

## EOSC Proposal Extract (CERN, WLCG, ...)

### Physical Sciences, Materials and Analytical Facilities

Proposed Science Demonstrators from the Physical Sciences, Materials and Analytical Facilities include:

...

CERN relies on the extensive data processing capacities of the WLCG for its locally generated data that is now being supplemented in a hybrid model by off-site Cloud IaaS resources (e.g. <http://www.hnscicloud.eu/>). While the scale of WLCG is currently unique, many of the issues involved are discipline-independent and hence profit from cross-domain collaboration. Some communities have addressed long-term data preservation as a pre-requisite of sharing and re-use, by adopting Open Science. Findability, re-usability (in terms of reproducibility of results in subsequent years) and scale (data volume and processing complexity) are key challenges. Even a small fraction of the data processing and the Open Science activities will require several to many petabytes of data.

...

### The 5 initial Pilot/Flagships will be developed that show benefit in the following domains:

...

- WLCG: Led by CERN, large-scale long-term data preservation and reuse of physics data.

**Key output:** *a number of PI driven science cases derived from in-flight RI services and partnerships*

### To be able to both work effectively from the very beginning of the project, 5 Science Demonstrators were pre-selected as candidates during the writing of the proposal:

...

WLCG: Led by CERN, through this demonstrator large-scale long-term data preservation and reuse across both private and public cloud resources and open science services and tools can be exercised.

...

Science Demonstrator Title

WLCG Open Science Demonstrator

Contact Name(s) and Organisation(s)	Bob.Jones@cern.ch; Helge.Meinhard@cern.ch; Jamie.Shiers@cern.ch
Demonstrator Description	<p data-bbox="826 302 1402 369"><i>Demonstrate scientific excellence and societal impact</i></p> <p data-bbox="826 369 1402 728">CERN is a practitioner of Open Science continuing on from the role played in terms of Open Access to publications. The LHC experiments all have data policies that call for their data to be preserved, for public releases of subsets of their data after a relatively short embargo period and for the ability to reproduce (at least key) analyses. (Data must be preserved before it can be re-used or shared).</p> <p data-bbox="826 728 1402 862">Experience from these Open Data releases has shown benefits not only in closely related disciplines (e.g. theoretical physics) but also in computer science in general.</p>
Science Area	High Energy Physics (Physical Sciences and Engineering according to ESFRI classification)
Relationship to European/national research organisations, Research Infrastructures, & equivalent Initiatives	<p data-bbox="826 940 1402 1041"><i>e.g. EIROforum, ESFRI, EU Flagship Projects, EU Centres of Excellence, evolving multi-national science initiative from several member states?</i></p> <p data-bbox="826 1041 1402 1265">CERN is a member of the EIROforum. The HL-LHC is on the ESFRI 2016 roadmap CERN is funded by 21 European Member states (including Israel) and is thus strongly linked to national research organisations in these member states and beyond.</p>
Broader Impact	<p data-bbox="826 1265 1402 1366"><i>How is this data challenge similar to others in your own or other research communities? By solving this problem what can others learn?</i></p> <p data-bbox="826 1366 1402 1512">Many communities are involved in activities involving long- term data preservation (a pre-requisite for sharing and re- use) but few have the same or similar challenges of scale.</p> <p data-bbox="826 1512 1402 1758">It has been shown, however, that many of the issues involved are discipline-independent and that much can be gained by working together. Open Data is a non-trivial issue, involving either standard formats and/or sufficient documentation and software to use the data in meaningful fashions.</p>