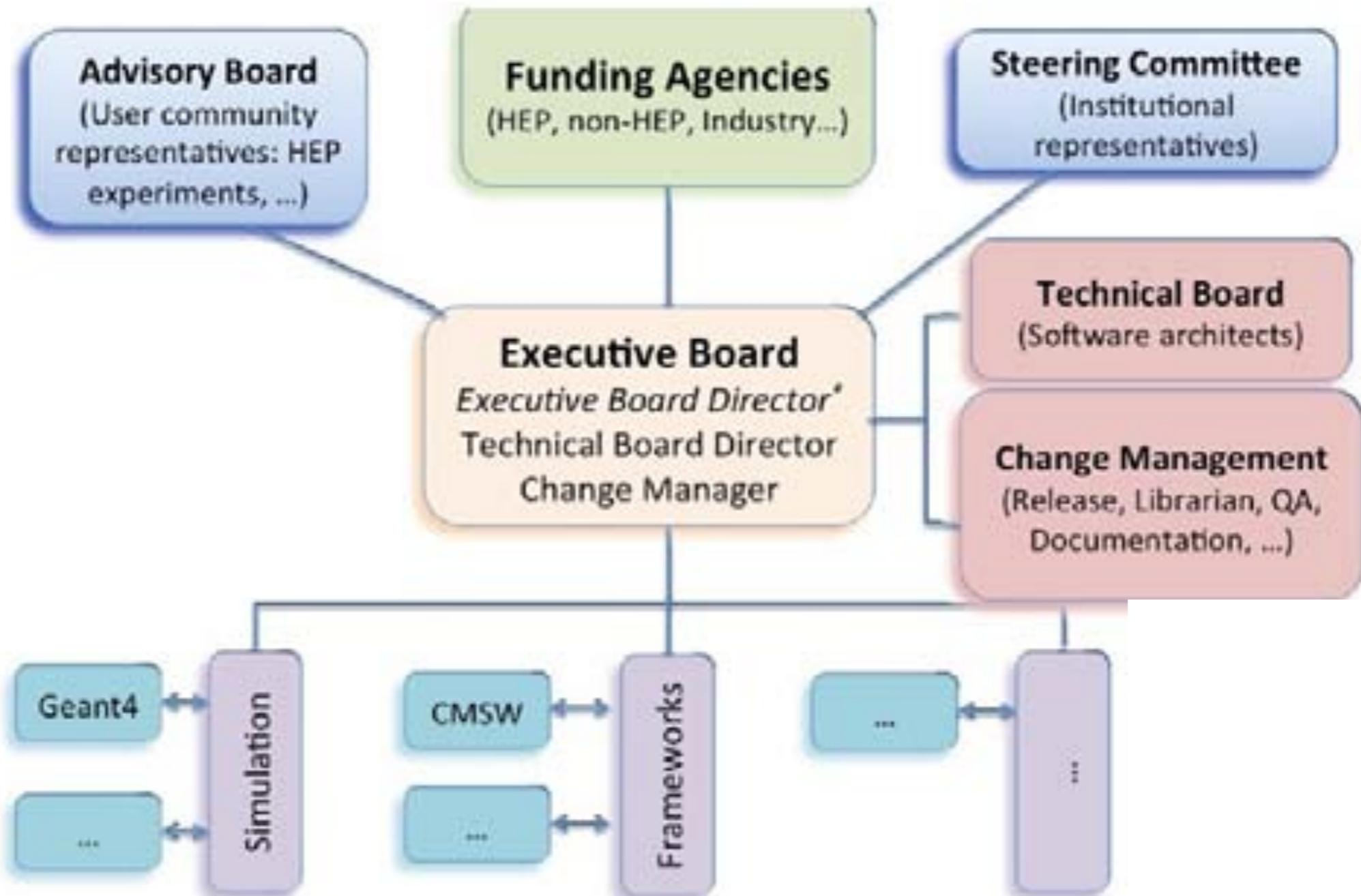


## Collaboration Stakeholders





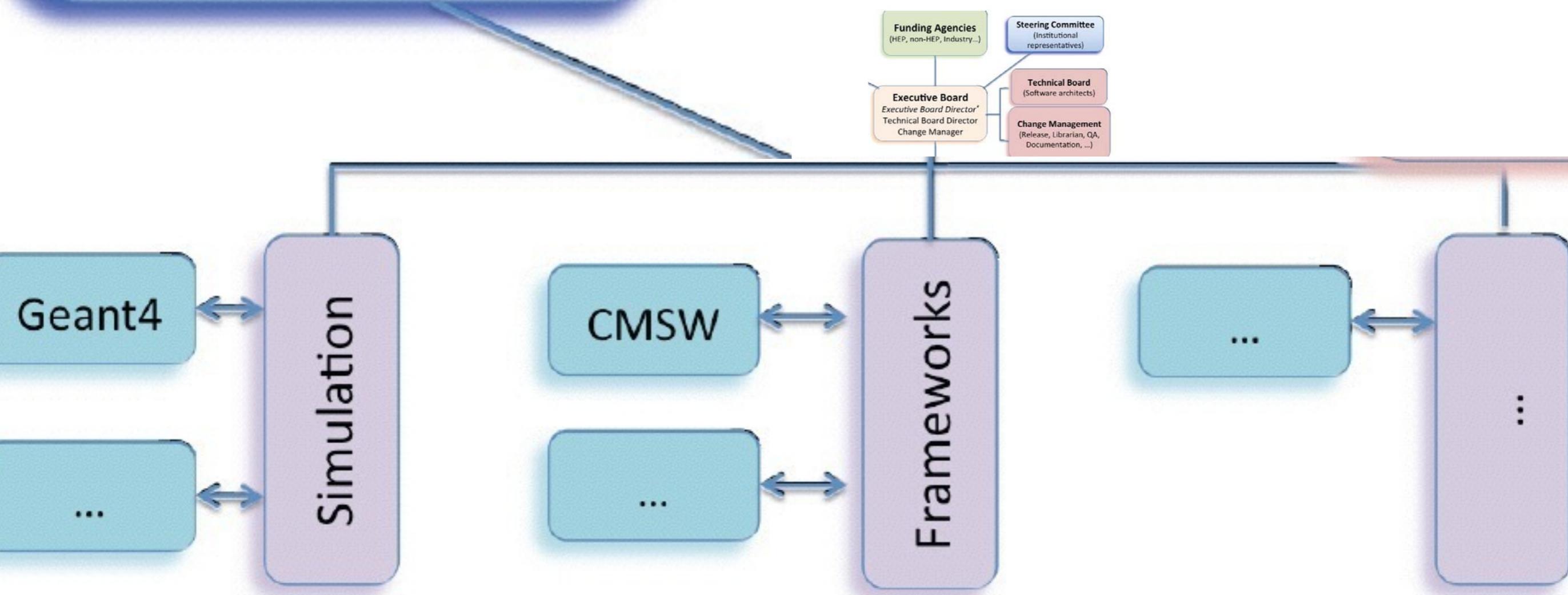
# Summary – Governance





# - Governance

**Advisory Board**  
(User community  
representatives: HEP  
experiments, ...)



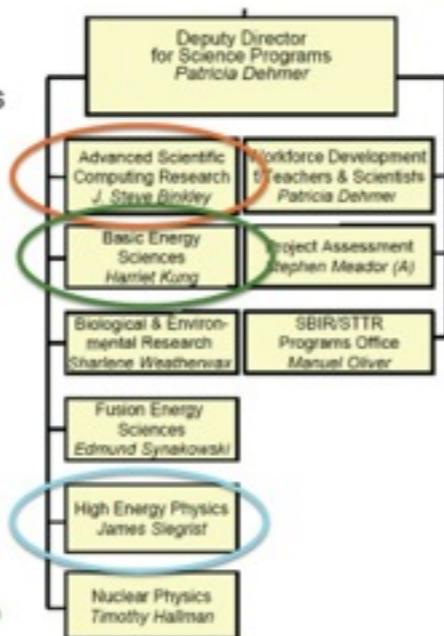


# Summary – Funding



## DOE funding landscape

- Computing and computational activities interdisciplinary, but DOE funding model not set up to easily support such programs
  - Different offices mainly focus on the portfolio of their specific domains
  - There are partnership programs for common or complementary activities
    - Not as common, harder to create
    - Funding usually remains segregated flowing to “domain” researchers from same “domain” offices
      - But at least resources are leveraged
- To first order, relevant offices
  - HEP (*direct*)
  - ASCR (*partnership* and *direct*, but partnership an advantage)
- BES a target of opportunity for *partnership*



## Excellent Science

<b>European Research Council (ERC)</b> Frontier research by the best individual teams	<b>13.1</b>
<b>Future and Emerging Technologies</b> Collaborative research to open new fields of innovation	<b>2.7</b>
<b>Marie Skłodowska-Curie actions (MSCA)</b> Opportunities for training and career development	<b>6.2</b>
<b>Research infrastructures (including e-infrastructure)</b> Ensuring access to world-class facilities	<b>2.5</b>

(all figures in billion euro, over the period 2014-2020)

3 apr. 2014

Software collaboration meeting

10



Fermilab  
EU-T0 topics

A strategy focusing on a list of main scientific and technological priorities:

- homogeneous evolution of current DCI computing model;
- shared archival and data access/analysis services, together with tools and applications for scientific data analysis;
- **federating interdisciplinary modern software developments;**
- issues, models, policies and services for data preservation;
- interfaces with the private sector aiming to be a “pilot” public/private data management and processing system;
- overtaking High Performance and High Throughput Computing paradigms;
- preparation of “data scientists” aimed to lead the major changes in e-Science.

## Funding opportunity for software development in Japan

- We need to carry out a research in borders of disciplines to seek more budget opportunity
  - Priority on software/computing is rather low in the Japanese HEP society
  - Services for other disciplines result more budget
- Borders of disciplines are new frontiers in research
  - Not only funding opportunity, but also new job opportunities
  - Feedbacks from new users will benefit people in HEP
- ILC will be another opportunity
  - The budget necessary for feasibility study and budget profiling is funded in JFY2014
  - No other budget has not been funded

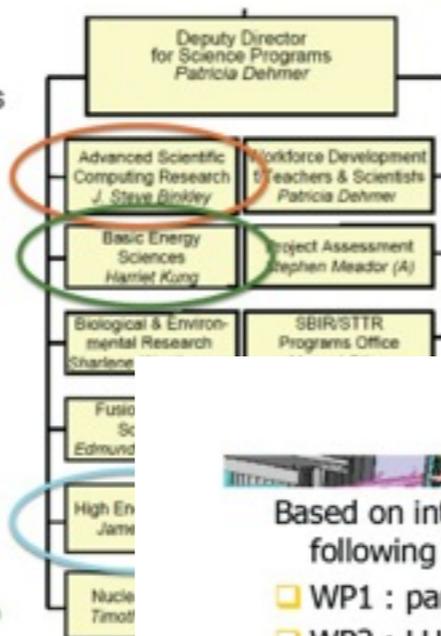


# Summary – Funding



## DOE funding landscape

- Computing and computational activities interdisciplinary, but DOE funding model not set up to easily support such programs
  - Different offices mainly focus on the portfolio of their specific domains
  - There are partnership programs for common or complementary activities
    - Not as common, harder to create
    - Funding usually remains segregated flowing to “domain” researchers from same “domain” offices
      - But at least resources are leveraged
- To first order, relevant offices
  - HEP (*direct*)
  - ASCR (*partnership* and *direct*, but *partnership* an advantage)
- BES a target of opportunity for *partnership*



## Excellent Science

<b>European Research Council (ERC)</b> Frontier research by the best individual teams	<b>13.1</b>
<b>Future and Emerging Technologies</b> Collaborative research to open new fields of innovation	<b>2.7</b>
<b>Marie-Curie actions (MSCA)</b> Training and career development	<b>6.2</b>
<b>Structures (including e-infrastructure)</b> World-class facilities	<b>2.5</b>

## Work packages

- Based on interest and experience of the partners the following work packages have been defined:
- WP1 : parallel framework (GaudiHive) (ATLAS, LHCb)
  - WP2 : LHCb upgrade tracking on accelerator or not, trigger and offline
  - WP3 : CMS portable software on accelerators
  - WP4 : statistical end-analysis, (confidence level computation)
  - WP5 : FastJet (original authors are “french” physicists originally from LPTHE)

Note: we intend to be part of existing working groups within or between experiments, not start our own.

David Rousseau, LPaSo, HEP sw collaboration 3 April 2014

5

(in billion euro, over the period 2014-2020)

Software collaboration meeting

10

development in Japan

in borders of disciplines to seek more

rather low in the Japanese HEP society

result more budget

- **Borders of disciplines are new frontiers in research**
  - Not only funding opportunity, but also new job opportunities
  - Feedbacks from new users will benefit people in HEP
- **ILC will be another opportunity**
  - The budget necessary for feasibility study and budget profiling is funded in JFY2014
  - No other budget has not been funded



A strategy focusing on a list of main scientific and technological priorities:

- homogeneous evolution of current DCI computing model;
- shared archival and data access/analysis services, together with tools a data analysis;
- **federating interdisciplinary modern software developments;**
- issues, models, policies and services for data preservation;
- interfaces with the private sector aiming to be a “pilot” public/private data management and processing system;
- overtaking High Performance and High Throughput Computing paradigms;
- preparation of “data scientists” aimed to lead the major changes in e-Science.



# Points of agreement

- Clearer collaboration guidelines / organisation will increase efficiency & bring new resources
- Collaboration with other sciences will further increase our opportunities to have better and more robust software and will increase HEP visibility



# Main issues under discussion

- University and lab participation will meet different view points from funding bodies
- Make management lightweight but functional
- Make sure that users (experiments) have a central role in driving the process
- Preserve successful projects and maintain stability and user investment
- Properly involve other communities and in particular Computer Science





# The problem

- The IT (r)evolution is accelerating & our needs expanding
- Taking advantage of the new opportunities is both necessary and more and more difficult
- Only via the mutualisation of the competences and the efforts of the whole scientific community we may provide real answers to problems like
  - Data preservation
  - Exploitation of new architectures
  - Distributed heterogeneous computing (Cloud)
  - Big data
- Failure to do this would leave us with sub-optimal solutions that will limit our scientific output and our scientific and societal impact

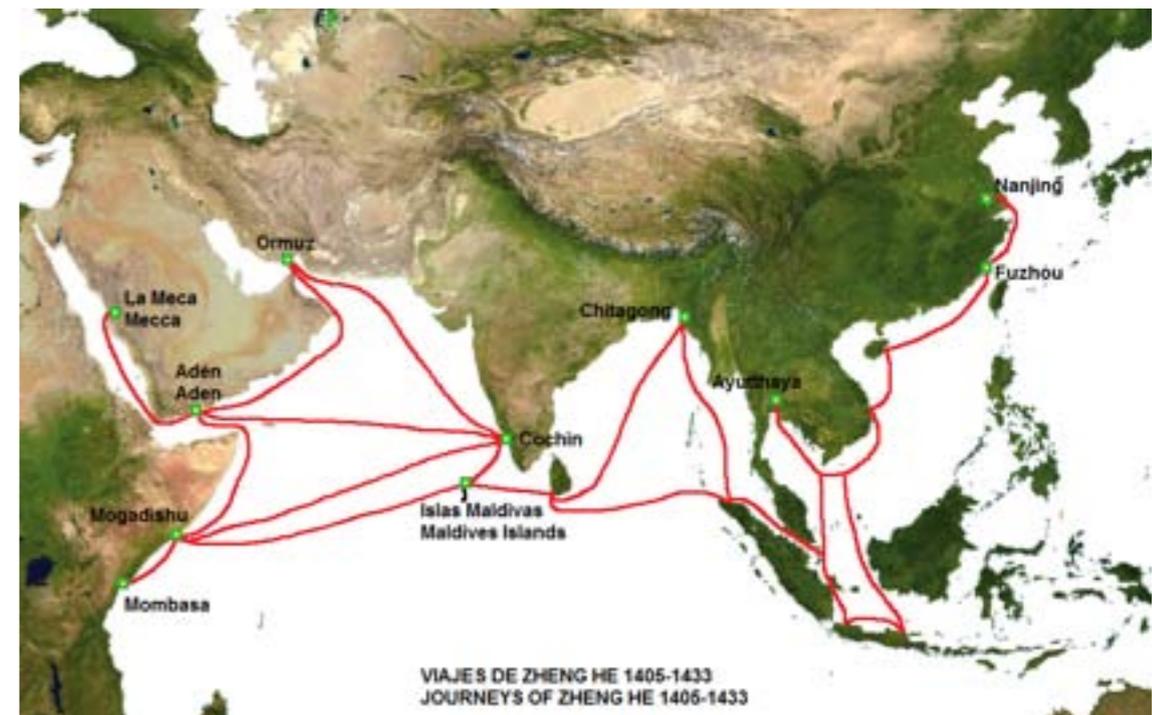
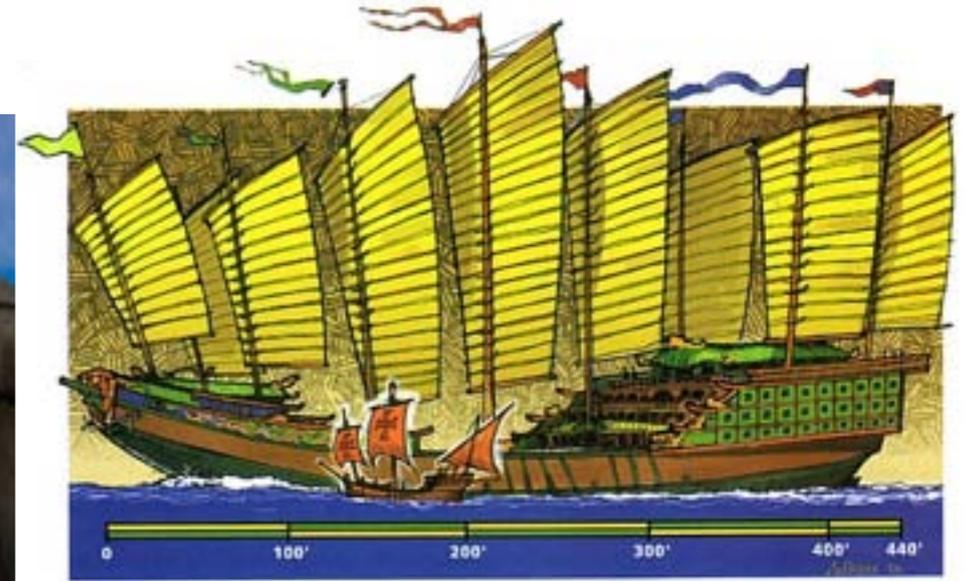


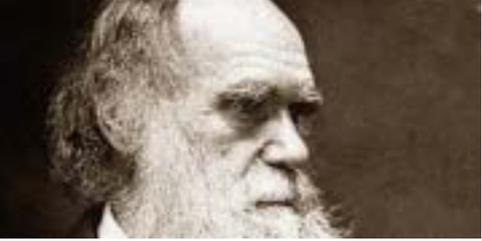
# The challenge

- Sharing software development will allow us to
  - Profit from developments done (better!) elsewhere
  - Avoid going down blind alleys and duplicate efforts
  - Increase the scientific and societal impact (spinoff) of HEP software
  - Increase robustness and functionality of our software thanks to extended user-base and cross-fertilisation of ideas and algorithms
  - Increase long-term sustainability and standardisation of our code

# Why China did not discover America?

- In 1426 the Xuande emperor ordered Zheng He's fleet to be destroyed after one last voyage
- Competition between European nations was essential to Columbus's trip





# A Darwinian world

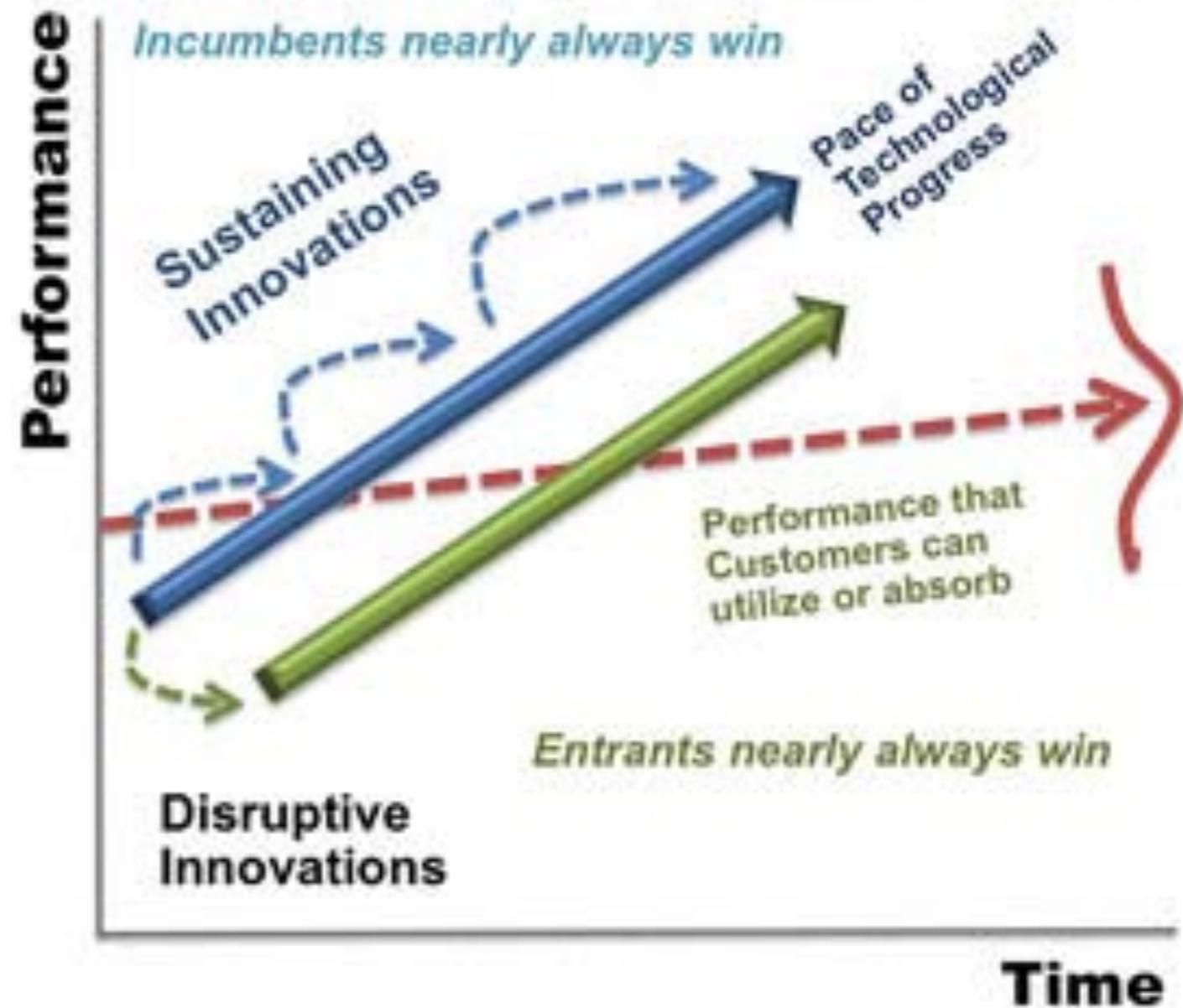
- If you allow NO variety and you kill “mutations” you kill evolution
- If you “protect” “unfit” mutations and do not let them die you also kill evolution
- “Competition between motivated developers”
- A Darwinian balance should be found between these two extremes



# Evolution

- This is now an acquired concept
- How can we build it into our work without “dramatisation”
- Creating, controlling and fostering internal competition
- Without disrupting people

## The Disruptive Innovation Model





# The scope

- The scope of the project is all the software that has to be developed because it cannot be found (yet) in the Open Source community
- This includes all common simulation, reconstruction and analysis software
- It also includes packages at the boundary between application software and middleware (e.g. XrootD or CernVMfs)



# The users

- The whole process should be user driven
  - Not design driven
  - Not Funding Agency driven
  - Not Committee driven
  - Not review driven

**USER DRIVEN**



# The users

- The whole process should be user driven
  - Not design driven
  - Not Funding Agency driven
  - Not Committee driven
  - Not review driven

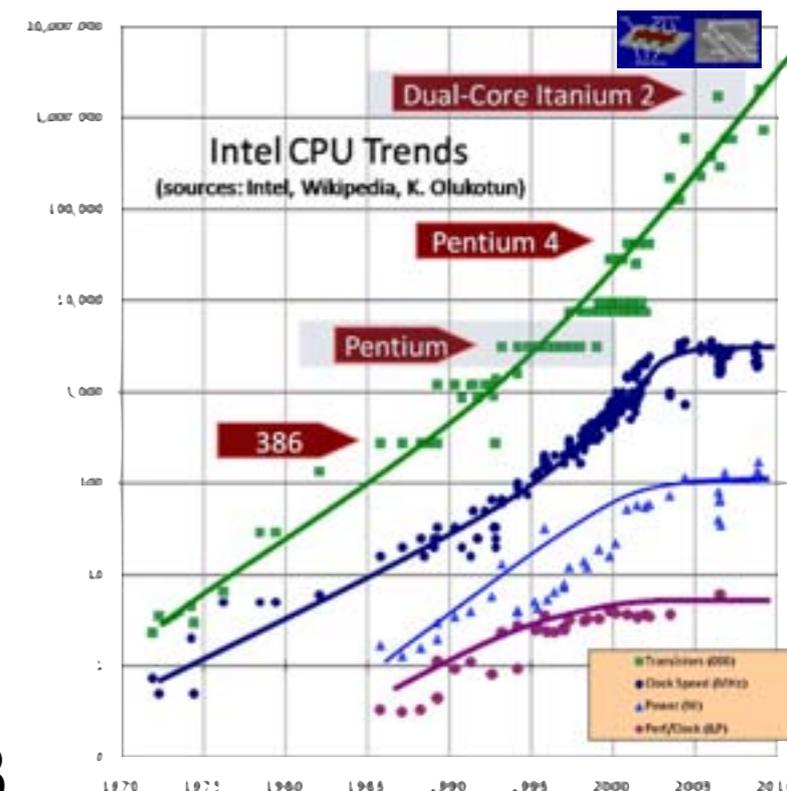
## USER DRIVEN

...WE HERE HIGHLY RESOLVE THAT A  
PROJECT OF THE USERS BY THE USERS AND  
FOR THE USERS SHALL NOT PERISH ...



# Why now?

- 1950 - 1990 the reign of FORTRAN
  - Attempts to move to C / Java were ignored / unsuccessful
- 1992 - 1996 soul searching (F90 / C++)
  - The community got together: MONARC, LCB...
- 1996 - 2000 the building of the frameworks
- 2000 - 2012 people get to work and build the software
- 2012 - 2013 realisation that clock cycle won't improve and that we need radical changes to access more computing power
- Time is now for a new for a new initiative within and beyond HEP





# The development process

- Modern Software Engineering, fuelled by the Open Software revolution and years of scientific software development, indicates
  - Open Source
  - Agile Technologies
  - Collaborative distributed development
- As the conditions of successful development
- The success or failure of the initiative will depend on how well we integrate this fact in our project
  - And we innovate even in this area



# A note about “sound software practices”



- Should they exist we would know about it



# The structure

- Major areas of work should be identified
  - Without building Chinese walls between them!!
- The work should be organised in limited-scope, limited-time projects each within one (or more) reference areas
- “Institutionalisation” of these projects should be avoided in order to maintain flexibility and do not “lock” manpower
- A viable model for the maintenance of existing products should be elaborated



# The management

- This initiative has to be managed and not simply coordinated, which means
  - Steer and even stop projects
  - (re-)deploy resources to critical areas
- This requirement is in partial contradiction with the one before
- Management must be lightweight as IT is a fast evolving environment
- A delicate mixture of bottom-up and top-down approaches must be implemented
- A powerless or paralysing management can doom the project



# The management

- The stronger and clearer the agreement on
  - Interfaces
  - Procedures
  - Conventions (in the wide sense) and “behaviours”
  - Component layout & interfaces
  - Standards for software documentation!
  - Release, testing and non regression process
  - Roles & procedures for architects and code librarians
- The wider the freedom of the developers within these limits

# The management

- A combination of
  - Shared vision
  - Clear (consensual) rules
  - Large delegation of responsibilities to projects
  - Serious project reviews
- Can lead us to success



# Extending the collaboration

- Creating a HEP software initiative is only the first step
- Involvement of other sciences is mandatory to reach critical mass and obtain meaningful results
- A rationalisation of our field is however the first step before we involve new partners
- Or otherwise said, if we fail even amongst ourselves...
- Existing collaborations with other fields should be of course preserved and even strengthened



# Reviewing

- Management should act based upon
  - The available resources
  - The stakeholders requirements
  - An independent assessment of the results
- Peer review is essential to maintain coherence and scientific and technical excellence
- Common practices in the scientific world offer a very good model here

# Outreach

- The activity of the project should be properly documented and exposed to the public
- Public actors and funding agencies should be fully aware of the impact of the project
  - We can give away the code NOT the IPRs!
- Prospect collaborators should be attracted by a clear outreach strategy
- The whole initiative should be clearly identified as the source of the supported products
  - A logo, a brand name etc.



# IPRs

- Intellectual Property Rights of HEP software are most probably “undefined”
- The new project should have a clear and sound policy for IPRs
- Full recognition for authors and institutions should be assured
- The software should be open and available to the world
- A model for commercial exploitation of the software should be established without hindering open sharing and scientific development



# Funding

- This will simply not be possible if existing entities (FAs, labs, experiments) do not buy in assigning resources
  - Which will be managed by the project!
- Entities participating to the initiative will agree to coordinate their requests for funding to the different FAs
  - European Union, NSF, DoE...
- There should be no “internal competition” (unless intended!) or contradictory or duplicate requests



# From here to there

- The half full glass
  - HEP software has been overall very successful (Grid, Geant4, CLHEP, Root, XrootD, EOS, CernVMfs...)
  - The “Higgs code suite”!
- The half empty glass
  - Two major systems (Root, Geant4) with a lot of room for improving synergy
  - Limited modularity of both systems
  - There could be more commonality between experiments



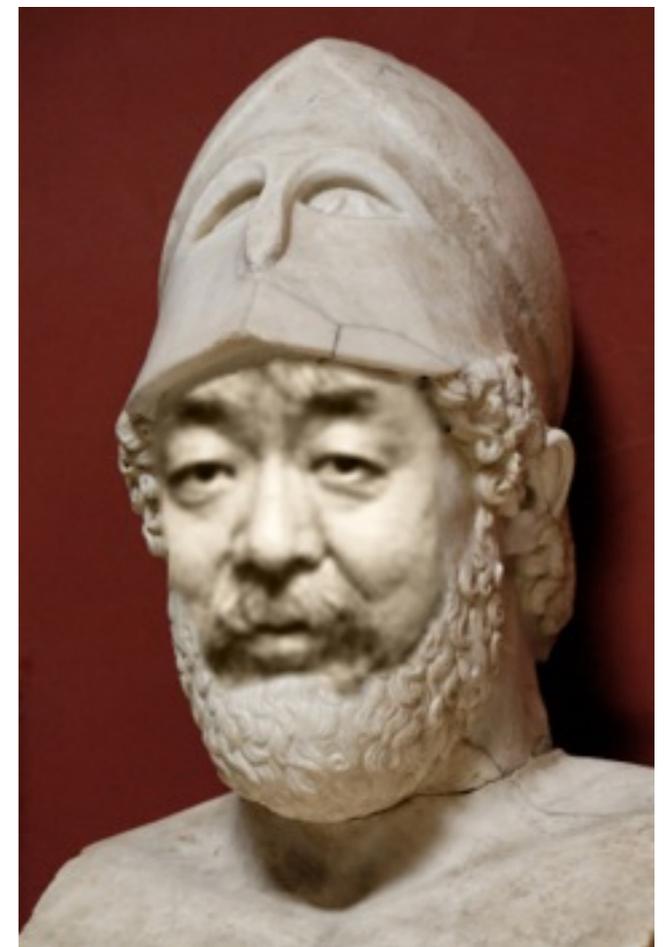
# Taking the first step

- The initial elements of the initiative are the current HEP software
  - Root, CLHEP, XrootD, Geant4, GeantV...
- Consensus should be found to “recast” these well established projects into the new structure
  - Streamlined governance, development and support model
  - Enhanced participation of the stakeholders (experiments!) in the governance
  - Enhanced logical and technical modularisation of the components
  - Assignment to project areas
  - “project-oriented” developments objectives for the different components
- This process must seamlessly preserve functionality and support for existing users



# Recasting projects

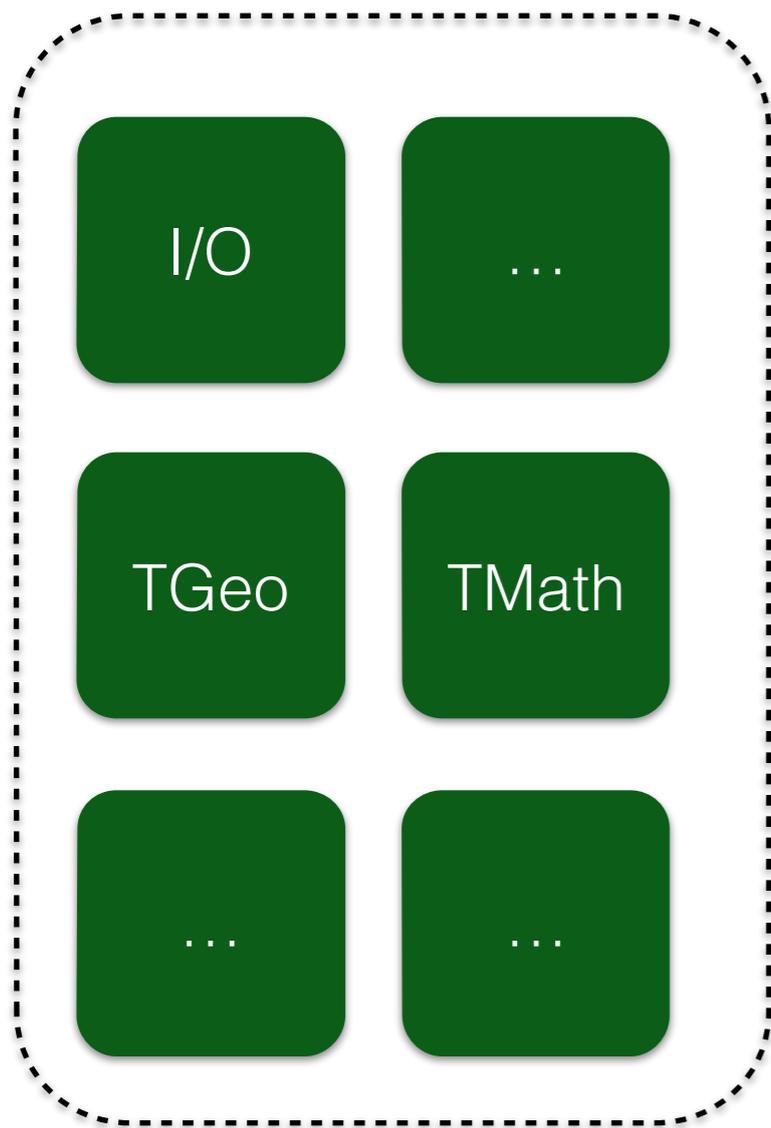
- The two larger components are ROOT and Geant4
- These have two very different “governance” models
- We should learn from both and find an effective synthesis



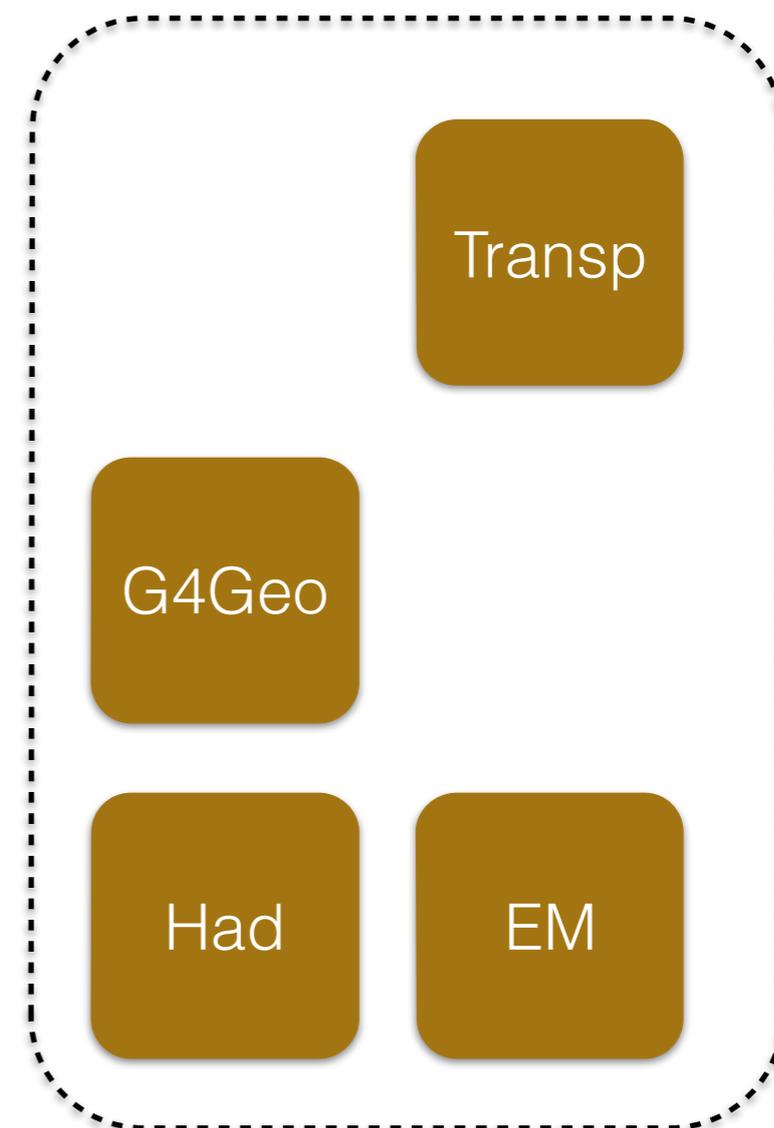


# Accelerating evolution

ROOT



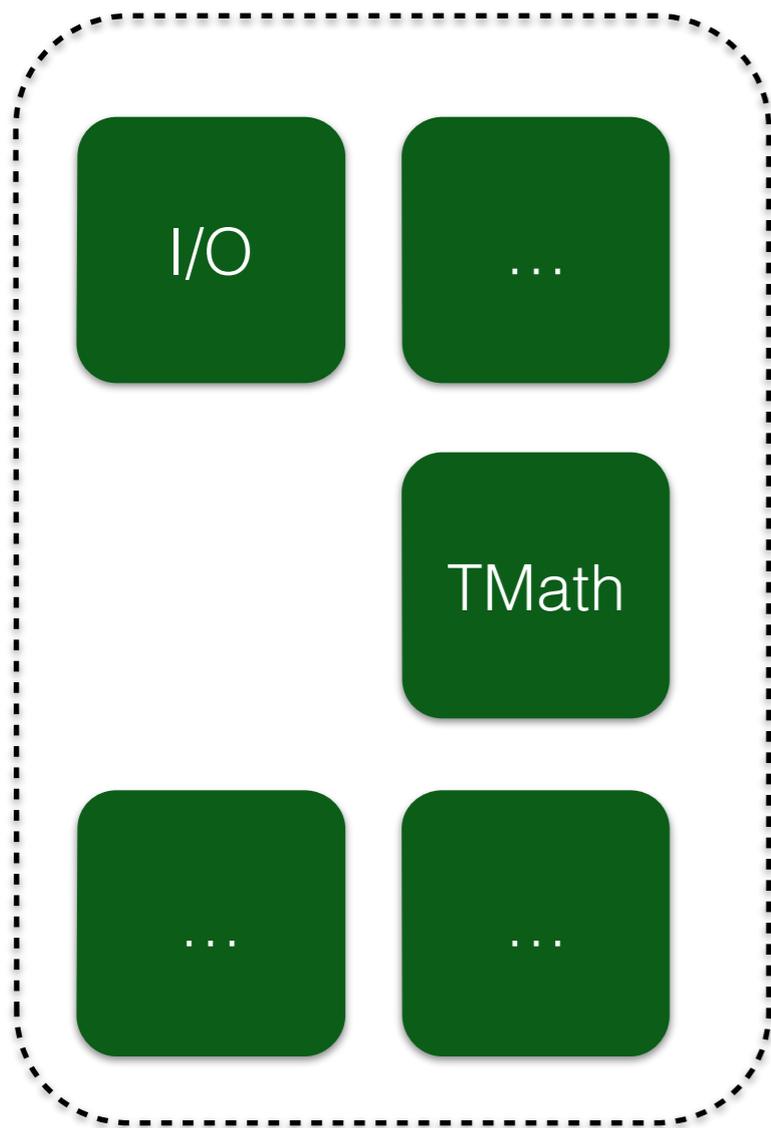
Geant4



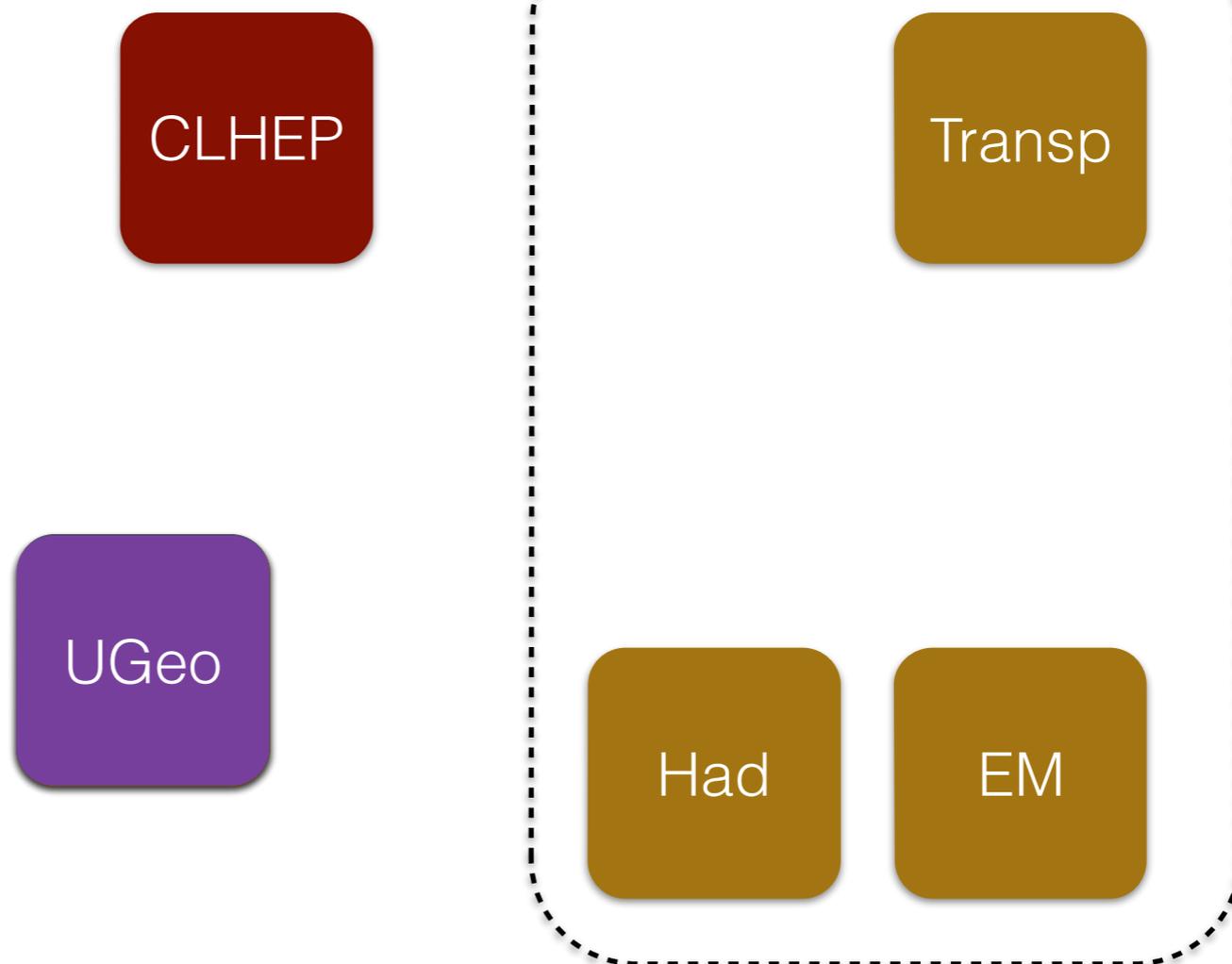


# Accelerating evolution

ROOT



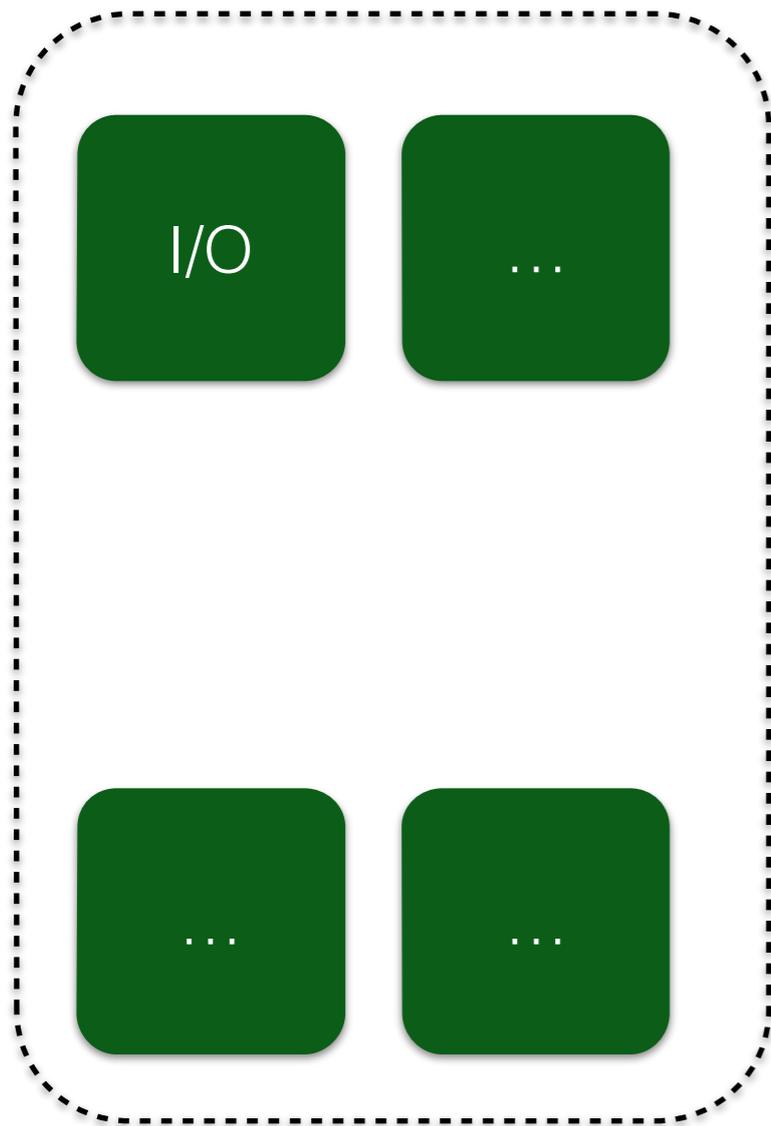
Geant4



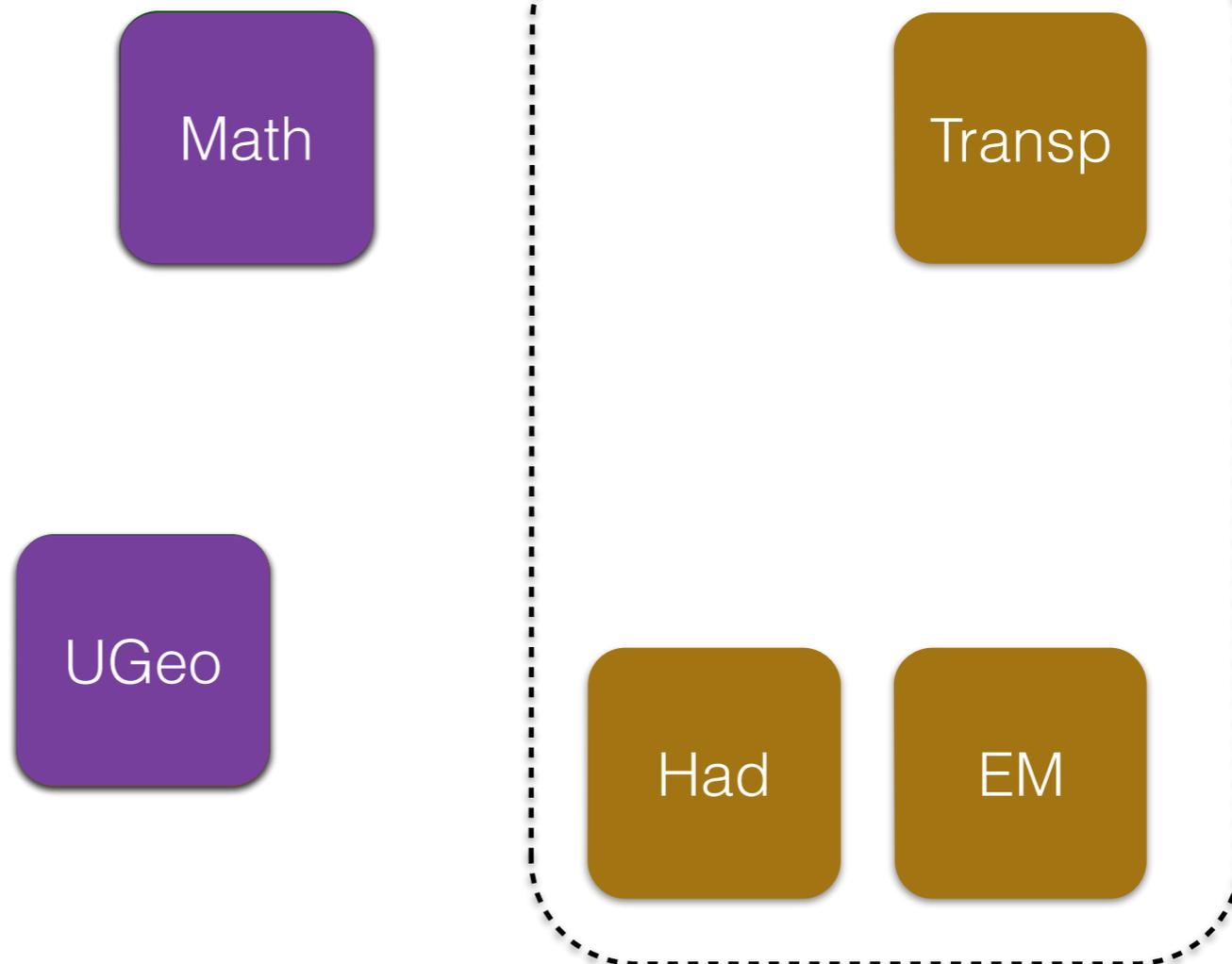


# Accelerating evolution

ROOT



Geant4





# Accelerating evolution

I/O

...

Math

Transp

...

...

UGeo

Had

EM



# Our next objective

- Define a “blueprint” of the collaboration (or better foundation) based on the consensus reached
- Meet again in Fall this year to finalise the agreement and start operation



# How do we get from here to there?

- Setup a blueprint editorial board
  - Working mostly virtually via email
- Produce a set of “chapters” by Summer
- Finalise the content in September
- Meet mid October to
  - Approve the document
  - Name the responsible persons in the organigram
  - Start



# Content of the document

- The general layout of the document could be
  1. Introduction
  2. Scope and duration
  3. Development model
  4. Components and frameworks
  5. Funding structure
  6. IPRs policy
  7. Reviewing structure
  8. Governance
  9. Conclusions

# Conclusions

