










WLCG transition from X.509 to Tokens: Progress and Outlook

Thomas Dack¹^{*}, Federica Agostini², Berk Balci³, Matthew Doidge⁴, Dave Dykstra⁵, David Kelsey¹, Maarten Litmaath³, Roberta Miccoli², Mischa Sallé⁶, Hannah Short³, and Enrico Vianello², representing the WLCG Authorization WG

¹Science and Technology Facilities Council (UKRI-STFC), United Kingdom

²Istituto Nazionale di Fisica Nucleare (INFN-CNAF), Italy

³European Organization for Nuclear Research (CERN) Switzerland

⁴University of Lancaster, United Kingdom

⁵Fermi National Accelerator Laboratory (FNAL), United States

⁶Nationaal Instituut voor Subatomaire Fysica (Nikhef), Netherlands

Abstract. Since 2017, the Worldwide LHC Computing Grid (WLCG) has been working towards enabling token-based authentication and authorization throughout its entire middleware stack.

Taking guidance from the WLCG Token Transition Timeline, published in 2022, substantial progress has been achieved not only in making middleware compatible with the use of tokens, but also in understanding the limitations of the WLCG Common JWT Profiles, first published in 2019. Significant scalability experience has been gained from Data Challenge 2024, during which millions of files were transferred with tokens used as credentials - a significant percentage of the total transfers completed.

Besides describing the state of affairs in the transition to tokens, revisions to the WLCG token profile, and the evolving road maps, this contribution also covers the corresponding transition from VOMS-Admin to INDIGO-IAM services, with continuing improvements in terms of functionality as well as deployment.

1 Introduction

More than six years ago, at CHEP 2018, in the plenary talk *Beyond X.509: Token-based Authentication and Authorization for HEP* [1], the use of the INDIGO Identity and Access Management Service (INDIGO IAM) [2] as a replacement for the existing X.509 and VOMS (Virtual Organization Membership Service) [3] architecture was publicly proposed. This plenary presented the initial work done by the **WLCG Authorization Working Group** (the AuthZ WG), formed in the summer of 2017, and which had begun investigating and defining architecture requirements in the course of that year. Since that initial phase, significant progress has been made towards this transition, with deployment of IAM instances and integration and testing with experiments and services. This contribution serves to describe the main achievements in 2024 and what remains on the road map for 2025 and beyond.

*e-mail: thomas.dack@stfc.ac.uk

1.1 Contributing Groups

The AuthZ WG was formed in 2017 and brings together experts from several projects and domains – including SciTokens [4], the INDIGO DataCloud project [?] [5] and EGI [6] – in order to ensure that the authorization infrastructure is both full-featured and, to the extent possible, interoperable, and that technical and policy challenges faced are tackled appropriately.

2 WLCG Token Infrastructure

The WLCG token issuers are instances of INDIGO IAM, initially deployed on the OpenShift infrastructure of the CERN IT Department. These were the services that issued LHC experiments with tokens during WLCG Data Challenge 2024 and that replaced the VOMS(-Admin) legacy services later that year. Both milestones are further discussed below.

As of October 2024, IAM services for the LHC experiments have also been made available as High-Availability (HA) deployments on the Kubernetes service of the CERN IT Department. Since this transition, the experiments have gradually moved use cases to the new deployments, with the aim of switching completely by early 2025 - when all relevant services have had their configurations adjusted to accept tokens as well as VOMS proxies, from the new services.

The move to Kubernetes provides important improvements to the operation of the WLCG IAM instances, regarding these aspects:

- deployment and configuration through GitOps
- load-balancing
- high-availability options
- logging
- monitoring

2.1 Deployment Status

There are production instances of INDIGO IAM for the following research communities:

- ALICE,
- ATLAS,
- CMS,
- LHCb,
- MoEDAL,
- FCC,
- ILC,
- CALICE,
- AMBER,
- COMPASS.

Furthermore, instances are operated also for the Virtual Organizations DTeam and Ops, which are used for service availability and other tests on the European Grid Infrastructure (EGI). To support all these instances, the CERN IAM service team works in close collaboration with the INDIGO IAM development team at INFN CNAF.

2.2 VOMS(-Admin) Retirement

As the VOMS(-Admin) legacy services were not foreseen to be supported on operating systems newer than CentOS 7, it became essential to ensure the IAM services would be able to take over all required functionality before the CentOS 7 end of life on 30 June, 2024. LHC experiments had been preparing for this transition for more than two years, steadily migrating critical use cases from VOMS-Admin towards the IAM services. The final use case was the actual management of each Virtual Organization (VO). For as long as VOMS-Admin was still in charge for a VO, its corresponding IAM service database was synchronized multiple times per day. The last of the VOMS(-Admin) services at CERN were finally switched off in the course of June, 2024. The IAM services continue supporting the creation of VOMS proxies, which will remain needed for each VO as long as its transition to tokens has not been completed yet.

3 Token Evolution in the WLCG

3.1 Lessons Learned from Token Usage in Data Challenge 2024

WLCG Data Challenge 2024 took place from 12 through 24 February of that year and featured many millions of file transfers between storage services. A significant fraction of those transfers were done using only tokens for authentication and authorization, thus allowing invaluable experience to be gained in token provisioning at scale.

During the challenge, it already became clear that some of the ways in which tokens were used were not advisable for operation at scale. For some of the LHC experiments, File Transfer Service (FTS) instances and IAM services became overloaded with token exchange and refresh operations, which were required to allow transfer requests still to succeed after potentially staying queued for up to several days. In the aftermath, file transfers in production workflows have followed two distinct models, each allowing the load on FTS and IAM instances to remain sustainable:

- In CMS workflows, the number of active tokens remains relatively low by letting the scope of each token refer to a complete data set instead of an individual file; those tokens are exchanged by the FTS and then refreshed as needed.
- In ATLAS workflows, each token refers to an individual file and is given a lifetime of several days, which usually should be sufficient to allow the corresponding file transfer to finish before the token runs out; those tokens are neither exchanged nor refreshed.

The scalability of token operations in IAM has been investigated. The acquisition of access plus refresh tokens from an IAM instance on Kubernetes has been demonstrated at 900 Hz, possibly limited by the unnecessary storage of access tokens in the IAM database, a feature of the underlying MITREid engine that cannot easily be avoided. Work is underway to replace that legacy engine with Spring Authorization Server, which then will allow the storage of access tokens to be made configurable. Because of uncertainty about the maximum sustainable token rates and concerns about the operational support level that can be afforded to the IAM services, ATLAS collaborators have started looking into letting their data management service Rucio mint its own tokens for high-rate usage, thereby avoiding time-critical dependencies on their IAM service. INDIGO IAM would still be necessary in order to authenticate with Rucio, and to act as the user information authority for Rucio. As of February 2025, small-scale tests of that model were successful. In the ALICE experiment, data management has been carried out following a similar model for 20 years already, and also the DIRAC experts in the LHCb experiment intend to give it a try.

3.2 Token Profile Improvements

Version 2.0 of the WLCG Token Profile is under preparation to fix a number of issues encountered with v1.0, some of which are being worked around already in various middleware products implementing support for WLCG tokens. Several open items still require further discussions between parties concerned and may only be addressed in subsequent versions.

WLCG is a stakeholder in the Grand Unified Token Profile WG, which aims to specify a common base for the various token flavors used to interact with services in WLCG, OSG and EGI. In particular, it is very desirable to devise a common definition of an attribute denoting by which Virtual Organization a token has been issued, for subsequent use e.g. in accounting systems. While progress has been made in that area, a specification has not yet been agreed at the time of writing.

4 Next Steps

4.1 Next Milestones

The main milestones with tentative dates are:

- **M.9 (March 2025): Grid jobs use tokens for reading and stageout**

There will be several implementations of that functionality across different experiments, each of which will need to deal with these questions:

- Will tokens be provided just in time, will tokens be regularly renewed, or will jobs be equipped with longer-lived tokens at submission time?
- What would the token scopes and audiences look like?
- How can potential scalability concerns in the provisioning of tokens be avoided or dealt with?
- How are jobs going to fall back on VOMS proxies during the transition period?

- **M.10 (March 2026): Users no longer need X.509 certificates**

To allow the user experience to be much improved in the process, several matters need to be looked into:

- Tools should be made sufficiently smart to obtain the correct tokens for specific operations.
- Auxiliary services such as HTVault [7] or MyToken [8] may be taken advantage of to simplify the user experience, used under the hood by tools for job and/or data management.

Investigations in that area are already underway within some of the experiments.

4.2 Other Developments & Focus Areas

Various IAM usability improvements are desirable, some of which will need to wait for major replacements of underlying legacy components, as presented at CHEP 2024 in *Evolving INDIGO IAM towards the next challenges* [9].

Closely related to the work of the WLCG Authorization WG is that of the Token Trust and Traceability WG, which aims to equip site administrators, VO experts, developers, and other interested parties with best practices for token usage, which will also provide input for

policy documents to be adopted by WLCG. These matters were presented at CHEP 2024 in *Preliminary findings and recommendations from the Token Trust and Traceability Working Group* [10].

Because of the current absence of a common VO attribute in the tokens used across WLCG, OSG and EGI (see 3.2), the grid job accounting software will need to obtain the VO information for each job through any of several ad-hoc methods. The LHC experiments still equip their jobs with VOMS proxies for the time being, allowing each job's VO to be determined from its proxy. A few other VOs, on the other hand, have already come to rely completely on tokens in their grid jobs, and the legacy method no longer applies to them.

5 Conclusions & Outlook

Collaborative efforts from all parties concerned and spread across several working groups will see the WLCG transition to tokens make steady progress in 2025 and beyond, with the intention to have the remaining uses of X.509 and VOMS phased out well before the start of LHC Run 4, currently planned for 2030. In the meantime, we look forward to an increasing reliance on the benefits of tokens, whilst steadily advancing in the following areas:

- Aiming to reach the next levels in data management.
- Equipping jobs for reading data and uploading results.
- Making the user experience both simpler and more secure.
- Improvements to stability, reliability, and resilience.
- Consolidation of best practices.

References

- [1] A. Ceccanti, E. Vianello, M. Caberletti, F. Giacomini, *Beyond X.509: token-based authentication and authorization for HEP* (2019), <https://doi.org/10.1051/epjconf/201921409002>
- [2] *INDIGO Identity and Access Management Service (IAM)*, <https://indigo-iam.github.io/>, <https://doi.org/10.5281/zenodo.8366226>
- [3] *Virtual Organisation Membership Service (VOMS)*, <https://italiangrid.github.io/voms>, <https://doi.org/10.5281/zenodo.1875371>
- [4] A. Withers, B. Bockelman, D. Weitzel, D. Brown, J. Gaynor, J. Basney, T. Tannenbaum, Z. Miller, *SciTokens: Capability-Based Secure Access to Remote Scientific Data*, in *Proceedings of the Practice and Experience on Advanced Research Computing (Association for Computing Machinery, New York, NY, USA, 2018)*, PEARC '18, ISBN 9781450364461, <https://doi.org/10.1145/3219104.3219135>
- [5] D. Salomoni, I. Campos, L. Gaido, J.M. de Lucas, P. Solagna, J. Gomes, L. Matyska, P. Fuhrman, M. Hardt, G. Donvito et al., *Indigo-datacloud: A data and computing platform to facilitate seamless access to e-infrastructures* (2018), <https://doi.org/10.48550/arXiv.1711.01981>
- [6] *EGI*, <https://www.egi.eu>
- [7] D. Dykstra, M. Altunay, J. Teheran, *Secure Command Line Solution for Token-based Authentication* (EPJ Web of Conferences, 2021), CHEP '21, <https://doi.org/10.1051/epjconf/202125102036>

- [8] G. Zachmann, *oidc-mytoken/server: mytoken-server 0.10.1* (2024), <https://doi.org/10.5281/zenodo.5154893>
- [9] E. Vianello, D. Marcato, D. Chung, F. Agostini, F. Giacomini, I. De Simone, J. Gasparetto, L. Bassi, M. Garai, R. Miccoli et al., *Evolving INDIGO IAM towards the next challenges*, Proceedings of the CHEP 2024 conference (to be published)
- [10] M. Doidge, D. Crooks, D. Kelsey, L. Cornwall, M. Litmaath, M. Hardt, M. Sallé, T. Dack, *Preliminary findings and recommendations from the Token Trust and Traceability Working Group*, Proceedings of the CHEP 2024 conference (to be published)