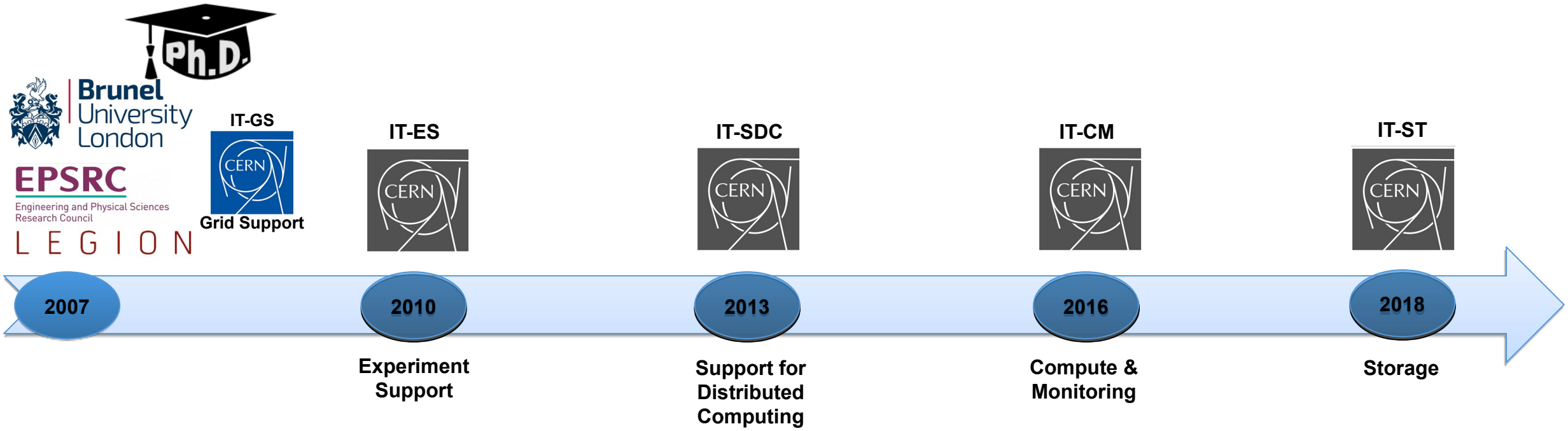
My Introduction

Edward Karavakis

NPPS Group meeting

21st October 2022

Past Experience

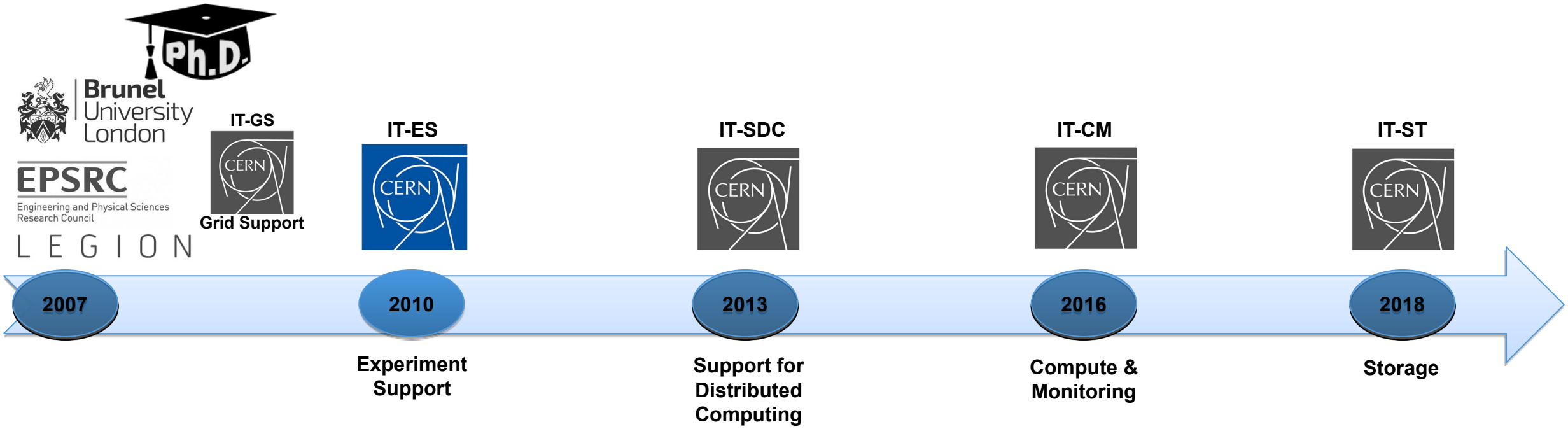


PhD in Computer Science

- Legion Ltd, London, UK (Jan. 07 – Jun. 09)
  - Developed a distributed version of a commercial pedestrian simulation software
  - Used for planning London Olympics, train stations and London metro stations
  - Pedestrian pathways were shaped using this software

- CERN, IT (May 08 – Apr. 10)
  - Developed two distributed job monitoring applications for CMS using the Dashboard Python Framework
    - In collaboration with several CMS physicists, supporting all CMS submission tools
  - > 10M. job status updates / day from > 120 computer centres
  - Web UIs displayed the status for users with different roles
    - Made it easy to identify problems and inefficiencies!
  - Used daily by more than 400 physicists – provided daily user support

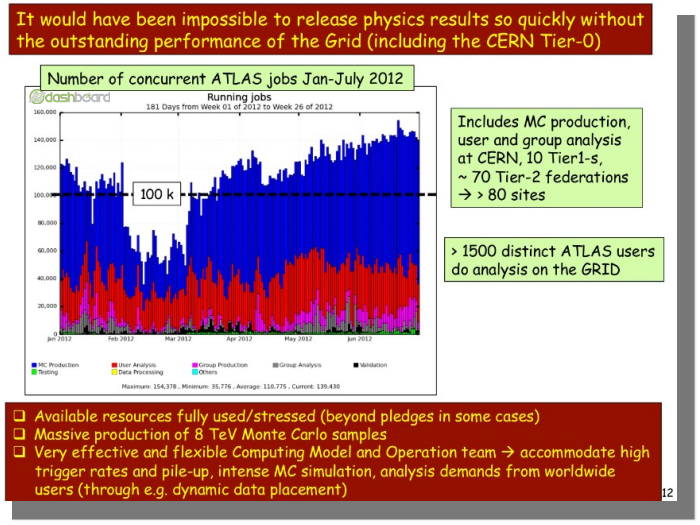
Past Experience



Senior Research Fellow

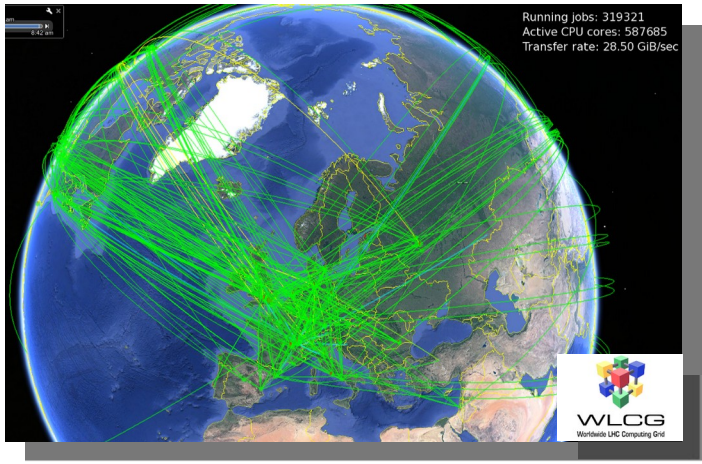
CERN IT-ES (2010 to 2013)

- Co-responsible for development and maintenance of Dashboard Python Framework and cluster – Led development and support of ATLAS and CMS job monitoring and accounting
  - Ported apps from CMS to ATLAS: Supporting Ganga, PanDA, ProdSys, TZero - translating all job monitoring data into one common schema
    - > 25M. job status updates per day from > 150 computer centres
    - > 1000 users daily from all over the world – provided daily user support
  - Used by individual scientists, operations teams, site admins and management



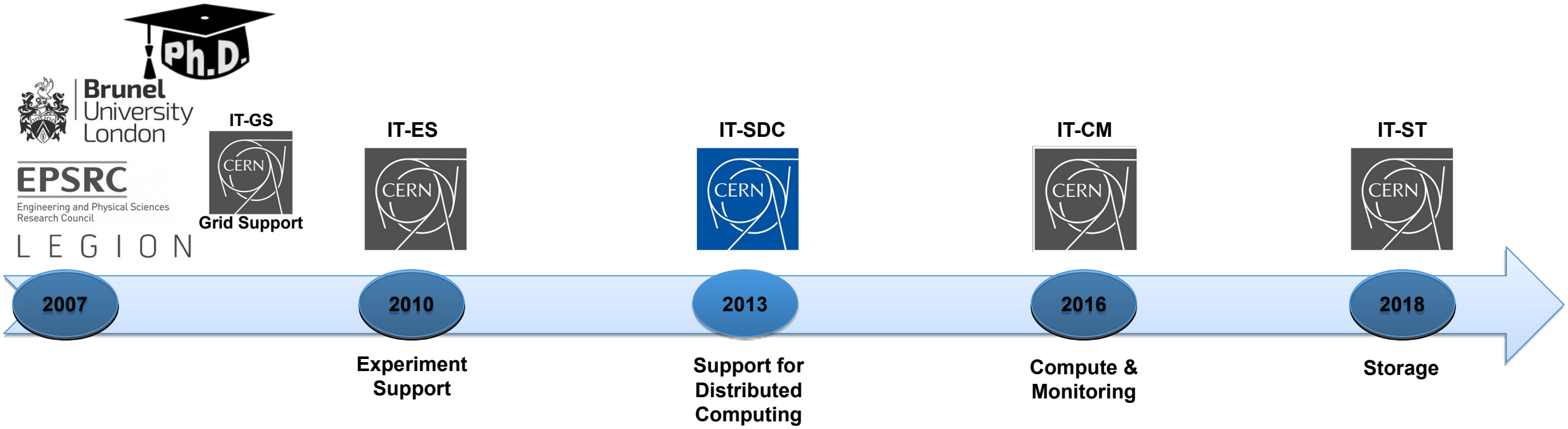
"Status of Standard Model Higgs searches in ATLAS" by Fabiola Gianotti on the Higgs Discovery day

- In charge of WLCG Google Earth Dashboard
  - For dissemination purposes at CERN and in other computing centres
  - Visual representation of complexity and scale of the Grid



- Member of "*Oracle Experts Task Force*" and section responsible for database optimisations
  - Important speed improvements by replacing DB triggers with stored procedures for bulk processing

Past Experience



Staff Computer Engineer

CERN IT-SDC (2013-2015)

- Project lead of WLCG Dashboards for real-time monitoring and accounting of job processing and distributed data management activities
  - Supervised team of 3 people with occasional contributions from ATLAS and CMS
  - Used by LHC Experiments Committee, Resources Review Board and LHC Computing Resources Scrutiny Group

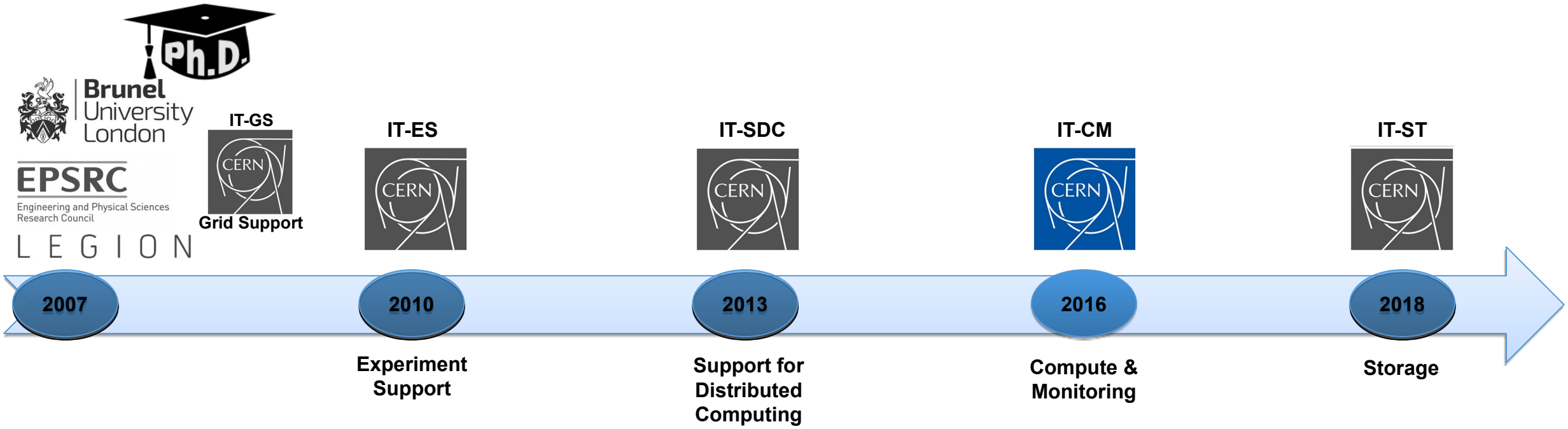
- Responsible for the official WLCG topology and resources/pledges system
  - Implemented requested features from the WLCG Office
  - Automated monthly report generation for the usage of WLCG

- First prototype of AGIS for CMS – CRIC

- "*NoSQL for WLCG Monitoring solutions*" Project Coordinator
  - Supervised team of 6 people with remote contributions from JINR, Moscow
  - Benchmarked performance of our Oracle cluster, compared it with ElasticSearch and HBase
  - Formed the basis for next generation unified monitoring system

- Led development, deployment and commissioning of gLExec within PanDA Pilot
  - Pilot switched to the user submitting the job to run payload within a secure sandbox
  - Deployed in half of ATLAS analysis sites - transparent to normal ATLAS operations
    - Handed over to the ATLAS Distributed Computing Operations team

Past Experience



Staff Computer Engineer

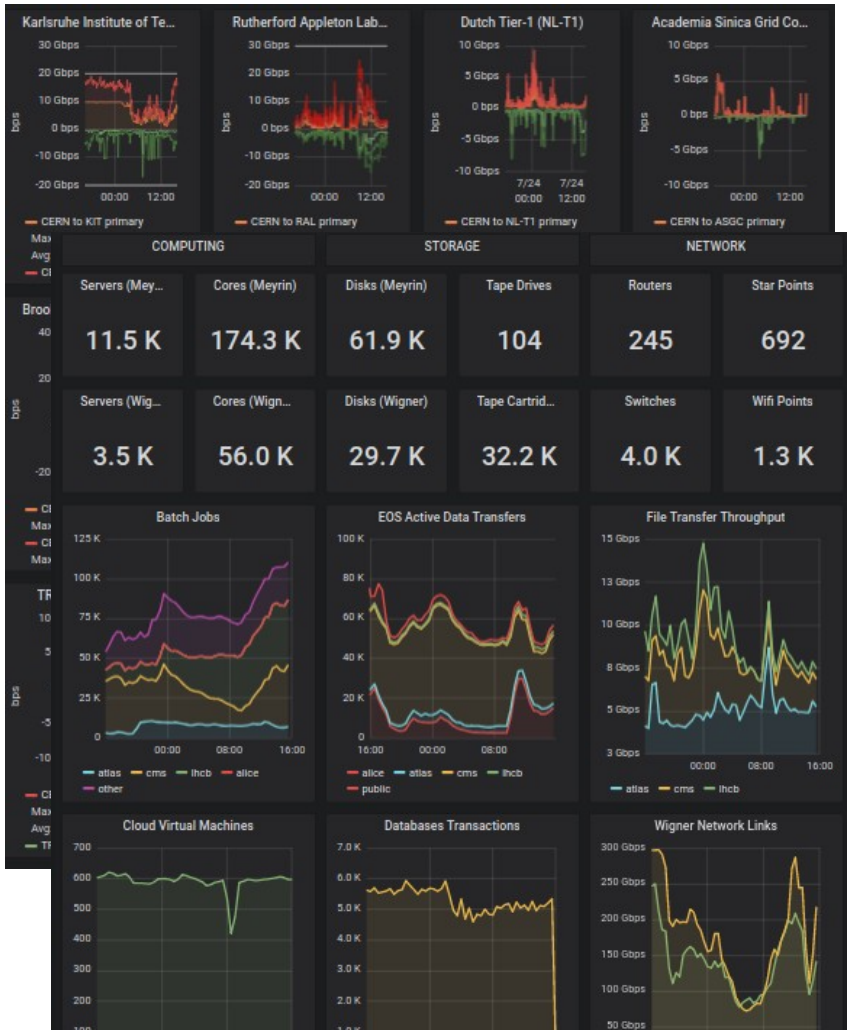
CERN IT-CM (2016-Feb. 2018)

- Merged WLCG and IT Data Centre monitoring
  - Key role in definition, evaluation and deployment of components for metrics and logs collection, analysis, storage and visualisation
  - Moved to established *open source technologies*
    - *Replaced Oracle and custom built Web User Interfaces*
    - *Replaced in-house developed LEMON with Collectd*
  - Used by LHC experiments for monitoring and CERN IT for host monitoring, KPIs, Service Availability and High Level Service dashboards

- Performing advanced data aggregation and enrichment using Spark in production environment with more than 3 TBs of data per day
  - Metrics and logs from 15K servers (~36K VMs) from CERN Data Centre
  - Job monitoring, data transfers and site status metrics from > 170 centres
  - Provided data analytics platform with Kafka/Spark and Interactive Notebooks

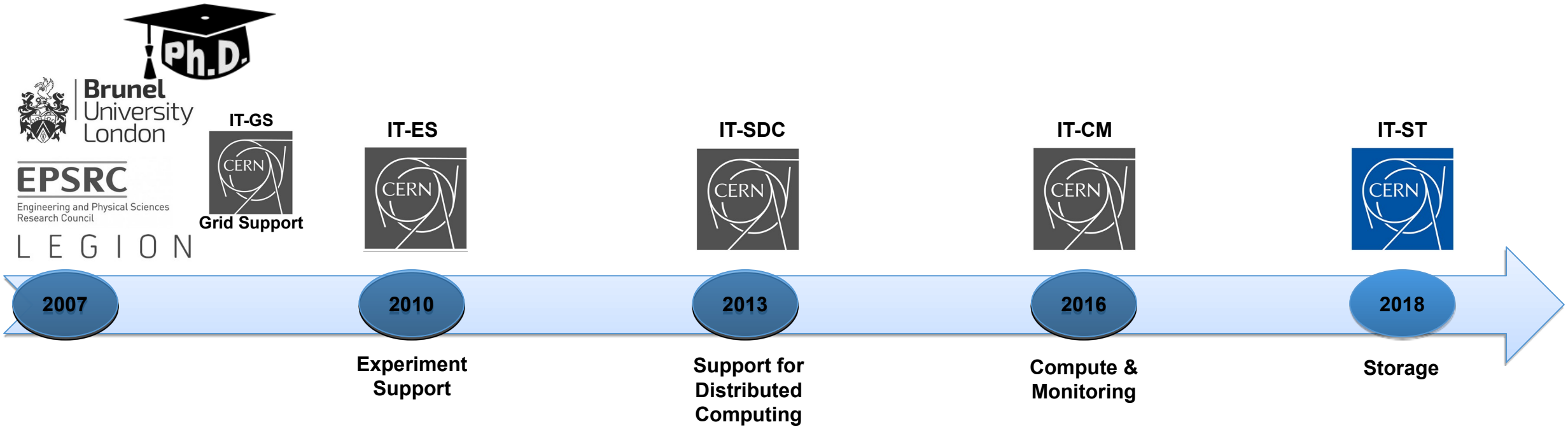
- (Co-)Supervised 5 people using Agile practices
  - Provided expertise and training for WLCG data sources

- Trainer of “Agile IT infrastructure / IT Tools for Service Managers”
  - Trained IT newcomers on OpenStack, Puppet and host monitoring



collectd kafka elasticsearch InfluxDB FLUME logstash Grafana Spark hadoop

Past Experience



Staff Computer Engineer

CERN IT-ST (Mar. 2018-Jun. 2021)

- Lead developer of the File Transfer Service (FTS) project distributing majority of LHC data
  - Transferred >1 billion files and 1 exabyte of data with transfer rates reaching ~60 GB/s in 2020
  - Used by more than 37 experiments at CERN and in other data-intensive sciences

- Optimised MySQL Database access to scale FTS for LHC Run-3
  - 30-35% speed improvement of scheduler: more transfers being scheduled and executed

- Appointed as Project Leader and Service Manager at CERN
  - Managed team of 4 people with occasional contributions from community
  - In close contact with data management teams of experiments and experts from CERN IT Services

- Migrated CERNBox to new EOS infrastructure by developing and operating the tool
  - Successfully migrated 16000 CERNBox users online transparently: more than 5 PBs and 1B. files!
  - Recovery time of EOS Server went down from 2 hours to tens of seconds!

- CERN IT-ATLAS Liaison (Sept. 2019 – Jun. 2021)
  - Ensured effective communication between CERN IT and ATLAS

Current role

Physics Department
NPPS

Brookhaven
National Laboratory

2022

Nuclear and Particle Physics Software

Recent/Current Assignments

- PanDA for Vera Rubin Observatory
  - Kubernetes deployment for PanDA monitoring: helm charts, secrets, ..
  - Improved container deployment for PanDA monitoring, fully automated procedure, incorporated with github
  - Worked on improving PostgreSQL compatibility within PanDA, solved numerous incompatibility issues between Oracle and PostgreSQL for PanDA monitoring
- PanDA Core
  - Proposal (and implementation) for PanDA DB upgrades/new PanDA DB installations and compatibility versions - since we have to handle multiple DB instances, experiments and DB backends (not only the ATLAS DB at CERN)
- ProdSys II
  - (ToDo) Review ATLAS Production System and come up with a proposal to simplify/rewrite/modernize/modularize/unify(?)

Thank you for your time and attention

-

Questions?