

PAPER • OPEN ACCESS

DIRAC in Large Particle Physics Experiments

To cite this article: F Stagni *et al* 2017 *J. Phys.: Conf. Ser.* **898** 092020

View the [article online](#) for updates and enhancements.

Related content

- [Belle II distributing computing](#)
P Krokovny
- [Computing at the Belle II experiment](#)
Takanori HARA and Belle II computing group
- [Belle II Software](#)
T Kuhr, M Ritter and Belle II Software Group

DIRAC in Large Particle Physics Experiments

F Stagni¹, A Tsaregorodtsev², L Arrabito³, A Sailer¹, T Hara⁴, X Zhang⁵

¹EP Department, CERN, CH-1211 Geneva 23 Switzerland

²Centre de Physique des Particules de Marseille, 163 Av de Luminy Case 902, CNRS/IN2P3, 13288 Marseille, France

³Laboratoire Univers et Particules, Université de Montpellier Place Eugne Bataillon - CC 72, CNRS/IN2P3, F-34095 Montpellier, France

⁴High Energy Accelerator Research Organization, 1-1, Oho, Tsukuba, Japan

⁵Institute of High Energy Physics, 19B Yuquan Road, Beijing 100049, P.R. China

On behalf of the DIRAC consortium

E-mail: federico.stagni@cern.ch

Abstract. The DIRAC project is developing interware to build and operate distributed computing systems. It provides a development framework and a rich set of services for both Workload and Data Management tasks of large scientific communities. A number of High Energy Physics and Astrophysics collaborations have adopted DIRAC as the base for their computing models. DIRAC was initially developed for the LHCb experiment at LHC, CERN. Later, the Belle II, BES III and CTA experiments as well as the linear collider detector collaborations started using DIRAC for their computing systems.

Some of the experiments built their DIRAC-based systems from scratch, others migrated from previous solutions, ad-hoc or based on different middlewares. Adaptation of DIRAC for a particular experiment was enabled through the creation of extensions to meet their specific requirements. Each experiment has a heterogeneous set of computing and storage resources at their disposal that were aggregated through DIRAC into a coherent pool. Users from different experiments can interact with the system in different ways depending on their specific tasks, expertise level and previous experience using command line tools, python APIs or Web Portals. In this contribution we will summarize the experience of using DIRAC in particle physics collaborations. The problems of migration to DIRAC from previous systems and their solutions will be presented. An overview of specific DIRAC extensions will be given. We hope that this review will be useful for experiments considering an update, or for those designing their computing models.

1. Introduction

The DIRAC project[1] enables communities to interact with distributed computing resources. It forms a layer that hides diversities across computing, storage, catalog, and queuing resources. DIRAC has been adopted by several HEP and non-HEP experiments' communities, with different goals, intents, resources and workflows.

DIRAC is a truly open source project: it has been started as an LHCb project, and open sourced in 2009 following interest of adoption from another community. Now, it is hosted



on GitHub¹ and is released under the GPLv3 license. There is no dedicated funding for its development: rather, communities using it are welcome to participate in its development.

DIRAC is publicly documented, with an active assistance forum animated by the users and developers themselves. A yearly user workshop and weekly open developers meetings gather users and experts. The project counts about five core developers, and a dozen contributing developers.

The DIRAC consortium has been established in 2014 as a representing body for the development and maintenance of the DIRAC software. The consortium counts a small set of active members, each of which elects a representative, while consortium members elect a Director, and a technical Director. Institutes that are part of the consortium engage in the maintenance and in the promotion of the DIRAC software.

The aim of this paper is to show how DIRAC can satisfy the requirements of different VOs, and which technical solutions have been adopted. This paper is organized as following: in section 2 we will give details on the DIRAC software organization. In section 3 some DIRAC users' experiences will be summarized. Conclusions are given in section 4.

2. Accommodating requirements

Communities that have chosen DIRAC as the distributed system of choice vary greatly: some of them have hundreds of users, others only a few. A community using DIRAC may run concurrently several tens of thousands of "jobs" on grids and clouds, and store and analyze petabytes of data on remote storage; others only run few hundreds of jobs per week, and barely need to interact with storage systems. Requirements vary, and DIRAC should be able to accommodate most of them.

A single DIRAC installation can serve one VO, or more than one. Few DIRAC installations can serve dozens of Virtual Organizations (VOs), or even single users. DIRAC needs to be flexible in accommodating the different requirements, while at the same time providing scalability, and a software easy to extend.

In order to accommodate different requirements, DIRAC has been developed with extensibility in mind. We recognize two types of extensibility, that we dub "horizontal" and "vertical". In Figure 1 we show a diagram that explains the so-called "horizontal extensibility": the core project ("DIRAC"), can be extended for accommodating different requirements.

Each of the projects in Figure 1 have been created with a specific goal and are independently versioned. At the same time a DIRAC release is composed by all those projects, so there is a strong dependency between them. All the projects in Figure 1 are hosted together on GitHub, share the same license and are maintained by the same set of users.

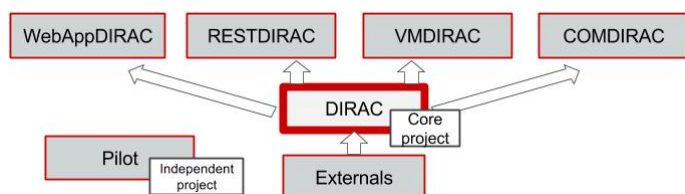


Figure 1. Horizontal Extensibility, for accommodating different requirements.

The "DIRAC" software is the "Core project", so it is the glue that keeps the satellite projects together. It depends on software not maintained by the DIRAC consortium, which we collect in "Externals". The code for the DIRAC web portal sits in "WebAppDIRAC". "RESTDIRAC" extends some DIRAC services by providing a REST (Representational State Transfer) interface. With "VMDIRAC" it is possible to create Virtual Machines on several clouds. "COMDIRAC" extends the DIRAC user interface. The "Pilot" is instead independent from all the other code.

¹ <https://github.com/DIRACGrid>

Figure 2 shows a graphical representation of vertical extensibility, that enables DIRAC users and VOs to extend the functionalities of the "vanilla" projects. In other words, the projects of Figure 1 can be extended by the communities in order to provide specific functionalities.

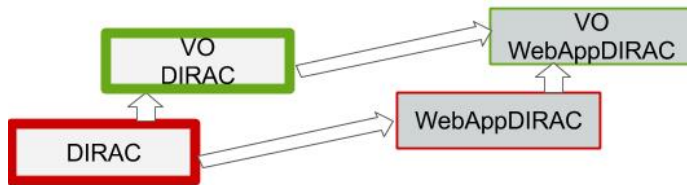


Figure 2. Vertical Extensibility, for accommodating requirements specific to a VO.

Coding for DIRAC also requires developing with flexibility in mind. There are several examples of DIRAC's flexibility, and one example regards DIRAC not imposing any computing model on its users. So for example "Tier" levels may not mean anything to certain VOs. DIRAC provides users the possibility for a "full mesh" computing model, where every job can, in principle, run everywhere. At the same time, a "real world" computing model is represented by putting limits on a full mesh, and this is simply a configuration parameter.

3. DIRAC Users

In this section some of the DIRAC users' experience is presented. There are dozens of communities, large and small, that use DIRAC for their distributed computing activities and workflows, but in the following sections we will take a closer look at some of them only, chosen among those which have the highest complexities.

3.1. LHCb

The LHCb collaboration² has been using DIRAC before it became an open source project: indeed LHCb is DIRAC's first developer, and as of today the collaboration is still its main maintainer. LHCb uses DIRAC for almost all its distributed computing needs, using a variety of computing resources including Grid resources (computing elements CREAM, ARC, HTCondorCE), private clusters (via SSH/GSISsh tunnels), batch systems (LSF, SGE, Condor, Torque), clouds[2] (where LHCbDIRAC pilots are spawned by vcycle), Vac[3], and opportunistic resources (BOINC[4], the LHCb HLT farm[5], a few HPC centers).

The high flexibility granted by DIRAC pilots[6] enabled LHCb to easily extend them in LHCbDIRAC pilots. LHCbDIRAC pilots create a suitable environment for running LHCb jobs everywhere, provided that CVMFS[7] is available (LHCb applications are deployed via CVMFS).

LHCb can interact, via DIRAC, with storage systems using protocols like SRM (gfal, gfal2) and DIP (the Internal DIRAC Protocol), while experimenting xRootD. LHCb uses DIRAC to submit and monitor transfer requests to FTS3[8]. The DFC[9] (DIRAC File Catalog) has been the replica catalog of choice for LHCb for the last 3 years, (previously, LHCb used the LCG File Catalog). LHCb also extended DIRAC for interacting with the LHCbDIRAC Bookkeeping[10], which is LHCb's data provenance and metadata catalog.

The DIRAC web portal is widely used for monitoring and management purposes, and partially extended. End Users often interface with LHCbDIRAC with Ganga[11].

Figure 3 shows a monitoring plot of concurrently running jobs on LHCb resources during December 2016 and January 2017, which also shows that a DIRAC installation can sustain more than a hundred thousands concurrently running jobs.

² <http://lhcb-public.web.cern.ch/lhcb-public/>

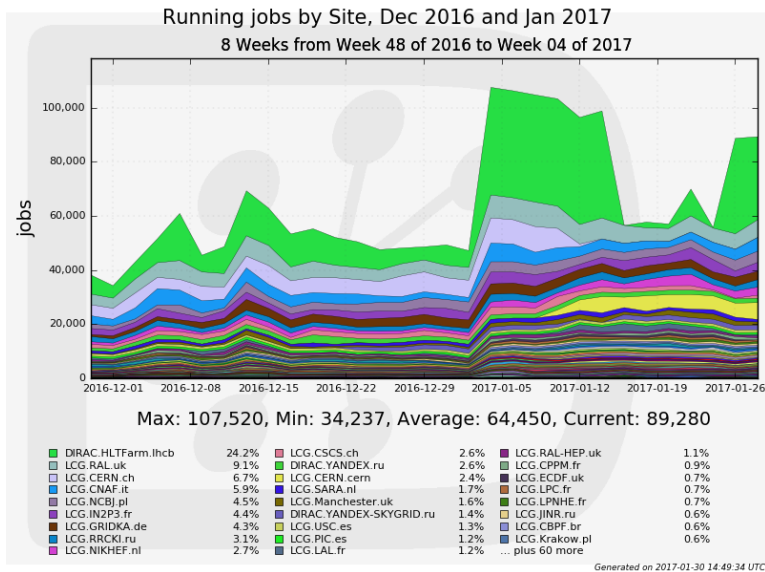


Figure 3. Concurrently running jobs in LHCb (2 months)

3.2. Belle II

Belle II is a super B flavor factory experiment at the asymmetric energy e^+e^- collider in Japan. The first physics run will take place in 2018, then the instantaneous luminosity will be increased up to the design value, which corresponds to 40 times higher than the predecessor experiment, Belle[12]. The Belle II computing system[13] therefore has to manage the process of massive RAW data, the production of copious simulation as well as many concurrent user analysis jobs. Belle II is a worldwide collaboration of about 700 scientists working in 23 countries, so naturally adopted a distributed computing model based on existing technologies. DIRAC was chosen as a workload and data management system and AMGA[14] is used as metadata service. For the file replica catalog, Belle II chose the LFC (LCG File Catalog)[15]. DIRAC provides the necessary interoperability of heterogeneous computing system such as grids with different middleware, academic/commercial clouds and local computing clusters.

DIRAC supports the functionality extension so that experiments can integrate necessary features by themselves. Belle II also developed an extension module, so-called BelleDIRAC, which provides not only the API but also basic tools including remote job submission and data management for the Belle II analysis software framework. On top of that, an automated production system based on DIRAC Transformation System was introduced.

In order to evaluate the computing model, Belle II used DIRAC for running Monte Carlo (MC) mass production campaigns. In particular, since the automated production system was introduced (from MC4 in Figure 4, which shows the history of MC production campaigns), the production conditions drastically improved. We recorded the highest number of concurrently running jobs (more than 20k) and the longest period of smooth operation. Throughout these campaigns more than 2PB data were transferred to the destination Storage Element. The production system is still being improved towards the coming data taking.

3.3. CTA

CTA[16] (Cherenkov Telescope Array) is the next-generation instrument in the field of very high gamma-ray astronomy. First telescope data are expected in 2018, while science operations will start in 2020 for a duration of about 30 years. In total, CTA will collect and process about 27 PB/year. Moreover, large computing and storage resources have been required during the preparatory and pre-construction phases (2010–2017), to perform several Monte Carlo studies. In order to cope with these requirements, during the past four years CTA used resources of the

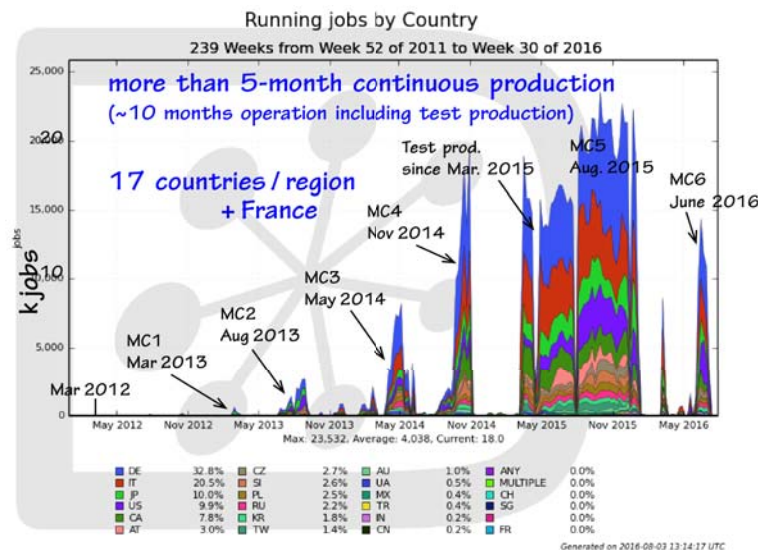


Figure 4. History of the MC production campaigns with Belle DIRAC

EGI grid. In order to efficiently manage large productions in a distributed environment, in 2011 CTA started to evaluate DIRAC [17]. CTA was not relying on any existing production system or particular service, so that no adaptations or extensions have been necessary.

In 2013, the CTA DIRAC instance was put in production and it was first employed for Monte Carlo studies dedicated to the selection of the two sites, hosting the CTA telescopes. Once the two sites have been selected, additional Monte Carlo simulations have been performed, for optimizing the telescope array layouts. This last Monte Carlo campaign (2015–2016) consisted in several thousands of jobs running over periods of 4–5 months, at about 10 different grid sites. About 11 PB have been transferred between the processing sites and the storage locations (see Figure 5). Finally, about 2 PB have been permanently stored.

Several DIRAC components have been employed to handle these productions. In particular, the DFC is used as replica and meta-data catalog. Today it contains more than 25 M replicas. For the production activities, CTA relies on the DIRAC Transformation System[18]. More recently, CTA also started to evaluate VMDIRAC[19] to access private and public cloud resources. Moreover, in order to facilitate end-users to run their specific analysis on the grid, CTA has developed a minimal CTADIRAC extension. This extension consists essentially of an API interface to configure and submit typical CTA workflows. Finally, CTA future plans consist in further developing the current CTA DIRAC instance, in view of the upcoming telescope data processing. Of particular importance, will be the integration of the DIRAC components with some external services, which may be adopted within CTA (e.g., science gateways, Authentication and Authorization systems).

3.4. The Linear Collider Community

The Linear Collider community consists of several collaborations – CLIC Detector and Physics study (CLICdp³), International Large Detector (ILD⁴), Silicon Detector (SiD⁵), working on detector optimisation and physics studies for the Compact Linear Collider (CLIC) or the International Linear Collider (ILC).

The iLCDirac extension of DIRAC [20] was created to support the linear collider software, and to customize the DIRAC Transformation System. The extension contains *Workflow Modules*

³ <http://cllcdp.web.cern.ch>

⁴ <http://www.ilcild.org>

⁵ <http://silicondetector.org>

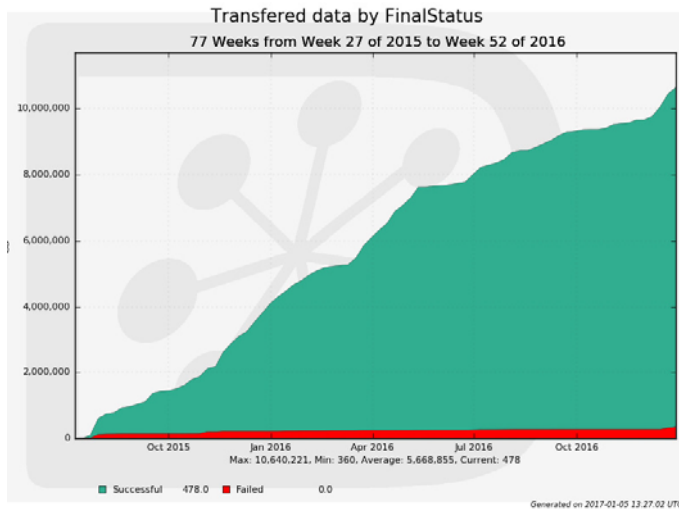


Figure 5. Cumulative data transferred during the MC campaign for the array layout optimization. In total, about 11 PB have been transferred. They include all the transfers between the processing sites and the storage locations, both for production and analysis jobs.

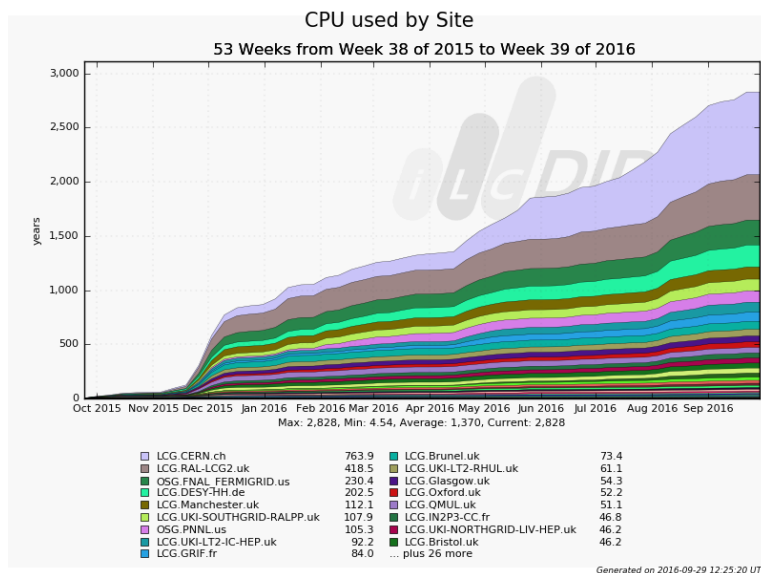


Figure 6. Cumulative CPU time used via the iLCDirac instance.

that offer `python` interfaces to control the MC generators, simulation, reconstruction and analysis software used in this community. The workflow modules also allow combining all these steps into a single job.

The computing resources are provided by grid sites in the OSG and EGI grids [21]. There are more than 40 grid sites accessed via iLCDirac. The largest provider of resources are DESY, FermiLab, RAL and CERN (see Figure 6). For accessing storage elements the XrootD and SRM protocols are used. The DIRAC file catalog (DFC) is used to store replica and meta data information. The (iLC)DIRAC transformation system is used for centralised productions. The transformation system uses meta data information in the DFC to chain generation, simulation, and reconstruction tasks.

In the 12 months leading up to October 2016 the iLCDirac instance was used to run 4 million jobs, which used a total of 2800 CPU years. 2 PB of new files were created.

3.5. BES III

The BESIII experiment studies electron–positron collisions in the tau–charm threshold region at the BEPCII storage ring, located in Beijing. The total data volume to be processed can reach

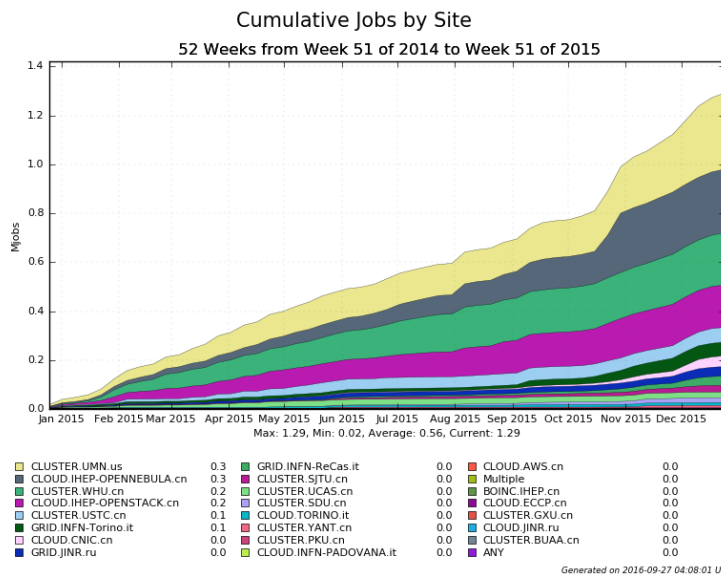


Figure 7. Cumulative jobs in the IHEP instance for 2015

1PB per year. For the past six years, the BESIII data processing activities have been mainly performed on the traditional cluster maintained by IHEP, where raw data is taken and stored. The BESIII distributed computing system aims to organize remote resources as complementary to the IHEP cluster, to meet peak needs. DIRAC has been proved to be the only existing grid solution that allowed the BESIII collaboration to build distributed computing in a fast way, with limited manpower and minimal efforts.

BESIII started to evaluate DIRAC for its distributed computing since the end of 2012, and later put in production at the beginning of 2014. Most of the computing resources comes from batch systems, while another 35% from Grids and Clouds. With VMDIRAC, the IHEP DIRAC instance is able to integrate cloud in an elastic way since the end of 2014. Besides pilot-based workload management, BESIII also uses the DFC (DIRAC File Catalog) as Replica Catalog, Metadata Catalog and Dataset Catalog to define and manage BESIII data. The BESDIRAC extension contains BESIII-specific services, which include Task Submission and Management System (TSMS), Massive Data Transfer System (MDTS), and Site Monitoring System (SMS), which is based DIRAC's RSS (Resource Status System). The MDTS is used to take care of the transfer of bulk datasets among sites, and includes accounting and error tracking. The SMS can report site status, and send warning messages when there are problems, in order to improve system operations. The TSMS has been developed to ease BESIII users to submit and manage massive jobs where the run-time workflow control adopts the concept of DIRAC workflow and allows applications to easily define their steps in one job and report status respectively. All these services were designed and developed within DIRAC well-defined services-oriented architecture.

Based on the multi-VO features of DIRAC, the IHEP DIRAC instance started to be shared among other IHEP experiments such as JUNO and CEPC. In 2015, the total job volume running in the instance reached 1.29 million, as seen in Figure 7.

4. Summary and prospects

DIRAC is an open source project, very actively used and developed. It satisfies the requirements and use cases of several communities by providing extensibility at several levels, the possibility to easily plug in new types of resources, and the flexibility requested for different use-cases. DIRAC also provides multi-VO support, which has proved to be particularly useful for smaller communities.

DIRAC can be extended to accommodate experiment-specific use cases, an option that gave

the possibility to several communities to develop their own extensions, commonly used for interfacing to experiment software, and production handling.

References

- [1] Tsaregorodtsev A and the Dirac Project 2014 *Journal of Physics: Conference Series* **513** 032096 URL <http://stacks.iop.org/1742-6596/513/i=3/a=032096>
- [2] Garca M U, Munoz V M, Stagni F, Cabarro B, Rauschmayr N, Charpentier P and Closier J 2014 *Journal of Physics: Conference Series* **513** 032099 URL <http://stacks.iop.org/1742-6596/513/i=3/a=032099>
- [3] McNab A, Stagni F and Garcia M U 2014 *Journal of Physics: Conference Series* **513** 032065 URL <http://stacks.iop.org/1742-6596/513/i=3/a=032065>
- [4] Himyr N, Blomer J, Buncic P, Giovannozzi M, Gonzalez A, Harutyunyan A, Jones P L, Karneyeu A, Marquina M A, McIntosh E, Segal B, Skands P, Grey F, Gonzalez D L and Zacharov I 2012 *Journal of Physics: Conference Series* **396** 032057 URL <http://stacks.iop.org/1742-6596/396/i=3/a=032057>
- [5] Cardoso L G, Gaspar C, Callot O, Closier J, Neufeld N, Frank M, Jost B, Charpentier P and Liu G 2012 *Journal of Physics: Conference Series* **396** 032052 URL <http://stacks.iop.org/1742-6596/396/i=3/a=032052>
- [6] Stagni F, Tsaregorodtsev A, McNab A and Luzzi C 2015 *Journal of Physics: Conference Series* **664** 062061 URL <http://stacks.iop.org/1742-6596/664/i=6/a=062061>
- [7] Meusel R, Blomer J, Buncic P, Ganis G and Heikkila S 2015 *Journal of Physics: Conference Series* **608** 012031 URL <http://stacks.iop.org/1742-6596/608/i=1/a=012031>
- [8] Riahi H, Salichos M, Keeble O, Andreeva J, Ayllon A A, Girolamo A D, Magini N, Roiser S and Simon M K 2015 *Journal of Physics: Conference Series* **664** 062051 URL <http://stacks.iop.org/1742-6596/664/i=6/a=062051>
- [9] Tsaregorodtsev A and Poss S 2012 *Journal of Physics: Conference Series* **396** 032108 URL <http://stacks.iop.org/1742-6596/396/i=3/a=032108>
- [10] Stagni F, Charpentier P, Graciani R, Tsaregorodtsev A, Closier J, Mathe Z, Ubeda M, Zhelezov A, Lanciotti E, Romanovskiy V, Ciba K D, Casajus A, Roiser S, Sapunov M, Remenska D, Bernardoff V, Santana R and Nandakumar R 2012 *Journal of Physics: Conference Series* **396** 032104 URL <http://stacks.iop.org/1742-6596/396/i=3/a=032104>
- [11] Currie R, Elmsheuser J, Fay R, Owen P H, Richards A, Slater M, Sutcliffe W and Williams M 2015 *Journal of Physics: Conference Series* **664** 062010 URL <http://stacks.iop.org/1742-6596/664/i=6/a=062010>
- [12] Abashian A 2002 *Nucl. Instrum. Meth.* **A479** 117–232
- [13] Hara T 2016 *J. Phys. Conf. Ser.* **These Proceedings**
- [14] Ahn S 2010 *Journal of the Korean Physics Society* **57** 715–724
- [15] URL <https://twiki.cern.ch/twiki/bin/view/LCG/LfcWlcg>
- [16] CTA Consortium 2011 *Experimental Astronomy* **32** 193–316
- [17] Arrabito L *et al.* 2016 *J. Phys. Conf. Ser.* **These Proceedings**
- [18] Arrabito L, Bregeon J, Haupt A, Diaz R G, Stagni F and Tsaregorodtsev A 2015 *Journal of Physics: Conference Series* **664** 032001 URL <http://stacks.iop.org/1742-6596/664/i=3/a=032001>
- [19] Muñoz V M, Albor V F, Diaz R G, Ramo A C, Pena T F, Arvalo G M and Silva J J S 2012 *Journal of Physics: Conference Series* **396** 032075 URL <http://stacks.iop.org/1742-6596/396/i=3/a=032075>
- [20] Grefe C, Poss S, Sailer A and Tsaregorodtsev A 2014 *J. Phys. Conf. Ser.* **513** 032077
- [21] Sailer A and Petric M 2016 *J. Phys. Conf. Ser.* **These Proceedings**