CERN Services for Long Term Data Preservation

Frank Berghaus, Jakob Blomer,
Germán Cancio Melia,
Sünje Dallmeier Tiessen,
Gerardo Ganis,
Jamie Shiers, Tibor Simko
CERN
1211 Geneva 23, Switzerland
+41 22 76 76111
{ Frank.Berghaus, Jakob.Blomer,
German.Cancio.Melia,
Sunje.Dallmeier.Tiessen,
Gerardo.Ganis, Jamie.Shiers,
Tibor.Simko }@cern.ch

ABSTRACT

In this paper we describe the services that are offered by CERN [3] for Long Term preservation of High Energy Physics (HEP) data, with the Large Hadron Collider (LHC) as a key use case.

Data preservation is a strategic goal for European High Energy Physics (HEP) [9], as well as for the HEP community worldwide and we position our work in this global content. Specifically, we target the preservation of the scientific data, together with the software, documentation and computing environment needed to process, (re-)analyse or otherwise (re-)use the data. The target data volumes range from hundreds of petabytes (PB – 10^{15} bytes) to hundreds of exabytes (EB – 10^{18} bytes) for a target duration of several decades.

The Use Cases driving data preservation are presented together with metrics that allow us to measure how close we are to meeting our goals, including the possibility for formal certification for at least part of this work. Almost all of the services that we describe are fully generic – the exception being Analysis Preservation that has some domain-specific aspects (where the basic technology could nonetheless be adapted).

Keywords

Data reuse; preservation; reproducibility; research data management; digital curation; virtualization; digital libraries; massive storage; open access; open data.

1. INTRODUCTION

CERN, the European Centre for Nuclear Research, is situated outside Geneva with much of its facilities spreading into neighbouring France. It has existed for over 60 years and has 21 member states with several more in various preparatory stages of membership. CERN performs research into the nature and structure of the Universe – the fundamental particles, e.g. the constituents of the constituents of the constituents of atoms[1], and the forces that act between them.

CERN has a diverse research programme based on a wide range of particle accelerators. The largest and most powerful of these is the LHC that entered production in late 2009 after many years of preparation. The LHC occupies a circular tunnel some 100m underground and 27km in circumference. The tunnel previously housed the Large Electron Positron collider (LEP) that operated from 1989 – 2000 and the data from LEP are still available and actively used. Although the total data volume from LEP is "only" around 500TB, this was also "Big Data" in its day. Also, lessons learned from LEP – where we expect the data to be usable up to around 2030 – point the way to what we can expect to achieve for the LHC.

Unlike many disciplines that make *observations* – by definition unrepeatable – HEP makes *measurements*[2]. It would, for example, be technically possible to build a new LEP machine and data from such a machine could make the existing data entirely redundant. This is a simplifying characteristic that is not shared, for example, by gravitational wave detectors, space telescopes and earth observing systems: there, if an event is missed or poorly measured, it can never be measured again – the opportunity has been lost for eternity.

The four main detectors at the LHC, named after the corresponding worldwide scientific collaborations with up to several thousand members each, take several tens of PB of data per year of LHC operation, even after highly de-selective triggers [21]. As the machine is progressively upgraded, these annual data volumes will increase giving a total data sample between 10 and 100EB by the end of active operation, around 2035 – 2040 according to current plans. The re-use of this data within the collaborations during this period is fundamental, including the ability to reproduce past analyses. Recently, all four of the main LHC collaborations (ALICE, ATLAS, CMS and LHCb) have agreed open access policies [27] whereby significant sub-sets of the data are released to other scientists, as well as the general public, after embargo periods. Amongst other things, this period allows the data to be fully prepared for public consumption, along with the corresponding documentation and computing environment – to do this in pseudo real-time would be impossible with the resources that are available.

[1] Electrons are believed to be fundamental, whereas nuclei are not, nor are their components.

[2] A simple analogy might be taking a photograph versus flipping a coin. If you take "the same" photograph tomorrow, or in a year's time, you are in fact recording something different. If you flip a coin ten times today, tomorrow, or in one year (statistically) it makes no difference.

The Computing needs of the LHC experiments are met via the Worldwide LHC Computing Grid (WLCG) [34] that consists of a Tier0 site at CERN (with a remote facility also in Budapest), some ten Tier1 sites elsewhere in Europe, North America and Asia and some hundred Tier2 sites around the world. Whilst a full description of WLCG is outside the scope of this paper it is important to point out that the purpose of the Grid is for rapid processing and analysis of the data by scientists and it is optimized for such. Its use by the general public and / or for educational outreach is not compatible with the way it is designed, resourced or run. This has some implications for the re-use of preserved data and these aspects are described in detail below.

Through a combination of certification for the core preservation services together with additional metrics associated with these key Use Cases we are able to measure how close we are to achieving and sustaining our goals. These Use Cases also form the basis of a Business Model for ensuring sustained funding over a period of (at least) several decades. Finally, we highlight where our experience differs from "conventional wisdom" and in particular tensions between on-going use of preserved data for which it was originally intended versus "Open Access" style usage for educational outreach and other purposes.

1.1 HEP-Wide Data Preservation

In this paper we focus on the data preservation services that are offered by CERN with the LHC experiments as a key Use Case. The current phase of data preservation in HEP was kick-started by a study group that was initiated in late 2008 by DESY [6], resulting in a Blueprint report [23] in May 2012. The study group evolved to include all major HEP laboratories worldwide and reported to the International Committee for Future Accelerators (ICFA) [14], emphasizing its truly global nature.

The Study Group made the following observation:

"Data from high-energy physics (HEP) experiments are collected with significant financial and human effort and are mostly unique. An inter-experimental study group on HEP data preservation and long-term analysis was convened as a panel of the International Committee for Future Accelerators (ICFA). The group was formed by large collider-based experiments and investigated the technical and organisational aspects of HEP data preservation. An intermediate report was released in November 2009 addressing the general issues of data preservation in HEP. This paper includes and extends the intermediate report. It provides an analysis of the research case for data preservation and a detailed description of the various projects at experiment, laboratory and international levels. In addition, the paper provides a concrete proposal for an international organisation in charge of the data management and policies in high-energy physics."

The DPHEP study group identified the following priorities, in order of urgency:

- ***Priority 1: Experiment Level Projects in Data Preservation.*** *Large laboratories should define and establish data preservation projects in order to avoid catastrophic loss of data once major collaborations come to an end. The recent expertise gained during the last three years indicate that an extension of the computing effort within experiments with a person-power of the order of 2-3 FTEs leads to a significant improvement in the ability to move to a long-term data preservation phase. Such initiatives exist already or are being defined in the participating laboratories and are followed attentively by the study group.*
- ***Priority 2: International Organisation DPHEP.*** *The efforts are best exploited by a common organisation at the international level. The installation of this body, to be based on the existing ICFA study group, requires a Project Manager (1 FTE) to be employed as soon as possible. The effort is a joint request of the study group and could be assumed by rotation among the participating laboratories.*
- ***Priority 3: Common R&D projects.*** *Common requirements on data preservation are likely to evolve into inter-experimental R&D projects (three concrete examples are given above, each involving 1-2 dedicated FTE, across several laboratories). The projects will optimise the development effort and have the potential to improve the degree of standardisation in HEP computing in the longer term. Concrete requests will be formulated in common by the experiments to the funding agencies and the activity of these projects will be steered by the DPHEP organisation.*

These priorities could be enacted with a funding model implying synergies from the three regions (Europe, America, Asia) and strong connections with laboratories hosting the data samples.

1.2 Worldwide HEP Data Preservation

Since 2013, the former study group has evolved to a Collaboration and has built partnerships with data preservation efforts and projects in other disciplines. We have benefitted significantly from such partnerships and believe that it is key to offering long-term sustainable services. A Status Report summarizing the progress made since the publication of the Blueprint is available here [22].

The main messages contained in that report are as follows:

- Significant progress has been made in the past years regarding our understanding of, and implementation of services and solutions for, long-term data preservation for future re-use;
- **However, continued investment in data preservation is needed: without this the data will soon become unusable or indeed lost (as history has told us all too many times);**
- **Some of this investment can be done centrally, e.g. by providing bit preservation services for multiple experiments at a given laboratory, whilst important elements need to be addressed on an experiment-by-experiment basis.**
- Funding agencies – and indeed the general public – are now understanding the need for preservation and sharing of "data" (which typically includes significant metadata, software and "knowledge") with requirements on data management plans, preservation of data, reproducibility of results and sharing of data and results becoming increasingly important and in some cases mandatory;
- The "business case" for data preservation in scientific, educational and cultural as well as financial terms is increasingly well understood: funding beyond (or outside) the standard lifetime of projects is required to ensure this preservation;
- A well-established model for data preservation exists – the Open Archival Information System (OAIS). Whilst developed primarily in the Space Data Community, it has since been adopted by all most all disciplines – ranging from Science to Humanities and Digital Cultural Heritage – and

provides useful terminology and guidance that has proven applicable also to HEP;

- **The main message – from Past and Present Circular Colliders to Future ones – is that it is never early to consider data preservation: early planning is likely to result in cost savings that may be significant. Furthermore, resources (and budget) beyond the data-taking lifetime of the projects must be foreseen from the beginning.**

1.3 DPHEP 2020 Vision

The "vision" for DPHEP – first presented to ICFA in February 2013 – a consists of the following key points:

- By 2020, all **archived data** – e.g. that described in DPHEP Blueprint, including LHC data – should be easily **findable** and fully **usable** by the **designated communities** with clear (Open) access policies and possibilities to annotate further
- Best practices, tools and services should be well run-in, fully documented and sustainable; built in common with **other disciplines**, based on standards
- There should be a **DPHEP portal**, through which data / tools may be accessed
- **Clear targets & metrics** to measure the above should be agreed between **Funding Agencies, Service Providers** and the **Experiments (Collaborations).**

Although there is clearly much work still to be done, this vision looks both achievable and the timescale for realizing it has been significantly reduced through interactions with other (non-HEP) projects and communities. This is an important message for other projects and disciplines – collaboration can benefit us all.

2. BUSINESS CASE FOR PRESERVATION

Successful data preservation can only be performed if firstly one understands the motivation for such preservation – who will be the eventual re-users of the data, what is or will be the knowledge base of these re-users and what types of re-use are desired, for example for scientific, educational or simply cultural reasons. Secondly, it is clear that it will require resources and so the potential benefits, ideally in terms of a cost-benefit analysis, are desirable.

Following numerous discussions, a set of common Use Cases has been agreed across the 4 main LHC experiments. With some small provisos, these are also valid for other HEP experiments worldwide.

The basic Use Cases are as follows:

1. Bit preservation as a basic "service" on which higher level components can build;
    - Motivation: Data taken by the experiments should be preserved at least during the lifetime of the experiments and preferably until "redundant".
2. Preserve data, software, and know-how[3] in the collaborations;
    - This is the foundation for the long-term DP strategy
    - Analysis reproducibility: Data preservation alongside software evolution
3. Share data and associated software with (wider) scientific community, such as theorists or physicists not part of the original collaboration
    - This brings additional requirements:
        - Storage for the released data, distributed computing resources to access and process it
        - Accessibility issues, intellectual property
        - Formalising and simplifying data formats and analysis procedures
        - Documentation targeted at the specific consumer communities (lower knowledge base).
4. Open access to reduced data set to general public
    - Education and outreach
    - Continuous effort to provide meaningful examples and demonstrations

2.1 REQUIREMENTS FROM FUNDERS

Increasingly, Funding Agencies (FAs) are requiring Data Management Plans (DMPs) as part of the project approval process. Although these differ in detail from agency to agency, there is nevertheless significant commonality. Using the Guidelines for Data Management Plans in the European Union's Horizon 2020 programme as an example, DMPs should cover, at a minimum, the following:

A DMP describes the data management life cycle for all datasets to be collected, processed or generated by a research project. It must cover:

- *the handling of research data during & after the project*
- *what data will be collected, processed or generated*
- *what methodology & standards will be applied*
- *whether data will be shared / made open access & how*
- *how data will be curated & preserved*

More details are given regarding data sharing and preservation. Furthermore, other Funding Agencies stress reproducibility of results.

For a worldwide project such as the LHC, compliance with the requirements from multiple funding agencies is required. We thus refer to "an intelligent superset" of these requirements that includes not only those from the funders but also those needed internally within the project for its own scientific needs. In fact, we see remarkable synergy between the Use Cases presented above and this superset of requirements.

We believe that the information required to produce and maintain a DMP is typically available within a project. Presenting it in a common format is much more than a "contractual" requirement – it can – and should – be of use within the project, for data sharing and outreach portals, as well as to compare data management and preservation strategies across even heterogeneous communities.

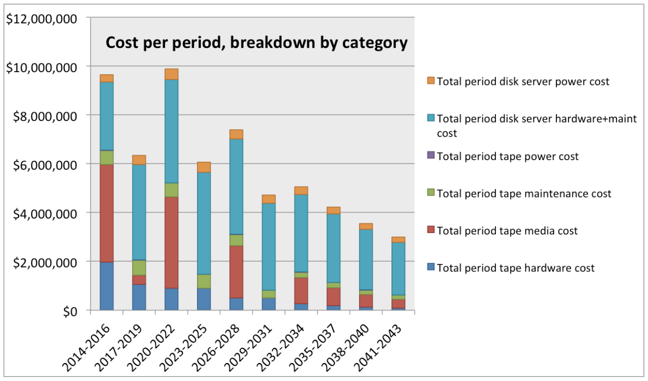
The DMP for the LHC experiments is given further below.

2.2 COST MODELS

In order to estimate the cost for bit preservation over a long period of time we have developed a simple cost model that is freely

[3] Additional Use Cases help to define whether the "know-how" has been adequately captured. See the Analysis Preservation section for further details.

available and can be readily adapted to the parameters of other projects. It is based on publicly available pricing information and technology predictions and – for the LHC – assumes an initial archive size with increasing growth rate, a 10% disk cache in from of the tape (robotic) store and regular (triennial) migration to new denser media. Whereas "conventional wisdom" is that the cost of storage will inevitably spiral out of control, at least for our predicted growth – luckily – this does not appear to be the case.



Here we see that the costs decrease with time despite a significant increase in the total data stored. Of course, it is possible to construct data growths that will exceed the benefits from increase in storage density but at least for one major project this does not seem to be the case.

Naturally, there are large uncertainties in such a model – will tape density continue to increase as foreseen, is there sufficient market demand and many other factors.

These numbers are nonetheless far from negligible – how do they compare with the overall cost of computing and / or the cost of the LHC project as a whole and what are the benefits of this bit preservation?

2.3 PRESERVATION BUSINESS CASE

The cost of bit preservation – certainly only one element of an overall data preservation strategy – can be seen to be small compared to the overall costs of computing for such a project as the LHC and smaller still when compared to the entire project (the collider, the detectors, the collaborations and so forth). On the other hand, the benefits can be measured quasi-directly: in terms of the scientific publications and PhDs that it enables, as well as indirectly in terms of technological, educational and cultural spin-offs. What is observed by all HEP experiments is that analysis of the data, publications and presentations continue many years after the end of data taking and their use for educational outreach continues even further. Again, we are "saved" by the fact that we make measurements and not observations – the duration for which the data (and associated software, metadata, documentation and so forth) should be maintained is perhaps a few decades – for example until a new and more powerful facility is available and first comparisons have been made – and not "forever" as is desirable with observations. (In fact, as the costs for bit preservation tend to zero there is no argument to delete the data even after this period and successful "data resurrection" has been achieved in the past – it requires only sufficient scientific motivation, such as a new or improved theoretical model and/or the discovery of a new particle that should have been visible in the old data).

3. DATA MANAGEMENT AND PRESERVATION PLAN FOR WLCG

Based on the Horizon 2020 guidelines [10], a Data Management Plan[4] for the 4 main LHC experiments – in the context of the Worldwide LHC Computing Grid – has been prepared. The following subsections specify this plan, together with the associated guidelines *quoted verbatim in italics*.

3.1 Data set reference and name

Identifier for the data set to be produced.

This Data Management Plan (DMP) refers to the data set generated by the 4 main experiments (also know as "Collaborations") currently taking data at CERN's Large Hadron Collider (LHC).

These experiments are ALICE, ATLAS, CMS and LHCb. For the purpose of this plan, we refer to this data set as "The LHC Data".

In terms of Data Preservation, the software, its environment and associated documentation must also be preserved (see below).

Further details can be found at the DPHEP portal site, described further below.

3.2 Data set description

Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.

The 4 experiments referenced above have clear scientific goals as described in their Technical Proposals and via their Websites. These are accessible through the official catalogue of all CERN experiments that is maintained by the CERN Research Board, the CERN Grey Book [25].

Hundreds of scientific publications are produced annually.

The data is either collected by the massive detectors of the above experiments (the raw data), is derived from it, or is the result of the simulation of physics processes according to theoretical models and the simulated response of the detector to these models.

Similar data – but at lower energies – have been produced by previous experiments and comparisons of results from past, present and indeed future experiments is routine.

The data behind plots in publications is made available since many decades via an online database, HEPData, described below.

Re-use of the data is made by theorists, by the collaborations themselves, by scientists in the wider context as well as for Education and Outreach.

3.3 Standards and metadata

Reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created.

The 4 main LHC experiments work closely together through the WLCG Collaboration on data management (and other) tools and applications. At least a number of these have found use outside the HEP community but their initial development has largely been

---

4 Much of this work is required as part of the formal approval process of an experiment and / or as part of data sharing and preservation planning.

driven by the scale and timeline of the above. The ROOT framework [20], in particular, is used as "I/O library" (and much more) but all LHC experiments and is a *de-facto* standard within HEP, also across numerous other laboratories.

The meta-data catalogues are typically experiment-specific although globally similar. The "open data release" policies foresee the available of the necessary metadata and other "knowledge" to make the data usable (see below).

3.4 Data sharing

Description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.).

In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

The 4 LHC experiments have policies for making data available, including reasonable embargo periods, together with the provision of the necessary software, documentation and other tools for re-use.

Data releases through the CERN Open Data Portal (see below) are published with accompanying software and documentation. A dedicated education section provides access to tailored datasets for self-supported study or use in classrooms. All materials are shared with Open Science licenses (e.g. CC0 or CC-BY) to enable others to build on the results of these experiments. All materials are also assigned a persistent identifier and come with citation recommendations.

3.5 Archiving and preservation

Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.

The long-term preservation of LHC data is the responsibility of the Tier0 and Tier1 sites that form part of the WLCG Collaboration. A Memorandum of Understanding [33] outlines the responsibilities of sites that form part of this collaboration (Tier0, Tier1s and Tier2s).

In the case of the Tier0 and Tier1s, this includes "curation" of the data with at least two copies of the data maintained worldwide (typically 1 copy at CERN and at least 1 other copy distributed over the Tier1 sites for that experiment).

The costs for data storage and "bit preservation" form part of the resource requests that are made regularly to the funding agencies. A simply cost model shows that the annual storage costs – even including the anticipated growth – go down with time and remain within the funding envelope foreseen. (The integrated costs of course rise).

Personnel from the Tier0 and Tier1 sites have followed training in ISO 16363 certification – A Standard for Trusted Digital Repositories – and self-certification of these sites is underway.

Any data generated on external resources, e.g. Clouds, is copied back for long-term storage to the Tier0 or Tier1 sites. The eventual long-term storage / preservation of data in the Cloud would require not only that such services are cost effective but also that they are certified according to agreed standards, such as ISO 16363.

The data themselves should be preserved for a number of decades – at least during the active data taking and analysis period of the LHC machine and preferably until such a time as a future machine is operational and results from it have been compared with those from the LHC.

The total data volume – currently of the order of 100PB – is expected to eventually reach 10-100 EB (in circa 2035 – 2040).

Additional services are required for the long-term preservation of documentation (digital libraries), the software to process and/or analyse the data, as well as the environment needed to run these software packages.

All such services are the subject of the on-going self-certification.

4. SERVICE PORTFOLIO

In this section we address the high level Use Cases in the order presented, namely:

1. Bit Preservation;
2. Preserving data, software and know-how within the "producer collaborations";
3. Share data and associated software with (larger) scientific community;
4. Open access to reduced data set to general public.

One key concern is the ability to "reproduce" physics analyses published in the past. The scientific data underlying publications by the CERN experiments is complex and requires a rich ecosystem of descriptive information. The data published in scientific papers is produced from derived data using software specific to that analysis. That software, a reference to and provenance of the derived data used, the computing environment, and the analysis documentation is catalogued in the CERN Analysis Preservation Portal. Some derived data is catalogued with the necessary access protocols on the CERN Open Data Portal. The information on the CERN Open Data Portal allows a virtual machine to access data on the CERN bit preservation infrastructure. These issues are discussed in more detail below.

4.1 Bit Preservation

We predict a total data volume of a few exabytes (EB) from the LHC experiments by the 2030s [18] and a final data volume between 10 and 100 EB. In fact, Influx rates from CERN's Large Hadron Collider experiment are expected to augment from currently 40 Petabytes / year to around 600 Petabytes / year in a few years time, therefore reaching archive volumes at the Exabyte-scale. The data from the rest of the CERN experimental programme can be archived easily alongside this massive LHC dataset. At this scale, bit-rot – the tendency of content in storage to become corrupt over time – becomes unavoidable. Reasons for bit rot are multiple, the most important ones being: wear out and breakage of media components (such as disk drive head crashes, snapping tapes, RAID controller failures); undetected bit flips during data transmission; hardware or media obsolescence; or environmental hazards (fire, water leaks, dust contamination).

- In order to minimise data loss and improve archive reliability, CERN has implemented storage verification and preservation services on top of its tape-based Mass Storage System (CASTOR [2]). These measures include notably:
- Regular Media verification: Every time a tape is written to, it will be subject to a verification process that consists in checking critical areas, namely the first and last ten files of

the tape, as well as 10 random files across the tape, and validating the metadata (such as file sizes and checksums). When a tape is filled, all its contents will be verified. In addition, all tapes are re-verified approximately every 2 years, ensuring also the correctness of older repository data.

- Controlled media lifecycle: Media at CERN is typically kept in production for not longer than two drive generations (typically 6-8 years). This is well below the physical media lifetime, which is around 20-30 years. While tape media is well-known for its longevity (30 years or more), the associated hardware infrastructure typically enters obsolescence after 4-6 years, after which it becomes difficult to find replacements for tape drives, firmware patches or software drivers for new operating system versions. In addition, newer media usually comes with increased reliability over older generations. Last but not least, by migrating existing data to newer-generation and higher-capacity media, less cartridges will be required and expenses in additional tape libraries and floor space can be avoided.
- Reducing tape mounts: In order to reduce media wear out and to increase efficiency, a policy-driven engine examines each tape read request and decides on whether to grant a tape mount or postpone it. This takes into account criteria such as user/group priority, number of files and amount of volume to be read, waiting time, and concurrent drive usage by the user/group. Since deployment in 2010, and despite continuous file and volume recall increases, the average number of daily tape read mounts has been reduced from over 4000/day to 1500/day.
- Data redundancy: For smaller communities, such as the former LEP experiments, secondary file copies can be created. These second data copies are stored in a separate library residing in a different physical building.
- Protecting the physical link: In order to increase the reliability of data transfers between the disk caches and tape servers, CERN has implemented support for SCSI Logical Block Protection [19]. This mechanism protects the path between the data source and the tape media (e.g. FC interface and physical link, internal drive data channels, etc.) against errors such as link-level bit flips. It works by pre-calculating and appending a CRC code for each data block sent to the tape drive, which is then re-calculated and verified at every transfer back and forth to tape.
- Protecting the operating environment: Tape media is vulnerable to contamination from airborne dust particles that can land on the rollers, reels or heads. These can cause scratches on the tape as it is being mounted or wound on the tape drive. With tape media bit sizes smaller than a bacterium or the particles emitted by a car exhaust, any damage to the tape can destroy significant amounts of data. CERN has prototyped and built custom environmental sensors that are hosted in the tape libraries, sampling the same airflow as the surrounding drives [8]. The sensor continuously samples the surrounding air and issues alarms if airborne particle density, humidity or temperature crosses configurable thresholds.

These measures have helped reducing the number of annual file losses by two orders of magnitude. For the period 2012-2015, the annual bit loss rate is in the order of $5*10^{-16}$. This rate can still be improved as it is still three to four orders of magnitude above the undetected bit error rate for enterprise-level tape drives, which can be considered as the upper ceiling to reach in terms of reliability.

4.2 Software Preservation

The HEP community has a long tradition of sharing and developing common, open-source software stacks within international collaborations. The software systems required to operate on LHC data comprise physics simulation and reconstruction algorithms to determine physics processes from detector signals, data analysis frameworks to extract (new) scientific knowledge from data sets, and distributed systems for data access and compute job management. Altogether, HEP software stacks add up to tens of millions of lines of code, half a dozen different languages and tens to hundreds of modules with dependencies on each other. Several millions of lines of code are specific to an experiment and there are numerous dependencies on standard software, most notably the GNU/Linux operating system and language compilers and interpreters. The support life cycle of the 3rd party software components is much shorter than the envisaged preservation period of several decades. Operating system versions are supported for a maximum of 5-10 years, for instance, and most developers abandon their software releases much earlier.

Running experiments invest many person months in the porting and the validation of their software stacks, coordinated by a dedicated software librarian. For a decommissioned experiment, that is one that is no longer in its data-taking phase, such an amount of effort would be impractical. Hardware virtualization (such as KVM [17], Xen [35] and VirtualBox [36]) and container virtualization (such as Docker [7]) provide a potential solution. Virtualization allows for execution of a frozen, historic software environment on contemporary hardware and operating systems. In a straightforward application of virtualization technology, a software environment is frozen in the form of a disk image, a large and opaque stream of bytes containing all the necessary software binaries. This approach tends to be clumsy and too rigid for HEP software. In order to be useful, even “frozen” environments need to stay open for minor modifications: patches to faulty algorithms, updates to physics models, updated tuning parameters for simulation algorithms, new configuration for data access software and so on. Software development communities have long solved similar problems by version control systems. Version control systems only store a track of changes to the source code and at the same time they can provide access to the state of a directory tree at any given point in the past.

In an attempt to get similar benefits for compiled and configured software stacks, since several years HEP experiments install all released software components in its final configuration on the versioning, open source, and distributed file system CernVM-FS [5]. By selecting different versions in the history provided by the file system, experiments can access any software state ever released. Thus we can separate virtualization – in this case handled by CernVM [26] – itself from the concerns of accessing software binaries. A minimal and stable virtual machine or container (~20MB) connects to the remote file system CernVM-FS that hosts the operating system and software binaries. By selecting different states of the versioned file system, experiments can go back and forth in time and create software environments compatible with Red Hat Enterprise Linux (RHEL) 4 to RHEL 7 (spanning 15+ years) with the very same virtual machine on the very same hardware. Concretely, we have demonstrated that by resurrecting the software of the ALEPH experiment at LEP more than 15 years[5] after the experiment was decommissioned. Contemporary virtual machines provide data access tools and middleware with support for the latest network protocols and security settings. Containers inside the virtual machines spawn historic operating system and application software environments. Data is provided from the container host to the historic applications through the very stable POSIX file system interface.

Among the currently active HEP experiments, many operate dedicated CernVM-FS services to satisfy their day-to-day needs for global software distribution. These services are operated in an "append-only" mode, so that software versions, once released to the production infrastructure, remain readily available for future use. Due to the file system's internal data de-duplication, this model proved to be sustainable even for the largest users. After more than five years of experience with LHC experiment software and more than hundred million registered files (software only!), the storage volume is still at only a few terabytes.

4.3 HEPData

HEPData [11] did not originate as a CERN service but deserves a specific mention as it has provided access to data behind physics publications for several decades. Founded and run by Durham University, it "*has been built up over the past four decades as a unique open-access repository for scattering data from experimental particle physics. It currently comprises the data points from plots and tables related to several thousand publications including those from the Large Hadron Collider (LHC)".*

Thus it is complementary to the portals offered by CERN and a transition will soon be made to a new HEPData site, hepdata.net, based on Invenio [15], developed in collaboration with INSPIRE [13].

4.4 Analysis Preservation Portal

Research outputs in physics range from data, software and documentation to the "traditional" publication – while so far only the latter is preserved and published openly. Currently, the user-generated content is scattered around various tools and services within the individual experiments and it is difficult to relate the individual elements to a specific analysis and research result. To enable others to build on collaborators' findings and to foster reproducible research it is important to preserve and provide (internal) access to the wider range of related and supplementary materials.

Hence, a new service is developed in close collaboration with the research community, i.e. the LHC collaborations (ALICE, ATLAS, CMS, LHCb). "CERN Analysis Preservation" is under development to capture research objects while researchers are in the process of conducting their research. The new service is being built on the latest release of Invenio, an open source digital library software developed at CERN and addresses the complex but essential Use Cases for Analysis Preservation [1]. For example:

- An analysis that is underway has to be handed over, e.g. as someone is leaving the collaboration;
- A previous analysis has to be repeated;
- Data from different experiments have to be combined.

The service aims at preserving the insider knowledge about a physics analysis. The researchers provide metadata about the data, software, configurations options, high-level physics information, documentation, instructions, links to presentations, quality protocols, internal notes, etc. The system connects to selected databases in the experiments so that information can be retrieved automatically, therefore information is up to date and researchers do not have to spend much extra effort on using this service. The service also facilitates versioning to accommodate the work in progress nature of the materials.

Once the service reaches production level, it will allow users of the collaboration for the first time to search for related user generated content they might be interested in. Furthermore, it is expected to allow internal analysis tools to plug into this new service to make use of the central information resource.

It should be noted that CERN Analysis Preservation is a "closed access" service, as it deals with the early stages of the data analysis process before the results are submitted for publications. The access is thus restricted to the individual LHC collaboration. After an analysis is approved for publication the CERN Analysis Preservation service may (upon request by the researcher) push parts of information to the public CERN Open Data portal and the INSPIRE services. Hence, in combination with the other preservation services at CERN, CERN Analysis Preservation should help fostering preservation and Open Science practices in the community.

4.5 Open Data Portal

Corresponding to the LHC data policies [27], a service was needed to serve large scale and complex datasets, together with underlying virtual analysis environment, an example software code, and supporting materials. Existing services, such as WLCG were not (by construction) suited to accommodate the needs for the sharing of complex and big datasets. Hence, the public CERN Open Data Portal [4] was launched in November 2014, providing data and accompanying software and tools for education and research purposes to the public. To give an example: the annual CMS data release of data from 2010 focused on primary and derived data, which amount to a volume of 27TB; the 2011 release comprised simulated data, detailed information about configuration and triggers and, hence, resulted in several hundred terabytes. All LHC collaborations have already shared data through this service.

Special emphasis was given on providing comprehensive metadata for the objects that are shared and on an appealing user interface. To serve the target groups, the physicists and the non-physicists, best the repository was presented with a modern website layout. A close collaboration of CERN IT, the Scientific Information Service, and the physics collaborations ensured that sufficient (interactive) tools, accompanying documentation and metadata are provided in an understandable way to facilitate future reuse. Following best practices, materials shared through this services are issued a persistent identifier so that reuse (i.e. citations) can be tracked on INSPIRE.

CERN Open Data Portal and INSPIRE are based on the Invenio digital library software.

4.6 DPHEP Portal

The portal of the DPHEP Collaboration [29] provides a human readable reference of sites and services used by the worldwide effort to preserve HEP data. The aim is to make the HEP data findable, accessible, interoperable, and re-usable [30]. The

[5] Data continue to have scientific value and underpin publications and even PhDs long after data taking has ended. Porting and validating the associated software as well as handling changes in data formats is a story in itself and outside the scope of this paper.

landing page lists the member institutes with their preferred data preservation portal. Members of the collaboration provide detailed information on current status of their data regarding: bit preservation, data, documentation, software, use cases, target audience, value, uniqueness, resources, issues, and outlook. The portal makes the applicable preservation and access policies available and provides relevant contact details. Agreeing on a common way of presenting the status at the different sites and laboratories took several years of elapsed time and helps to highlight commonalities and areas for shared developments, services and / or tools.

Furthermore, the portal provides a central reference point to meetings, documents, and services organizing the HEP data preservation effort.

Institutes / organisations referenced from the portal include Brookhaven National Laboratory, Fermi National Laboratory and Stanford Linear Accelerator Laboratory, all in the US; CERN, CSC, DESY, INFN, IN2P3 and STFC in Europe as well as IPP in Canada, IHEP in China and KEK in Japan.

5. CERTIFICATION OF REPOSITORIES

It is widely accepted that certification of repositories is at least a best practice as part of a long-term data preservation strategy. Whilst there are a number of certification frameworks in use [12], that covered by ISO 16363 [16], based on the OAIS reference model [31], is considered to be the most comprehensive and even ambitious. Moreover, it matches our current practices more closely than other such frameworks.

In the context of the LHC experiments, the “repository” for long-term storage consists of the WLCG Tier0 site (CERN) plus the Tier1 sites spread around the world. A copy of all “archive” data is maintained at CERN with at least one additional copy being spread over the Tier1 sites that serve that experiment.

Representatives of these sites have undergone training in the application of ISO 16363 and self-certification is underway with a goal of covering at least the WLCG Tier0 prior to iPRES 2016. This would be a first step – formalizing some of the issues in terms of official policies will not be achieved on this time frame. Similarly, including all of the Tier1 sites will take some additional time, as would extending this certification to cover all experiments whose data is archived at CERN and/or all projects supported by the Tier1s [37], many of which are multi-disciplinary.

We believe that this will greatly enhance the transparency and long-term sustainability of our overall long-term data preservation strategy. In particular, it will formalize and institutionalize many of the practices, services and strategies described in this paper. Furthermore, any gaps identified will help us improve our strategy for the future.

6. CONCLUSIONS

We have presented the services offered by CERN for the long-term preservation of the data from the LHC experiments, along with a business case and a draft Data Management Plan. Key to this DMP is the on-going self-certification of the archive sites according to the ISO 16363 standard – the most rigorous of the various certification standards currently available. Goals for data sharing and reproducibility of results have been shown and by constantly monitoring these we are able to measure we are meeting are targets. Whilst ISO 16363 is discipline agnostic, at least some details of our requirements and practices for sharing and reproducibility require domain-specific knowledge. Furthermore, our experience with data sharing and re-use is still relatively young: as more data is released, crossing the PB threshold and well beyond, new issues will arise and fresh lessons will be learned. However, we strongly believe that the more these issues are addressed in common the more everyone benefits. We have also shown the need for separate (but linked) infrastructures for different purposes: WLCG provides the main processing, analysis and archival facilities whilst the Portals perform tasks related to reproducibility, data sharing and outreach. Such a separation is particularly important during the active data taking stage of an experiment and may become less critical with time but we believe that it should not be overlooked. Finally, our cost model shows that, at least for bit preservation, and assuming no major technological surprises, our overall budget is sustainable in both medium and long term.

7. ACKNOWLEDGMENTS

Although this paper focuses on the services offered by CERN, the work is not done in isolation and owes much not only to the DPHEP Collaboration [28] but also to many others active in the data preservation world, including members of the Alliance for Permanent Access (APA) [24], the Research Data Alliance (RDA) [32] and its many Working and Interest Groups as well as numerous data preservation projects, workshops and conferences worldwide.

Furthermore, it is the experiments themselves that not only drive the requirements described above but also need to do a significant amount of work to prepare, clean, document and share their data as well as support access to it – these services are for them, as well as potential re-users of the data.

8. REFERENCES

[1] CAP Use Cases: http://dx.doi.org/10.5281/zenodo.33693.

[2] CASTOR homepage http://cern.ch/castor.

[3] CERN – the European Organisation for Nuclear Research – see http://home.cern/.

[4] CERN Open Data Portal – see http://opendata.cern.ch/.

[5] CernVM-FS – see https://cernvm.cern.ch/portal/filesystem.

[6] Data Preservation in High Energy Physics (DPHEP): see http://www.dphep.org/.

[7] Docker – see https://www.docker.com/.

[8] Dust sensors: Data Centre Environmental Sensor http://www.ohwr.org/projects/dces-dtrhf-ser1ch-v1/wiki.

[9] European Strategy for Particle Physics – see https://indico.cern.ch/event/244974/page/1736-welcome.

[10] Guidelines for Data Management Plans in Horizon 2020: see http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

[11] HEPData – see http://hepdata.cedar.ac.uk/; https://hepdata.net/.

[12] http://www.trusteddigitalrepository.eu/Memorandum%20of%20Understanding.html.

[13] INSPIRE: the HEP information system – see http://inspirehep.net/.

[14] International Committee for Future Accelerator (ICFA) – see http://icfa.fnal.gov/.

[15] Invenio – see http://invenio-software.org/.

[16] ISO 16363: Audit and Certification of trustworthy digital repositories – see http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510.

[17] KVM – a Kernel-based Virtual Machine – see http://www.linux-kvm.org/page/Main_Page.

[18] LHC data volume predictions: Bird I et al. 2014 Update of the Computing Models of the WLCG and the LHC Experiments *CERN-LHCC-2014-014*.

[19] Logical block protection: Butt K 2007 Tape end-to-end data protection proposal http://www.t10.org/ftp/t10/document.07/07-374r0.pdf.

[20] ROOT – a modular scientific software framework – see https://root.cern.ch/.

[21] See http://home.cern/about/computing/processing-what-record for further details.

[22] Status Report of the DPHEP Collaboration, February 2016: https://zenodo.org/record/46158#.Vvo8zBL5jBI.

[23] Status Report of the DPHEP Study Group: Towards a Global Effort for Sustainable Data Preservation in High Energy Physics: http://arxiv.org/pdf/1205.4667.

[24] The Alliance for Permanent Access (APA) – see http://www.alliancepermanentaccess.org/.

[25] The CERN database of experiments – “the Grey Book”: see https://greybook.cern.ch/greybook/.

[26] The CernVM Software Appliance – see https://cernvm.cern.ch/portal/filesystem.

[27] The data (access) policies of the four LHC experiments are available here: http://opendata.cern.ch/collection/Data-Policies.

[28] The DPHEP Collaboration – see https://hep-project-dphep-portal.web.cern.ch/.

[29] The DPHEP Portal http://dphep.web.cern.ch.

[30] The FAIRport initiative http://www.datafairport.org.

[31] The Open archival information system (OAIS) -- Reference model – see http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284.

[32] The Research Data Alliance (RDA) – see https://rd-alliance.org/node

[33] The WLCG Memorandum of Understanding – see http://wlcg.web.cern.ch/collaboration/mou.

[34] The Worldwide LHC Computing Grid (WLCG) – see http://wlcg.web.cern.ch/.

[35] The Xen project – see http://www.xenproject.org/.

[36] VirtualBox – see https://www.virtualbox.org/.

[37] WLCG Tier1 sites: see http://wlcg-public.web.cern.ch/tier-centres.