21st International Conference on Computing in High Energy and Nuclear Physics **CHEP2015** Okinawa Japan: April 13 - 17, 2015

Planning for the future

Ian Bird

WLCG Workshop

Okinawa, 12th April 2015

60 YEARS/ANS CERN

WLCG

LHC Upgrades and computing

- CERN Council asked for better understanding of costs for detector upgrades and computing
- At December Council it was agreed that HL-LHC should make a proposal to be part of the ESFRI roadmap
  - ESFRI==European Strategy Forum for Research Infrastructures
    - E.g. ELIXIR, ESS, XFEL, CLARIN, CTA, SKA, …
  - Implies recognition as a major research infrastructure and opens way to significant EC funding lines
  - Computing (E-NEEDS) are a part of the overall project
- The proposal to ESFRI was submitted at end of March, including section on computing ➔

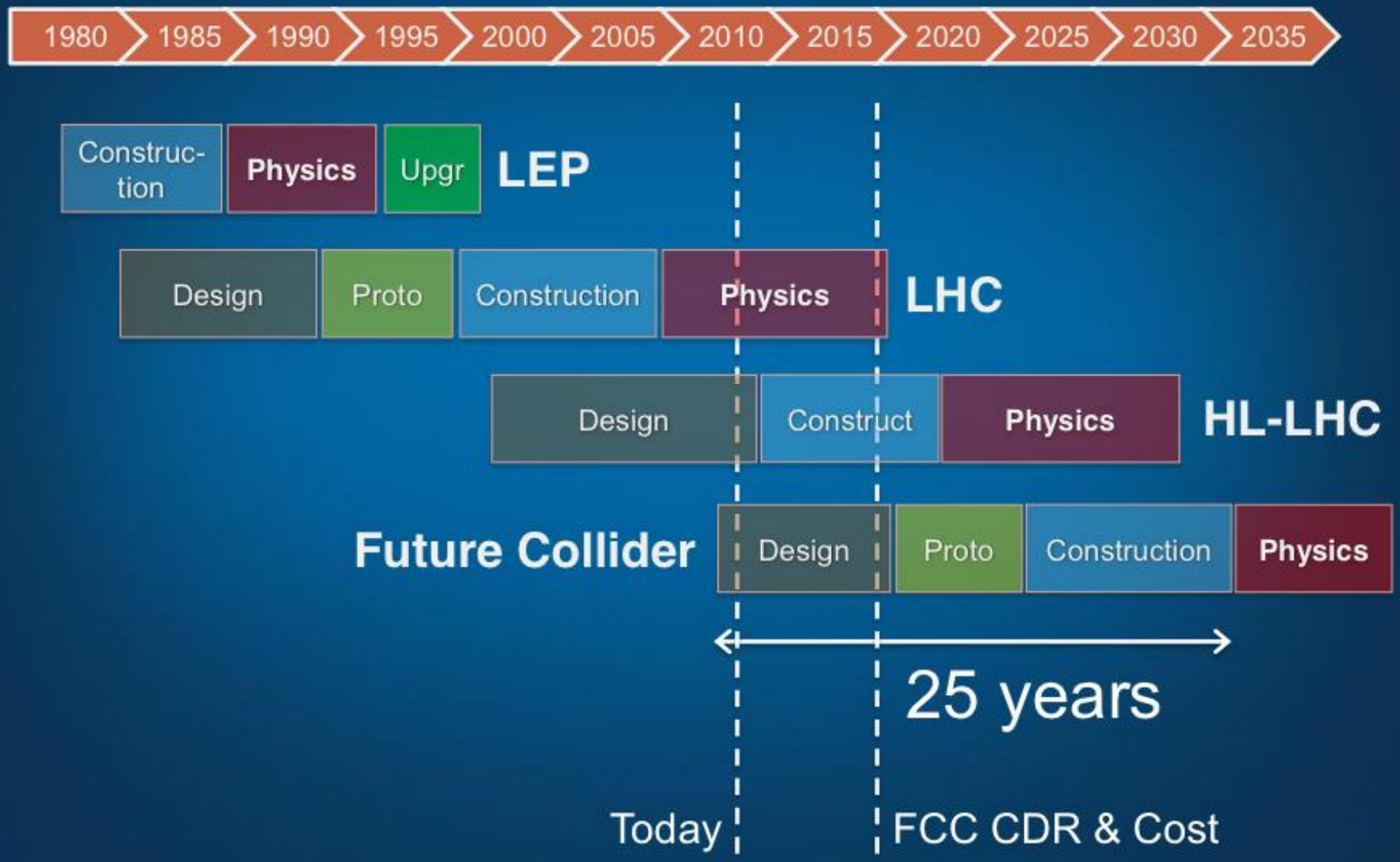
Computing part of ESFRI proposal

- E-NEEDS: (computing part)
  - What will be the data management and open data policy of the RI? (Would data become accessible from a repository to the public? Would the RI be interfaces to e-infrastructures for science?)
  - What is the plan for supporting advanced data management and how will it be funded?
  - What is needed (if applicable) from external e-infrastructure services (resources for storage, computing, networking, tools for data management, security, access, remote analysis etc.)?
  - Will the RI contribute to the development of e-infrastructure commons in the field or in general?
  - Will the RI policy on data include training services for “data practitioners” to enable the effective use of data repositories and data analysis tools by non-scientists?
- Other: (relevant for the entire project)
  - What will be the access policy of the RI …… access to data etc?
  - What are the linkages with existing platforms, and networks, …
  - What is the expected contribution of the RI to address H2020 societal challenges?
    - Computing can surely contribute here…

Planning towards HL-LHC
&(later)

- We need to start to develop the computing models for evolving towards HL-LHC so that we can (eventually) make some statements about costs and requirements
- Need to agree common working baseline expectations and performances
- Probably need to provoke some difficult discussions
  - Do today's models scale to HL-LHC? Suspect not.
  - Distinction between "online" and "offline"?
  - Physics performance vs costs of computing?
  - Etc.

HEP Timescale



Trends in HEP computing

- Distributed computing is here to stay
  - Actually we had it 30 years ago, and seriously 15-20 years ago
- Ideal general purpose computing (x86 + Linux may be close to the end
  - May be more effective to specialise
    - GPU and other specialised farms
    - HPC machines
    - Commodity processors (“x86”, ARM, etc)
  - Used for different purposes – lose flexibility but may gain significantly in cost

Trends – Data centres

- Moving data around the world to 100's of sites is unnecessarily expensive
  - Much better to have large scale DC's (still distributed but O(10) not O(100)) – connected via v high bandwidth networks
  - Bulk processing capability should be located close or adjacent to these
  - Data access via the network – but in a truly "cloud-like" way – don't move data out except the small data end-products

Data centres

- Our Data Centres may become exactly that – dedicated to data
- Compute resources are quite likely to be commercially available much cheaper
  - Don't know how they will be presented (hosted, cloud, xxx, …)
  - Already see today commercial compute costs are comparable to our costs
- Not likely, or desirable, that we will give up ownership of our data
  - Will still need our large data facilities and support

“Tier 2”-like resources

- Today these are crucial
  - >50% of CPU provisioned here
  - More importantly today these give access to the experiment data
    - And get us synergistic use of spare resources
- And, engagement of skilled people
- Don’t want to lose this
  - But there are many workloads that are still suited to this type of resource

Opportunistic resources

- Today this has become more important
  - Opportunistic use of:
    - HPC's
    - Large cloud providers
    - Other offers for "off-peak" or short periods
    - Etc.
    - All at very low or no cost (for hardware)
  - But scale and cost are unpredictable
- Also growing in importance:
  - Volunteer computing (citizen science)
    - BOINC-like (LHC@home, ATLAS/CMS/LHCb@home, etc)
    - Now can be used for many workloads – as well as the outreach opportunities

Trends – software

- Recognizing the need to re-engineer HEP software
  - New architectures, parallelism everywhere, vectorisation, data structures, etc.
- Set up HEP Software Foundation (HSF)
  - Community wide – buy in from major labs, experiments, projects
  - Goals:
    - Address rapidly growing needs for simulation, reconstruction and analysis of current and future HEP experiments,
    - Promote the maintenance and development of common software projects and components for use in current and future HEP experiments,
    - Enable the emergence of new projects that aim to adapt to new technologies, improve the performance, provide innovative capabilities or reduce the maintenance effort,
    - Enable potential new collaborators to become involved,
    - Identify priorities and roadmaps,
    - Promote collaboration with other scientific and software domains.

Evolution of facilities

- Today we have LHC/WLCG as the computing facility
- Recognise that between now and FCC, we have potentially many international facilities/collaborations involving global HEP community
  - Bearing in mind we have possibly many international or global HEP challenges: Neutrino facility, ILC, CLIC, FCC and others as well as large experiments such as Belle-II that ask to use "WLCG"
  - And not forgetting the possible commonalities with related projects (SKA, LSST, CTA, etc) where facilities may be heavily shared
- What is the process to build on our working infrastructure to evolve towards HL-LHC, FCC, etc. serving the needs of these facilities and learning from them?
- How should WLCG position itself to help build a common global infrastructure that evolves through these coming facilities?

Evolution of structure

- Distinguish between infrastructure and high level tools
- We need to continue to build and evolve the basic global HEP (+others) computing infrastructure
  - Networks, AAA, security, policies, basic compute and data infrastructure and services, operational support, training, etc.
  - This part MUST be common across HEP and co-existing science
  - This part must also be continually evolving and adapting with technology advances
- Need a common repository/library of proven and used middleware and tools
  - A way to help re-use of high and low level tools that help an experiment build a computing system to make use of the infrastructure
  - The proto-HSF today could be a seed of this
- We must try and make this a real common effort and remove a lot of today's duplication of solutions
  - While retaining the ability and agility to innovate
  - The cost of continuing to support unnecessary duplication is too high

This workshop:

- Start some discussions and perhaps organise working groups or task forces to start to work on preparation for the future
- Have as input the work that was done for the ECFA workshops
- Needs strong input from the experiments
- We need to understand what the role of WLCG will be in this
- Failing to develop commonalities will lead to too high costs
  - LHCC, Funding agencies, & etc. are following this and asking hard questions now