



# Paris Interop, May 2019

Overview from PADDC/VAMDC activities in RDA  
*Federation Identity Management*  
*Data Citation & Scholix*

C.M. Zwölf, N. Moreau, Y.A. Ba  
Paris Observatory



# The FIM-IG

- Federation Identity Management for research is a forum of communities to share ideas on FIM for research communities.
- FIM-4R was established prior to the birth of the RDA.
  - A first whitepaper was produced in 2012;
  - FIM-4R members agreed to join the RDA and proposed this IG.
- Coordination between FIM-IG and FIM-4R forum
  - Discussed during the 10th RDA plenary In Montreal;
  - Whitepaper v2. produced in synergy and close interaction (cf. session at plenary 11 in Berlin) and endorsed by the RDA;
- Ambitions
  - Increase the impact of the Whitepaper v2 for leveraging identity federation
  - Become a central point within the RDA for Authentication and Authorization linked issues

# Presenting the 2nd FIM Whitepaper

*Every researcher is entitled to focus on their work and not be impeded by needless obstacles nor required to understand anything about the FIM infrastructure enabling their access to research services.” FIM4R version 2*

- Data and services providers may want to grant access to resources only to a given category of users.
  - Instead of adopting «home made/specific» solutions for solving common problems, FIM aim is to promote a community-consensual approach for Authentication & Authorisation
  - Federation of identity
    - Will help data and services providers in implementing AAA strategy, since part of the work may be delegated to third parties.
    - Will help users in consuming different services by providing only one identity (avoid to identify separately on each different data-service)
- FIM V2 whitepaper is a RDA supporting output.

# Presenting the 2nd FIM Whitepaper

ABOUT RDA ▾ GET INVOLVED ▾ GROUPS ▾ RECOMMENDATIONS & OUTPUTS ▾ RDA FOR DISCIPLINES ▾ F

## Federated Identity Management for Research Collaborati

### Federated Identity Management

**Group co-chairs:** Carlo Maria Zwölf, Hannah Short

**Supporting Output title:** Federated Identity Management for Research Collaborations

**Authors:** Christopher John Atherton; Thomas Barton; Jim Basney; Daan Broeder; Alessandro Costa; Mirjam van Daalen; Stephanie Dyke; Willem Elbers; Carl-Fredrik Enell; Enrico Maria Vincenzo Fasanelli; João Fernandes; Licia Florio; Peter Gietz; David L. Groep; Matthias Bernhard Junker; Christos Kanellopoulos; David Kelsey; Philip Kershaw; Cristina Knapic; Thorsten Kollegger; Scott Koranda; Mikael Linden; Filip Marinic; Ludek Matyska; Tommi Henrik Nyrönen; Stefan Paetow; Laura A D Paglione; Sandra Parlati; Christopher Phillips; Michal Prochazka; Nicholas Rees; Hannah Short; Uros Stevanovic; Michael Tartakovsky; Gerben Venekamp; Tom Vitez; Romain Wartel; Christopher Whalen; John White; Carlo Maria Zwölf

**DOI:** [10.5281/zenodo.1296031](https://doi.org/10.5281/zenodo.1296031)

# Whitepaper V2 overview

- Cartography of the different elements in FIM and AA architecture
  - Identity providers, proxies, roles & attributes providers,...
- **Recommendations**
  - Remove interoperability barriers in eduGAIN metadata processes
  - Enable researcher mobility
  - Organizations participating in Research & Education federations should apply best practices in operational security to their federated entities
  - They should participate in security incident response frameworks such as Sirtfi and should be supported by their R&E federation operators in doing so.
  - Each R&E federation operator, and the eduGAIN operator, should have a security incident response plan
  - Follow the proxy model and related AARC guidelines
  - Re-use shared AAI and related services

# The Research Data Alliance and the Data Citation WG

VAMDC is a distributed global infrastructure for sharing A&M data

- It is a distributed set of TAP services with no central management system
- The “V” stands for Virtual in the sense that the infrastructure does not contain data
  - A wrapping software transforms existing autonomous DBs into VAMDC federated resources, accepting queries in a standard grammar and serving standardized output files.

Scientists usually cite the database, not the data producers

- Data Citation was an issue for VAMDC
- We started working with the RDA and in 2014 joined “Data Citation” WG as an early pilot

The Data Citation Recommendation comes for standalone databases and warehouses.

- From 2014 to 2017 we validated the recommendation from the conceptual point of view → No blocking point for adapting it to the VAMDC distributed infrastructure.
- Implemented during 2017, co-founded by the RDA-EU3 adoption grant.
- From the technical point of view, the VAMDC query store is a kind of smart log service.

# Query Store in a nutshell:

- The Query Store is a central real-time repository for all the queries served by any TAP service
  - Independently by the used client for querying the infrastructure
  - Queries are stored together with their metadata:
    - Production/extraction context and timestamp + references + link to the generated XSAMS file + **Unique Identifier & DOI**
  - Provides live monitoring of the entire infrastructure
    - Data providers may measure their impact and have detailed statistics of usage
  - Increases the quality of the data
    - All the produced XSAMS are parsed for extracting references & validated

The VAMDC Query Store is a technical hub for innovative development:

- It has been connected with the Zenodo open science repository. Zenodo implements Scholix and VAMDC inherit from Zenodo the Scholix capabilities
- VAMDC indirectly implemented Scholix via Zenodo.
- When you cite a VAMDC extracted data using the Query Store feature, the data producers directly receives bibliographic credits.
- Digital certification based on Blockchains technologies has been added to the Query Store (partnership with [fireblock.io](https://fireblock.io)).

# Resources

## References:

- Implementing in the VAMDC the New Paradigms for Data Citation from the Research Data Alliance Data Science Journal, 18(1), p.4. DOI: <http://doi.org/10.5334/dsj-2019-004>
- The VAMDC Portal as a Major Enabler of Atomic and Molecular Data Citation Galaxies 2018, 6(4), 105; <https://doi.org/10.3390/galaxies6040105>
- New model for datasets citation and extraction reproducibility in VAMDC Journal of Molecular Spectroscopy, Volume 327, September 2016, Pages 122-137 <https://doi.org/10.1016/j.jms.2016.04.009>
- <https://www.rd-alliance.org/improving-citation-evolving-data-distributed-asynchronous-infrastructures>

## Demos:

- <https://www.youtube.com/watch?v=js4d-b9U0II>
- <https://www.youtube.com/watch?v=OLe-qcqCcCw>

## Source code:

- <https://github.com/VAMDC/QueryStore>