

WLCG Draft Data Management Plan – Annex 2

This draft data management plan has been prepared according to the H2020 (and other) guidelines for input to the DPHEP Workshop in Lisbon in February 2016: <https://indico.cern.ch/event/444264/>.

H2020 Annex 2 Guidelines		
Guideline	Guidance	
Discoverable	<i>Are the data and associated software produced and/or used in the project discoverable (and readily located), identifiable by means of a standard identification mechanism (e.g. Digital Object Identifier)?</i>	The released data from the LHC experiments are discoverable and readily locatable through the CERN open data portal (http://opendata.cern.ch/). This also contains the necessary software and documentation and/or virtual machine for running the software and accessing the data.
Accessible	<i>Are the data and associated software produced and/or used in the project accessible and in what modalities, scope, licenses (e.g. licencing framework for research and education, embargo periods, commercial exploitation, etc.)?</i>	See above. Embargo periods for the 4 main LHC experiments and access conditions are explained in http://opendata.cern.ch/about (see “Data and Re-use” and “Data Policies”).
Assessable and intelligible	<i>Are the data and associated software produced and/or used in the project assessable for and intelligible to third parties in contexts such as scientific scrutiny and peer review (e.g. are the minimal datasets handled together</i>	The medium to long-term goal of the 4 LHC experiments is to permit verification of results by suitably qualified personnel (including peer review). This is encapsulated in the following statement ¹ , originally from the ALICE experiment but subsequently endorsed by the Data Preservation coordinators of the others: “The goal is to require the

¹ Does this document have a DOI?

	<i>with scientific papers for the purpose of peer review, are data is provided in a way that judgments can be made about their reliability and the competence of those who created them)?</i>	<i>reproducibility of analysis in such virtualized environment as a prerequisite for publishing results.”</i> All publications from all HEP experiments undergo comprehensive internal reviews before being submitted to external review.
Useable beyond the original purpose for which it was collected	<i>Are the data and associated software produced and/or used in the project useable by third parties even long time after the collection of the data (e.g. is the data safely stored in certified repositories for long term preservation and curation; is it stored together with the minimum software, metadata and documentation to make it useful; is the data useful for the wider public needs and usable for the likely purposes of non-specialists)?</i>	The data is preserved – together with all necessary software, documentation, metadata and so forth – to permit re-use for well-defined Use Cases ² that are common across the 4 main LHC experiments, as well as other High Energy Physics experiments worldwide. The LHC data is stored in (self) certified digital repositories, namely the Tier0 and Tier1 centres of the Worldwide LHC Computing Grid (WLCG). Re-use by non-specialists is one of the targets of the CERN Open Data portal (although this requires considerable work by experts to prepare the data / documentation / software for such re-use).
Interoperable to specific quality standards	<i>Are the data and associated software produced and/or used in the project interoperable allowing data exchange between researchers, institutions, organisations, countries, etc. (e.g. adhering to standards for data</i>	42

² Maybe we need to document these with a DOI.

	<i>annotation, data exchange, compliant with available software applications, and allowing re-combinations with different datasets from different origins)?</i>	
--	---	--