

ADVANCING ACCELERATOR SCIENCE THROUGH DATA-INTENSIVE RESEARCH AND TRAINING*

C. P. Welsch^{†, 1}, Cockcroft Institute, Warrington, UK
¹also at University of Liverpool, Liverpool, UK

Abstract

The Liverpool Centre for Doctoral Training for Innovation in Data Intensive Science (LIV.INNO) continues to make significant progress in developing precision diagnostics for accelerator facilities. This contribution presents recent results from four projects that collectively demonstrate how data-intensive methods, advanced modelling and modern instrumentation can enhance measurements under challenging conditions.

It is shown how detailed studies into detector dead time, after-pulsing and timing for the LHC Longitudinal Density Monitor provide quantified error budgets and mitigation strategies that improve satellite and ghost bunch measurements; how a comprehensive evaluation of multimode-fiber imaging techniques offers practical guidance on model selection, robustness and achievable accuracy for radiation-tolerant transverse profile diagnostics; how measurements of optical transition radiation (OTR) from low-energy electrons can be used to establish clear dependencies on target roughness and incidence angle, and how this allows experiment optimization; how the LHC Beam Gas Curtain has developed into a reliable, minimally invasive beam monitor, used for monitor calibration and machine optimization. In addition, this contribution gives an overview of the comprehensive LIV.INNO training, its recent and forthcoming events, and shows how structured skills development and collaboration across science, healthcare, industry and sustainability drive innovation.

INTRODUCTION

LIV.INNO was established in 2022 to train almost 40 PhD students in data intensive research, stretching across Nuclear, Particle and Astrophysics, as well as Accelerator Science [1]. The center builds on the success of its predecessor, the LIV.DAT Centre for Doctoral Training, which successfully trained another 40 PhD students between 2017 and 2023 [2]. Both centers leverage Liverpool's long-standing leadership in postgraduate training from major training programs, including several Marie Skłodowska-Curie Actions networks [3].

RESEARCH

The research activities within LIV.INNO are unified by a common focus on extracting reliable physical insight from complex, high-volume and often imperfect experimental or simulated datasets. The projects presented

in the following sections demonstrate how data-intensive methodologies enable measurements and optimization under conditions where conventional approaches face significant limitations. Common themes include the application of machine learning for reconstruction and inference, Monte Carlo techniques for modelling detector response and uncertainty propagation, statistical analysis of low-signal measurements, and the integration of simulation with experimental data to improve robustness and predictive capability. These approaches are complemented by advanced instrumentation, real-time data acquisition and high-performance computing techniques that together support precision diagnostics for next-generation accelerator facilities.

Longitudinal Density Monitor for the LHC

The Beam Synchrotron Radiation Longitudinal density monitor (BSRL) at the Large Hadron Collider (LHC) provides non-invasive measurements [4, 5] of the longitudinal charge distribution of circulating beams using time-correlated single-photon counting of synchrotron radiation. These measurements are particularly important for quantifying nominally empty low-population ghost and satellite bunches, which can influence machine protection and luminosity calibration. Recent work by Alex Jury has focused on characterizing the full measurement chain, combining historical measurements, dedicated laboratory measurements, cross-comparison with independent diagnostics, and simulation.

Comparison of BSRL bunch-length measurements with independent diagnostics suggested that the use of the Beam Synchronous Timing (BST) reference was consistent with an observed effective Gaussian timing broadening of approximately 300 ps [6]. This was identified as a dominant source of timing uncertainty. Following this analysis, the BSRL timing input was taken directly from the RF system during the 2023-2024 technical stop. Post-intervention data show improved timing consistency and more reliable bucket assignment.

Operational data showed a gradual reduction in detected counts per unit beam charge over the course of Run 3. Additionally, measurements of the optical fiber used to transport the synchrotron radiation to the detector in a radiation shielded area, together with wavelength-dependent transmission behavior and partial recovery during beam-free periods, are consistent with radiation-induced attenuation in the current fiber system [7]. Chromatic dispersion in the transmission line remains a significant contributor to pulse broadening. Current estimates indicate total spreading of approximately 569 ps for Beam 1 and 629 ps for Beam 2. These results motivate fiber replacement or removal.

* This work was supported by the Science and Technology Facilities Council (STFC) under grant agreement ST/W006766/1.

† carsten.welsch@cockcroft.ac.uk

Laboratory measurements confirmed that dead-time effects can distort measured populations when count rates become excessive. For the legacy acquisition chain (Detector and electronics), the combined effective dead time is of order 14-15 ns. This particularly affects trailing structure within a slot and can bias low-level populations if uncorrected. A replacement *PicoQuant* Time to Digital Converter (TDC) has been installed in parallel with the existing system [8]. The upgraded unit offers substantially shorter intrinsic dead time (~ 1 -2 ns) together with finer timing precision. Parallel measurements show broad agreement for corrected bunched observables, generally at the few-percent level, while larger differences remain in total intensities. This suggests improved sensitivity of the new system to low-level background or edge populations rather than disagreement in the core bunch signal.

A modular Monte Carlo simulation has been developed to reproduce photon generation, optical attenuation and dispersion, detector efficiency, after-pulsing, dark counts, TDC dead time, jitter, and histogram formation. The model provides a controlled environment for testing upgrade scenarios and estimating systematic biases in derived observables. It is now being used to compare mitigation strategies including revised filtering, new fibers, and updated correction algorithms.

Reconstruction of Transverse Beam Distribution using Machine Learning

Screen-based imaging is widely used for measuring the transverse beam distribution in accelerators, but in radiation-exposed areas cameras and associated electronics can degrade significantly. This has become more relevant with the discontinuation of the radiation-hardened *Vidicon* systems historically used at CERN. A proposed solution is to relay scintillation light to a remote low-dose area through a single large-core multimode fiber and reconstruct the original distribution from the fiber output [9]. The challenge is that multimode-fiber propagation scrambles the spatial information through mode coupling, modal interference and scattering, producing a speckle pattern rather than a directly usable image. Since representative real beam data are limited and recalibration may be needed under environmental perturbations, a reconstruction approach with reduced dependence on large real-beam datasets is of interest [10, 11].

To address the limited availability and uneven coverage of real beam data, the work of Qiyuan Xu and co-workers investigates a synthetic-data-driven reconstruction approach. A stochastic Gaussian mixture (SGM) model was developed to generate large numbers of high-variance synthetic patterns spanning a broader region of the relevant parameter space than the available experimental dataset. These synthetic patterns were displayed on a laser-illuminated digital micromirror device and transmitted through a 5 m multimode-fiber optical system. The same setup was then used to evaluate previously acquired real beam images from CERN's CLEAR facility, replayed on the DMD. For reconstruction, a modified convolutional autoencoder was adopted. The model removes skip

connections and pooling layers, uses stride-2 convolutions for learnable down sampling, and relies on a compact bottleneck representation to favor generalization to the main beam structure rather than recovery of all high-frequency details. The aim was stable reconstruction of physically relevant transverse beam parameters under realistic multimode-fiber scrambling.

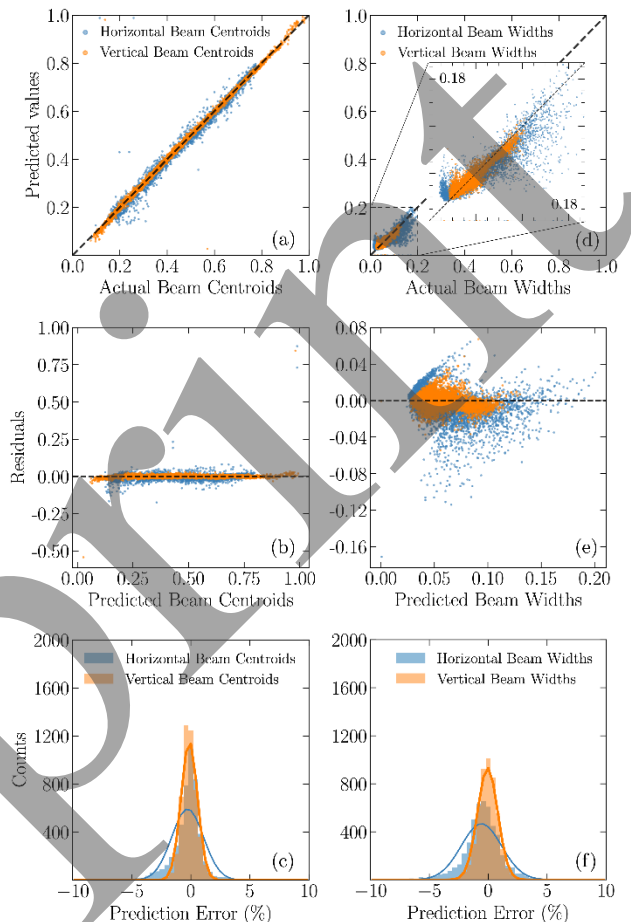


Figure 1: Statistics of the reconstruction errors for four transverse beam parameters on the test set.

A comparison of the two training strategies shows that using the original real-beam dataset alone for training leads to larger reconstruction errors and a visible prediction bias in the beam centroid estimates, consistent with the limited coverage and uneven parameter distribution of that dataset. In contrast, the same model trained only on the synthetic SGM dataset gives predictions that are closer to the ground truth and more evenly distributed about the ideal agreement line. This behavior is consistent with the broader coverage of the relevant parameter space provided by the synthetic training data, which helps reduce the effect of data scarcity and distributional imbalance in the reconstruction task. Fig. 1 (a)–(c) illustrates the beam position reconstruction on test set, with blue and yellow representing the horizontal and vertical parameters, respectively. In Fig. 1 (a), the scatter points closely follow the 45° line, indicating strong agreement between predicted and actual values. The residuals, computed as the

difference between actual and predicted values, are shown in Fig. 1 (b). The residual plot shows that most samples have minimal error, with only a few exhibiting larger deviations. Figure 1 (c) displays the residual distribution, clipped within $\pm 10\%$ for better visualization, with 99.3% of predictions concentrated within $\pm 5\%$. Similarly, beam width reconstruction results are shown in Fig. 1 (d)–(f), where errors remain within a smaller range. As seen in Fig. 1 (f), 98.7% of residuals also lie within $\pm 5\%$, and more importantly, the centered distribution indicates no systematic prediction bias.

Single-Ion Detection via UV-VIS Emission from Relativistic Heavy Ions

Reliable, fast, and radiation-hard particle counting is a prerequisite for high-intensity heavy-ion beam operation at GSI-FAIR. The project of LIV.INNO PhD candidate Rupeshkumar Ghagi investigates two ultraviolet-visible (UV-VIS) photon emission mechanisms: beam-induced fluorescence (BIF) in air and optical transition radiation (OTR) from metallic surfaces as the physical basis for a novel class of single-ion detectors.

Theoretical estimates of photon yields were used to guide detector design, and single-ion detection was demonstrated in proof-of-concept experiments [12]. Building on these results, further detector iterations were developed, informed by experimental data and optical transport simulations. These iterations target improved detection efficiency and quantitative comparison of the radiation mechanisms contributing to the UV-VIS signal.

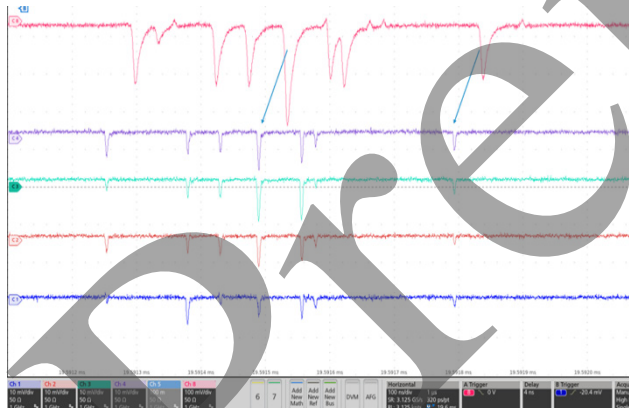


Figure 2: Measured signals during slow extraction.

Figure 2 shows representative oscilloscope waveforms recorded during slow extraction of Ag^{45+} at 500 MeV/u at GSI. It can be seen that several individual ion transits are clearly resolved within the acquisition window. A scintillator-based reference detector (C8) provides the trigger, while channels C1-C4 record UV-VIS emission signals from four detector configurations under simultaneous test. The pulse-by-pulse correspondence between the reference and detector channels confirms single-ion sensitivity across all configurations. Variations in pulse amplitude between the channels reflect differences in the optical yield and collection efficiency across different configurations.

Beam Gas Curtain (BGC) Monitor for High Luminosity LHC

Beam profile measurements for high-intensity high-power beams are very challenging due to the destructive beam power. Invasive methods, such as wire scanner and secondary beam monitor, only work with reduced intensity for machines such as the Large Hadron Collider (LHC). Most high-intensity facilities don't have beams with energy high enough to produce synchrotron radiation for diagnostics purposes. Even for the LHC, although the visible synchrotron radiation monitor provided a non-invasive profile measurement only for the proton beam, it suffers from resolution drifting and requires regular calibration. Previously, students from LIV.DAT and the early phase of the LIV.INNO managed to install the beam gas curtain monitor (BGC) into the LHC and measure the full cycle of both proton and lead ion beams in the LHC, including the injection, energy ramping-up and stable operation phases of the LHC operation, see Fig. 3 [13]. The method utilizes a thin supersonic gas curtain to interact with the LHC beams to produce the fluorescence signal, which can be analyzed to reveal the beam profile [14].

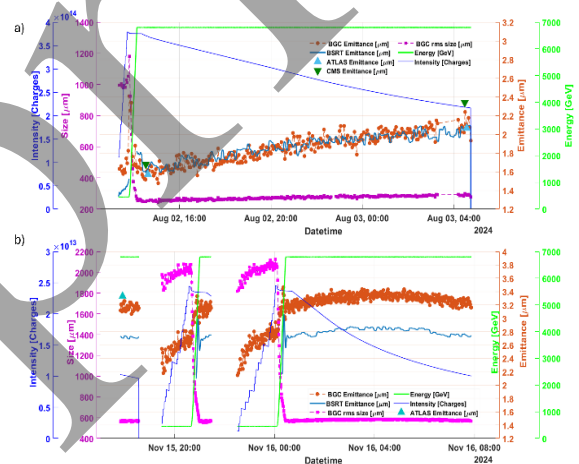


Figure 3: Horizontal emittance measurements from the BGC monitor for a) the LHC proton beam and b) the lead ion beam compared with other methods [13].

The PhD project of Ravi Gupta consists of delivering and commissioning two additional BGC monitors for CERN. The first BGC monitor was recently commissioned at the Cockcroft Institute (CI). Whilst ultimately targeting measurements in the LHC, it will initially be used for beam profile monitoring of the hollow electron beam at the Electron Beam Test Stand (EBTS) at CERN. An earlier version of the monitor was successfully used to characterize the hollow electron beam successfully, albeit over a limited beam size of only ~ 18 mm [15]. The new system that was shipped to CERN in early 2026 is based on an optimized curtain with a width of up to 42 mm, i.e. large enough to cover the whole range of electron beams produced in the EBTS. Work will now focus on paving the way for a monitor that can simultaneously characterize an electron and ion beam. A second BGC monitor will be installed in beam 2 of the LHC. Compared with the one already installed in beam 1, the space is much tighter. This

requires installing the device at an angle of 45 degrees and install two optical lines where light will be captured from two axes to increase the signal, whilst reducing distortion. This requires detailed simulations of gas curtain forming and optimization using machine learning techniques to increase the signal level and mitigate the risk of misalignment caused by mechanical tolerances. This monitor is expected to be assembled and commissioned at the Cockcroft Institute in the second half of 2026 and installed in beam 2 during the LHC Long Shutdown 3.

TRAINING

The scientific training in LIV.INNO is organized around three core Work Packages covering Monte Carlo Methods and High-Performance Computing (HPC), Artificial Intelligence (AI) and Machine Learning (ML), and Data Analysis. These underpin a structured first-year training program requiring 120 credits of coursework delivered jointly by the University of Liverpool and Liverpool John Moores University, complemented by contributions from external industry specialists and international researchers. The training provides students with advanced knowledge in statistical methods, simulation, AI/ML techniques, scientific computing and data analytics, together with domain-specific applications across accelerator science, healthcare, astronomy and particle physics.

From Year 1 onwards, each student undertakes a personalized Development Needs Analysis (DNA) and Career Development Plan (CDP), enabling training activities to be tailored to individual research projects and career aspirations. Training delivery combines lectures, seminars, specialist workshops, schools and collaborative research activities. Students also benefit from access to LIV.HUB, a dedicated multi-purpose training and collaboration environment that supports technical training, hackathons, industrial engagement and interdisciplinary interaction.

A central component of the LIV.INNO model is the mandatory 6-months industry or public-sector placement, during which students apply their expertise to challenges outside their core PhD research. These placements span sectors including healthcare, rail transportation, cybersecurity, software engineering, digital innovation and environmental science. Recent and past projects have included optimization of low-dose 3D X-ray imaging systems with Adaptix Ltd, development of safety and productivity tools for railway operations, surrogate modelling for epidemiological applications at IBM Research (Hartree Centre), and commercial natural language processing tools for software tender analysis. Students have also contributed to digital mental-health platforms and industrial data classification systems, demonstrating the broad applicability of data intensive science across multiple sectors. In addition to technical expertise, LIV.INNO places strong emphasis on transferable skills training, including project management, leadership, scientific communication, networking, entrepreneurship and public engagement. These activities are designed to equip students for careers in both academia

and industry. The Centre also collaborates closely with international training initiatives, including the EuPRAXIA Doctoral Network, with which it delivers joint schools and introductory skills training [16].

LIV.INNO maintains an active program of seminars, workshops and outreach activities that connect academia, industry and the wider public. The center also hosted the 2024 STFC Summer School on Data Intensive Science in Liverpool [17], followed by the AI for Innovation Summit in May 2025, which brought together more than 100 delegates from academia and industry [18]. A workshop on Data Science in Healthcare and Health Technologies took place at The Spine, Liverpool, in June 2025 [19]. Additional activities include the hosting of the RL4AA'26 Workshop on Reinforcement Learning for Autonomous Accelerators [20], a Symposium on Innovations for a Sustainable Tomorrow [21], and public engagement initiatives such as the "AI & eyes" outreach activities and data science exhibitions at the British Science Festival.

Through this integrated and multidisciplinary training model, LIV.INNO equips doctoral researchers with advanced scientific expertise, industrial experience and the transferable skills required to address future challenges in accelerator science and other data-intensive research domains. This is widely regarded as an extremely successful model for training the next generation of scientists and engineers, and as a template for others to follow. The center director was invited to speak about the state of doctoral training in Europe at the European Commission-hosted event on "More than a PhD: The added value of MSCA doctorates" celebrating the 30th anniversary of the MSCA in Brussels on 24/54 June 2026.

More research and training is needed in this important area and will be carried out within the new Horizon Europe projects TwinRISE [22], iRIS [23] and EPITA [24]. In addition, dedicated training schemes have been proposed including a new STFC Center for Doctoral Training with a focus on AI and a European Doctoral Network on AI for Accelerators, ARTIFACT-DN.

SUMMARY AND OUTLOOK

This paper presented recent advances in accelerator diagnostics enabled through data-intensive methodologies. Studies of the LHC Longitudinal Density Monitor quantified key systematic effects and informed upgrades for improved low-intensity bunch measurements. Synthetic-data-driven machine learning demonstrated robust and unbiased reconstruction of transverse beam parameters from multimode-fiber imaging, highlighting the potential of AI-assisted radiation-tolerant diagnostics. In addition, UV-VIS single-ion detection studies and ongoing BGC developments illustrated how simulation, statistical analysis and modern instrumentation support precision beam characterization under challenging conditions. Together with LIV.INNO's interdisciplinary training, these activities demonstrate how data science techniques are driving innovation in accelerator diagnostics and future data-driven accelerator operation.

REFERENCES

- [1] LIV.INNO, www.livinno.org
- [2] C. P. Welsch, "Using Data Intensive Science for Accelerator Optimization", in *Proc. IPAC'22*, Bangkok, Thailand, Jun. 2022, pp. 980-982.
[doi:10.18429/JACoW-IPAC2022-TUPOST051](https://doi.org/10.18429/JACoW-IPAC2022-TUPOST051)
- [3] A. Jeff, "A longitudinal density monitor for the LHC", PhD thesis, University of Liverpool, UK, 2012.
- [4] A. Jeff, *et al.*, "Longitudinal density monitor for the LHC", *Phys. Rev. Spec. Top. Accel. Beams*, vol. 15, p. 032803, 2012. [10.1103/PhysRevSTAB.15.032803](https://doi.org/10.1103/PhysRevSTAB.15.032803)
- [5] M. Gonzalez Berges, *et al.*, "Beam synchrotron radiation for the longitudinal density monitor (BSRL), error analysis", CERN, Geneva, Switzerland, Rep. CERN-EDMS-3324712, 2025.
- [6] A. Jury *et al.*, "Performance analysis of the LHC BSRL and possible improvements", in *Proc. IBIC'25*, Liverpool, UK, Sep. 2025, pp. 35-40.
[doi:10.18429/JACoW-IBIC2025-MODC02](https://doi.org/10.18429/JACoW-IBIC2025-MODC02)
- [7] G. Sourpi, "Development of the longitudinal density monitoring system for the LHC particle accelerator", Diploma thesis, National Technical University of Athens, Greece, 2025. [doi:10.26240/heal.ntua.31013](https://doi.org/10.26240/heal.ntua.31013)
- [8] G. Trad and S. Burger, "Artificial Intelligence-Assisted Beam Distribution Imaging Using a Single Multimode Fiber at CERN," in *Proc. IPAC'22*, Bangkok, Thailand, Jun. 2022, pp. 339-342.
[doi:10.18429/JACoW-IPAC2022-MOPOPT041](https://doi.org/10.18429/JACoW-IPAC2022-MOPOPT041)
- [9] Q. Xu, *et al.*, "Estimation of Beam Transverse Parameters Through a Multimode Fiber Using Deep Learning," in *Proc. IBIC'24*, Beijing, China, Sep. 2024, pp. 170-173.
[doi:10.18429/JACoW-IBIC2024-TUP54](https://doi.org/10.18429/JACoW-IBIC2024-TUP54)
- [10] Q. Xu, *et al.*, "Evaluating machine learning models for multimode-fiber-based transverse beam profile reconstruction," in *Proc. IBIC 2025*, Liverpool, UK, Sep. 2025, pp. 373-377.
[doi:10.18429/JACoW-IBIC2025-TUPC011](https://doi.org/10.18429/JACoW-IBIC2025-TUPC011)
- [11] Q. Xu, *et al.*, "Development of radiation-tolerant beam imaging via multimode fiber and synthetic data-driven machine learning," *Phys. Rev. Accel. Beams*, vol. 29, p. 032801, 2026. [doi:10.1103/wddb-m371](https://doi.org/10.1103/wddb-m371)
- [12] R. Ghagi, *et al.*, "UV-VIS emissions from high energy heavy ions: mechanisms and applications to single-ion detection," in *Proc. IBIC 2025*, Liverpool, UK, paper MOPMO28.
[doi:10.18429/JACoW-IBIC2025-MOPMO28](https://doi.org/10.18429/JACoW-IBIC2025-MOPMO28)
- [13] O. Sedlacek, *et al.*, "Full-cycle, noninvasive emittance monitoring with the beam gas curtain monitor at the LHC", *Phys. Rev. Research* vol. 7, no. 4, p. 043144, Nov. 2025.
[doi:10.18429/5ggy-f81m](https://doi.org/10.18429/5ggy-f81m)
- [14] A. Salehilashkajani, *et al.*, "A gas curtain beam profile monitor using beam-induced fluorescence for high intensity charged particle beams", *Appl. Phys. Lett.*, vol. 120, no. 17, p. 174101, 2022, [doi:10.1063/5.0085491](https://doi.org/10.1063/5.0085491)
- [15] O. Sedláček, *et al.*, "Gas jet-based fluorescence profile monitor for low energy electrons and high energy protons at LHC", in *Proc. IBIC 2023*, Saskatoon, Canada, Sep. 2023, pp. 312-317.
[doi:10.18429/JACoW-IBIC2023-WE3I01](https://doi.org/10.18429/JACoW-IBIC2023-WE3I01)
- [16] LIV.INNO Introductory Skills School , indico.ph.liv.ac.uk/event/1839/
- [17] LIV.INNO Summer School on Data Science, indico.ph.liv.ac.uk/event/1639/
- [18] LIV.INNO AI for Innovation Summit, <https://bit.ly/ai-for-innovation-summit>
- [19] LIV.INNO Data Science in Healthcare and Health Technologies Workshop, indico.ph.liv.ac.uk/event/1999/
- [20] LIV.INNO Workshop on Reinforcement Learning for Autonomous Accelerators (RL4AA'26), indico.ph.liv.ac.uk/event/2025/
- [21] Symposium on Innovations for a Sustainable Tomorrow, agenda.infn.it/event/49199/
- [22] TwinRISE project, artifact-network.org/initiatives/twinrise/
- [23] iRIS home page, www.irisproject.eu
- [24] EPITA home page, epita.web.cern.ch/