Notes on the 11th FIM4R Workshop Montreal September 18.-19. 2017

Participants

Name	Email Address	Affiliation
Tom Barton	tbarton@uchicago.edu	UChicago and Internet2
Jim Basney	jbasney@illinois.edu	NCSA
Darren Boss	darren.boss@computecanada.ca	Compute Canada
Tangui Coulouarn	tangui.coulouarn@deic.dk	DeIC
Stephanie Dyke	stephanie.dyke@mcgill.ca	McGill University
Heather Flanagan	hlf@sphericalcowconsulting.com	REFEDS
Licia Florio	licia.florio@geant.org	GEANT
Patrick Fuhrmann	patrick.fuhrmann@cern.ch	DESY, dCache.org INDIGO-DataCloud
Peter Gietz	peter.gietz@daasi.de	DAASI International/DARIAH
David Groep	davidg@nikhef.nl	Nikhef National institute for s (NL)
Christos Kanellopoulos	christos.kanellopoulos@geant.org	GEANT
David Kelsey	david.kelsey@stfc.ac.uk	STFC - Rutherford Appleton
Ken Klingenstein	kjk@internet2.edu	Internet2
Scott Koranda	skoranda@gmail.com	LIGO

Marco Leonardi	m.leonardi@rheagroup.com	RHEA
Tommi Nyrönen	nyronen@csc.fi	CSC / ELIXIR
Benjamin Oshrin	benno@sphericalcowgroup.com	Spherical Cow Group
Stefan Paetow	stefan.paetow@jisc.ac.uk	Jisc
Chris Phillips	chris.phillips@canarie.ca	CANARIE Inc.
Marc Rousseau	marc.rousseau@computecanada.ca	Compute Canada
Andre Schaaff	andre.schaaff@astro.unistra.fr	UniversitÃ© de Strasbourg,
Hannah Short	hannah.short@cern.ch	CERN
Derek Simmel	dsimmel@psc.edu	Pittsburgh Supercomputing (
Matthew Viljoen	matthew.viljoen@egi.eu	EGI Foundation
Tom Vitez	tom.vitez@canarie.ca	CANARIE.ca
Chris Whalen	cwhalen@mail.nih.gov	NIH
Nancy Wilkins-Diehr	wilkinsn@sdsc.edu	San Diego Supercomputer C
Carlo Zwölf	carlo-maria.zwolf@obspm.fr	Paris Observatory - VAMDC

Notes

Intro (Dave Kelsey)

RDA Session Summaries will be given by:
Monday Morning - Hannah
Monday Afternoon - Hannah
Tuesday Morning - Peter
Tuesday Afternoon - Peter

Paris Observatory Update (Carlo Maria Zwölf)

VAMDC architecture Requirements:
works with scripts (GUI should just be a frontend to that)
Considering a scientific identity hub, linking identities
Not using eduGAIN - how is LoA obtained?
Main requirement = easy to use, eduGAIN is not chief “IdP”
Proxy model
Identity should not be a problem for a scientific service - happy to outsource
Want identities to be linked back to identity source services, e.g. store data back on Google Drive
Size of organisation - 15 research institutes some technical resources in each, + 4 people centrally (but time is dedicated to scientific services). Thousands of users.
Being a consortium (MoU) may cause pain at some stage- technical services should be available to non-legal-entities.
Could one of the MoU signatories be agreed represent other legally? Probably not.
- Not others, but the consortium as a whole, yes, provided the mandate exists from the MoU steering (management) committee.
VAMDC is a scientific service provider willing to trust identity (and authorisation) management to be provided for them as a service

LIGO (Scott Koranda)

Exciting times in astrophysics (black hole detection)
Ligo will introduce a proxy following the AARC blueprint architecture. The added value of a proxy is that it offers a central place to manage heterogeneous information coming via eduGAIN/IdPs. This also signals a shift of paradigm: instead than asking eduGAIN and ID feds to provide more attributes, LIGO takes the view to ask edUGAIN to provide the basic info; support for RS, Sirtfi, etc. is then added via the proxy.

Proxy is for < 100 SAML SPs in the global collaboration. SIRTFI adoption will be critical -- IdPs will not show in the service provider service unless they are compatible with SIRTFI. Proxy reviews if the connected SAML IdPs are compatible with SIRTFI. When this is enforced, a lot of user feedback may be due.

Interest in looking at tools like SaToSa.

DARIAH (Peter Gietz)

Dariah services have 4080 users (local user database) divided into 300 user groups increasing.
The AARC Blueprint Architecture is being followed by implementing an SP-Proxy to hide the complexity (attribute aggregation, checking for Terms of Use, etc.) from the services, that now can use a plain vanilla SP.
New work on Accounting is on the way, where usage metering will be done on the service side, but will have to be aggregated centrally. Challenge is how to integrate accounting into PDP e.g. if the quota authorised for a user has been exceeded.

UmbrellaID & CALIPSOplus (Stefan Paetow)

UmbrellaID started in PaNdata, does not leverage eduGAIN but eduGAIN bridge pilot developed that should be taken forward along with Moonshot IdP support for PaN facilities that need it. Funding gap but facilities told they would have to fund with Time&Manpower, but now part of CALIPSOplus where the aim is to become sustainable. As of 2017, Umbrella ID is the exclusive IdP for the PSI industrial-science collaboration, collaboration wants to see more use of it (in addition to others). Suggestion by Tommi that ELIXIR would be interested in using Umbrella ID as an additional ID source. Pilot with eduTEAMS to investigate use of eduTEAMS as AA for Umbrella ID (for things like ORCID and other attributes that might be useful under R&S and eduGAIN use). CALIPSO+ project has kicked off with JRA2 and NA1.

WLCG (Hannah Short)

PKI based grid technology Globus Toolkit will from now on be supported only on the base of community effort, which on the long term will be an issue.

WLCG will also follow AARC Blueprint Architecture using a proxy, leveraging CERN's infrastructure. They are also considering replacing ADFS - however there are a lot of requirements for that replacement (including kerberos).
Authorization still based on VOMS but alternatives are looked at
Token Translation pilot based on STS, WATTS (the INDIGO Token Translation Service) could be an alternative.
CERN has a strong requirement for delegation (need to provide ways for software to act on behalf of users). There are technologies to do that, OIDC is not one of them at the moment.
Suggested approaches build on RCAuth and WATTS.
Data protection is a special issue for CERN given their status (not being in the EU).
Assurance - multifactor expected to happen in house.

Plans are to revamp the authorisation, but no clear decisions yet. WLCG Authorization working group started this July.

ELIXIR & GA4GH (Stephanie Dyke (McGill.ca), Tommi Nyronen (CSC))

Elixir compute platform becomes more important
Human genome data are very sensitive, thus committee decision based access control
Elixir AAI also has commercial relying cloud service providers (e.g. IBM, SixSq)

ProxyIdP, central directory, group/role management (PERUN), attribute self-management and integration with eduGAIN and common social IdPs including ORCID are in production.
Integration of REMS for Data Access Committee access approval is on its way

In future consolidating AAI with the whole life science community within ESFRI cluster project CORBEL, where AAI is no core business for any BMS ESFRI, but required component in all. A pilot is currently worked on in the frame of AARC2.

GA4GH Global Alliance for Genomics & Health, aiming at a common framework, with > 475 member organisations

Data sharing policy: either complete free Open access or very restrictive committee based controlled access. In between there will be FIM based registered access, with identity proofing

NIH (Chris Whalen)

Project on diseases in Africa, South Asia and China
Collection of data, e.g. on TB
First AAI approach was fIM based, but enrolled only 17% of users
Thus an own IdM was set up disconnecting from eduGAIN

Identity aggregation was needed because a lot of researchers switch jobs, but not the project

IdP of Last Resort must not use google authenticator because that does not work in China

New federation development based on RENU (Uganda), CARSI (China), SAFIRE (South Africa) and INFLIBNET (India)

NIAID VRO Discovery was problematic due to name conflicts now includes previous choice, common choices and only if these two don’t have the right IdP, “find my login server”.

Requirements Gathering

https://docs.google.com/spreadsheets/d/125KRVnZ16ZKUhwlHB7MFgCcxUv6Ywsr_2ag-a7L63rM/edit?usp=sharing

Research Community Drivers

AARC has community engagement forum, input welcome for training courses.

How do we want to spin FIM4Rv2 so that it has maximum impact?

- AARC could support pilots for clearly defined requirements
- GEANT input
- Internet2 welcomes input on research community needs

We should

- qualify how well we have satisfied the FIM4Rv1 requirements
- provide metrics of what “success” means

AARC (Christos Kanellopoulos (GÉANT) , Dave Kelsey (STFC - Rutherford Appleton Lab. (GB)) , Licia Florio (GEANT))

AARC Blueprint architecture gives a holistic view of the AARC endorsed approach. There are currently 4 reference implementations that can give tips on which components to use (although there is not yet a component catalogue). eduTeams is a reference implementation.

Snctfi defines a policy framework to ensure that all elements behind the proxy behave in the desired way and that security best practices are followed. Snctifi framework has been developed within the AARC project. In the future (when AARC will finish) Snctifi will be managed by the IGTF group. . EGI (and WLCG) policies to be updated to comply with Snctfi.

AARC also contributed effort to Sirtfi. The group originates in SCI and evolved into a REFEDS WG. Sirtfi defines a framework to handle security incidents in entities particiapting in eduGAIN.

FIM4R communities asked to try applying Snctfi to our infrastructures/communities.

The Blueprint Architecture solves defines a meta layer; there are a different implementations of the blue print architecture. The AARC Engagement Group for Infrastructures (AEGIS), recently created,wants to increase collaboration and interoperation of these

implementations, with participants from EUDAT, EGI, Elixir, Xsede, GÉANT and PRACE, open only for infrastructure operators.

Canadian Projects (Chris Phillips (CANARIE))

CANARIE besides network provides services on identity, cloud and research software

Identity federation has 46 IdPs (having 92 Universities and a couple of 100 colleges), all in eduGAIN. Next milestone is to have all IdPs support R&S entity category

Large interest in Service Catalogue.

Aims at simplifying Installation and improving SAML software guidance. Perhaps FIM4R can influence technology providers?
Assessing currently used IdP Software in the federation: Shib, SimpleSAMLphp and ADFS
Future areas: non-web SSO, discovery, sirtfi and authorization services, which all are also of interest in FIM4R

US Projects (Jim Basney (University of Illinois), Scott Koranda (Ligo))

New IdPs of Last Resort, for R&S, Sirtfi, ECP, MFA (but does have Google recaptcha)
ECP is used in the US, must not be forgotten.
ORCID seems to have a big uptake, some sites have ORCID as only possibility for signing in. Although it is not organisational provided, the scholars themselves take it very seriously (career ID).
OAuth/OIDC for authentication and authorization also gains importance

New CTSC call on IAM consulting with dead-line Okt 2.

New collaborations leveraging FIM: Murchison Widefield Array, Open Storage Research Infrastructure (OSIRIS) and Humanities Commons (10000+ users

UK Project Announcement (Stefan Paetow)

Long history of trying to get moonshot on Mac OS X, although a very Big Apple development firm took it up, Apple did (and does) not support it, but now workarounds were found, so now it works on Sierra and Capitan.

HNSciCloud (Hannah Short)

IaaS cloud service from different commercial cloud providers for different communities (e.g. physics, life sciences), with AAI based on SAML. Internal generation of tokens for API access. Supporting both Elixir & eduGain.

SP is troubled with joining eduGain, different attributes to map for each IdP. Compliance with R&S attribute bundle by IdPs is important and authorization decisions are to be made locally in the cloud. One issue is that some SPs expect user blocking by the IdP - conformance with Sirtfi important for an SP to be able to contain an incident.

Question whether there should have been (or should be) a HNSciCloud Procurers Proxy set up to shield the SPs from eduGAIN's complexity. However several factors influence choice:

- The cloud services should be available to the community in general, after the HNSciCloud project has finished
- All development work is expected to be managed by the cloud providers
- Tenant agreements are handled by the cloud providers,
- A proxy per community could make sense but would require the same level of custom integration

The current prototypes are in line with the AARC blueprint architecture, all components managed by the cloud providers (Q whether the AARC blueprint architecture adds much value as is very generic and flexible). Perhaps there should be a proxy involved, complete with HNSciCloud Attribute Authority, that the cloud provider can plug into. However, when is this AARC model appropriate, and when is a service just an SP?

The amount of testing that is required for each cloud service connecting to eduGAIN is a huge task. This should be streamlined by tools from technology providers - an area to highlight in FIM4Rv2. Actually there already is a whole set of tools to test eduGAIN. They are now listed on this page: https://wiki.geant.org/display/eduGAIN/Tools+and+Services

EGI Check-in (Matthew Viljoen)

EGI had a great history in using X.509, but SAML and OIDC is now part of the AAI strategy. Check-in is an AAI proxy (IdP proxy and SP proxy). Development lead by GRNet, with features such as authZ based on VO/group, attribute aggregation, account linking, LoA and interoperability. Architecture conforms to the AARC Blueprint Architecture. Can be used to integrate other research infrastructures such as Elixir. But also a full AAI platform as a solution for communities that do not have AAI infrastructure. It can also function as Authentication proxy.
Supports different LoA schemes.
In production (as theoretical Beta version) since almost a year.

Question discussed who and how many communities will operate a proxy service like Check-in. It should be a limited number of big and sustainable organisations such as EGI and CERN.

FIM4R All Communities Survey

Survey is for reaching out to research communities (not federation operators, not single institutions and IdP operators) not yet participating in FIM4R. Current version is https://docs.google.com/document/d/1zwthTAL4-wkWlWv8c3neFQYDH3A0kq3h6eKW4P7N7wE/edit

Which communities shall be targeted?

- All 29 ESRI projects
- Activities in the Asia-Pacific
- List of EU projects at CORDIS
- Different domains, such as Humanities, Chemistry, Botany, Radio Astronomy
- Communities connected to TAC
- RDA has a lot of involved research communities

How should the survey be done: just send it out, do interviews?

- The more target groups the less will be single one hour plus interviews with each
- So may be first approach to just send the survey and offer help and interviews as fallback
- The introduction and / or email text should show how influential FIM4R can be and that the mentioned requirements in the answers will be, e.g. taken care of in future EU projects
- Piggyback on GEANT team that is already engaging with most of the ESFRI projects
- Start approaching the AARC communities participating in the pilots (including those rejected).

May be an alternative would be that communities just provide a small text (such as in FIM4R paper V1) on what they are doing now and where they want to be in future.

FIM4R v 2

What is the right order (survey and next version of the FIM4R paper)?

- Paper V 2.0
- Survey with sending V 2.0
- Paper V2.1 including new requirements of the new communities

Paper should first contain a narrative on what already has been achieved since V1 of the paper, and then to describe which requirements have not yet been met and which new requirements have popped up since then. Top ten priorities.

The paper should summarize the common requirements all participating communities agree to.

We should work on:

1. A table on attributes of the single communities
2. Success story on what we have achieved
3. A text on new recommendations to
    a. Research infrastructures
    b. Federation providers
    c. Funders

Funders will listen, if it fits into their frameworks and action items

Starting point will be the post-its of this meeting categorised in groups

Deadlines:
Good draft by beginning of December 2017
Final Version of Version 2.0 in March 2018

A next FIM4R face to face could be

February at TIIME in Vienna Feb. 7-8, expecting especially life sciences AAI stakeholders (ELIXIR, BBMRI-ERIC (biobanking), Life Sciences to host if needed (Tommi) (Austria)

Other possible meeting points:

- Chicago offer by Tom is still open
- I2 TechEX17 will be held in San Francisco, CA, October 15–18, 2017.
- NSF Washington August 2018
- PERK in July in Pittsburgh?

Program officers in DC

Places to evangelize:

- TNC?
- Conference in Taipei in Feb 2017?
- APAN?
- GA4GH https://genomicsandhealth.org/working-groups/security-working-group