FIM4R

Presenting the 2nd Whitepaper, TNC18

With thanks to the FIM4R Authors and Contributors for their collaboration on the whitepaper and the following slides.

*** Not all contributors' logos represented*

Agenda

Topic	Speaker	Duration
Introduction & Motivation	Hannah Short (CERN/WLCG)	10m
Progress since 2012	David Kelsey (STFC-RAL)	15m
Research Community Use Case Highlights	Scott Koranda (LIGO)	15m
Common Requirements	Scott Koranda (LIGO)	10m
Recommendations	Peter Gietz (DAASI International/DARIAH)	15m
Next Steps & Discussion	Hannah Short (CERN/WLCG)	20m

Introduction & Motivation

Motivation

- Research Communities provide complex use cases where federated identity can be leveraged to great effect
  - Distributed users
  - Distributed services
- The specific way of working brings specific challenges that go beyond the functionality typically offered by federations and interfederation (more to follow from Scott!)

FIM4R provides a forum for Research Community representatives to exchange experiences of implementing AAI and by combining our voices we hope to influence the future direction of FIM in a way that meets the needs of our Users

"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**

2012: FIM4R version 1

- Published a whitepaper in 2012 that guided the direction of identity federation for research
https://fim4r.org/documents/
- Specified a common vision together with common requirements and recommendations
- Revised (just to specify priorities) in 2013

2018: FIM4R version 2

- In early 2017 FIM4R decided to start work on a Version 2 paper
- 5 years on, much had changed & time to review progress
- The FIM4R Community was called to action once more…

https://www.lonelyplanet.com/austria/vienna

2018: FIM4R version 2

- Representatives of more than 20 research communities have provided input
- Five face to face meetings in Europe and North America
  - Vienna (TIIME) x 2
  - Montreal (RDA)
  - Barcelona (RDA)
  - San Francisco (Internet2 TechEx)
- A new distillation of specific requirements and a set of recommendations is the result of this process

Dept. of Physics, McGill University, Montreal (Sep 2017)

Who is represented?

Research Fields (14)

- Arts and Humanities
- Climate Science
- Earth Observation
- European Neutron and Photon Facilities
- Gamma-Ray Astronomy
- Gravitational Wave Astronomy
- High Energy Physics
- Ionospheric and Atmospheric Science
- Infectious Disease Research
- Life Sciences
- Linguistics
- Nuclear Physics
- Radio Astronomy
- Virtual Atomic and Molecular Data Centre

Others

Research Driven Services

- HNSciCloud
- ORCID

Identity Federation Projects/Communities

- AARC(2)
- GÉANT-GN4
- InCommon/Internet2
- REFEDS

Progress since 2012

Evolution of FIM

Aim: "Authenticate locally, authorize globally"

- deliver good user experience
- extend the reach of campus credentials
- reduce the number of credentials needed by users
- leverage campus credential management practices
  - focus more energy on their own services
- make federation a global infrastructure
  - academic collaboration, itself global, benefits

In a very substantial sense, this has been achieved (for Authentication)

- FIM now spans the national R&E federations from many countries into an aggregated whole known as eduGAIN

The more difficult part

"Authorise Globally" is more difficult (as identified in FIM4R v1)

- Access control/Authorisation to Research SPs
  - perform checks of the attributes assigned to the user
  - user needs to get those attributes from somewhere
- Initially just a singular infrastructure
  - that operated by the national R&E federations
- But the campus IDP is not authoritative for research attributes

Concept of FIM gradually began to evolve

- Social ID gateways admitted users without waiting on their campus
- Credential providers for unaffiliated users started to pop up
- Research e--Infrastructures experimented with Proxies

The Research Community SP/IdP Proxy
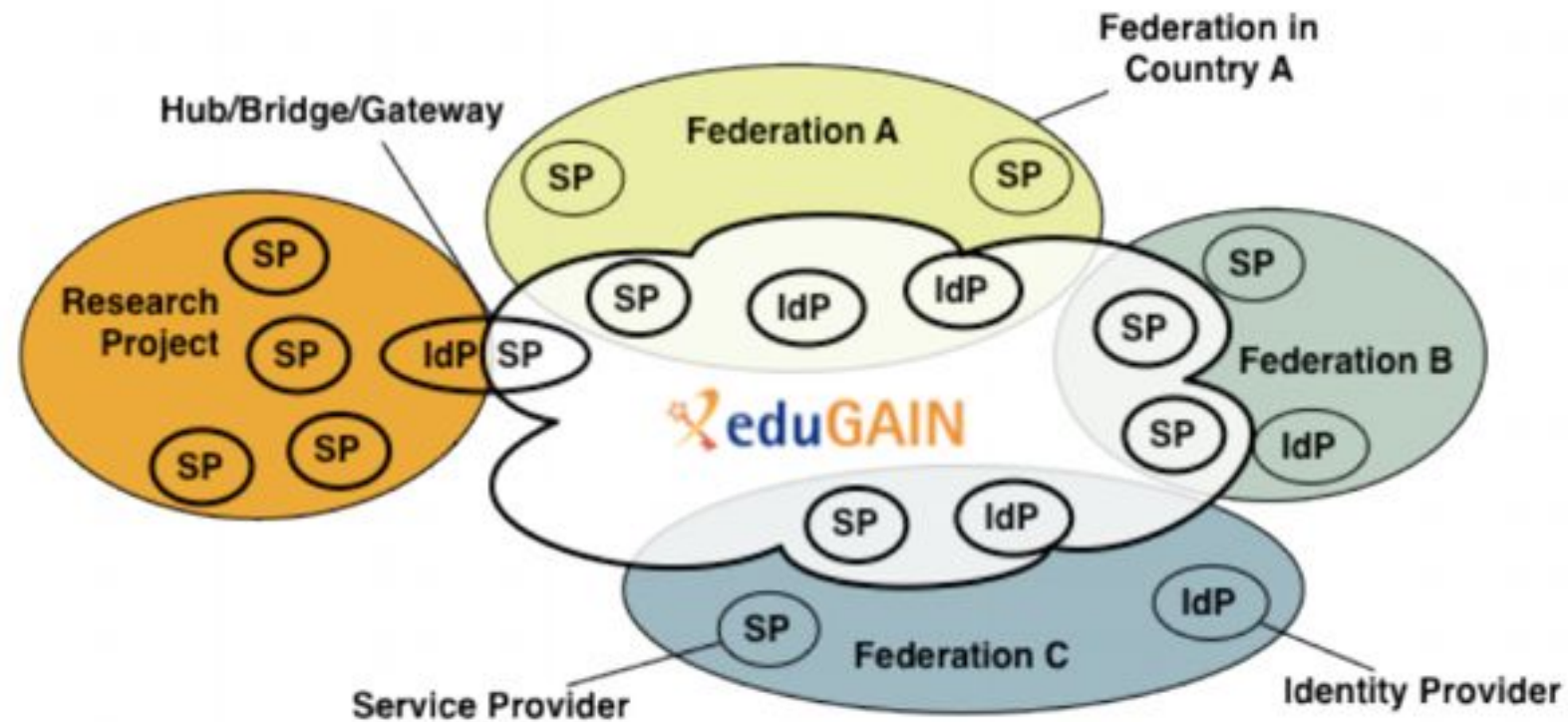
What are the successes since 2012?

Much has changed since 2012. The FIM4R paper was taken seriously and AAI is now more mature with many successes

- European Commission funding (H2020) for the AARC/AARC2 projects
- Federations and interfederation have found their role as an Authentication infrastructure; Authorisation managed by Communities
- eduGAIN's operational support capability is now in place and maturing
- e-Infrastructures are deploying shared AAI services (EGI, EUDAT, GÉANT, EOSC-hub, …)
- Specific successes include the Sirtfi and Snctfi trust frameworks
- Standardisation and best practices for the emerging trend of "proxy" architecture (The AARC Blue Print Architecture)
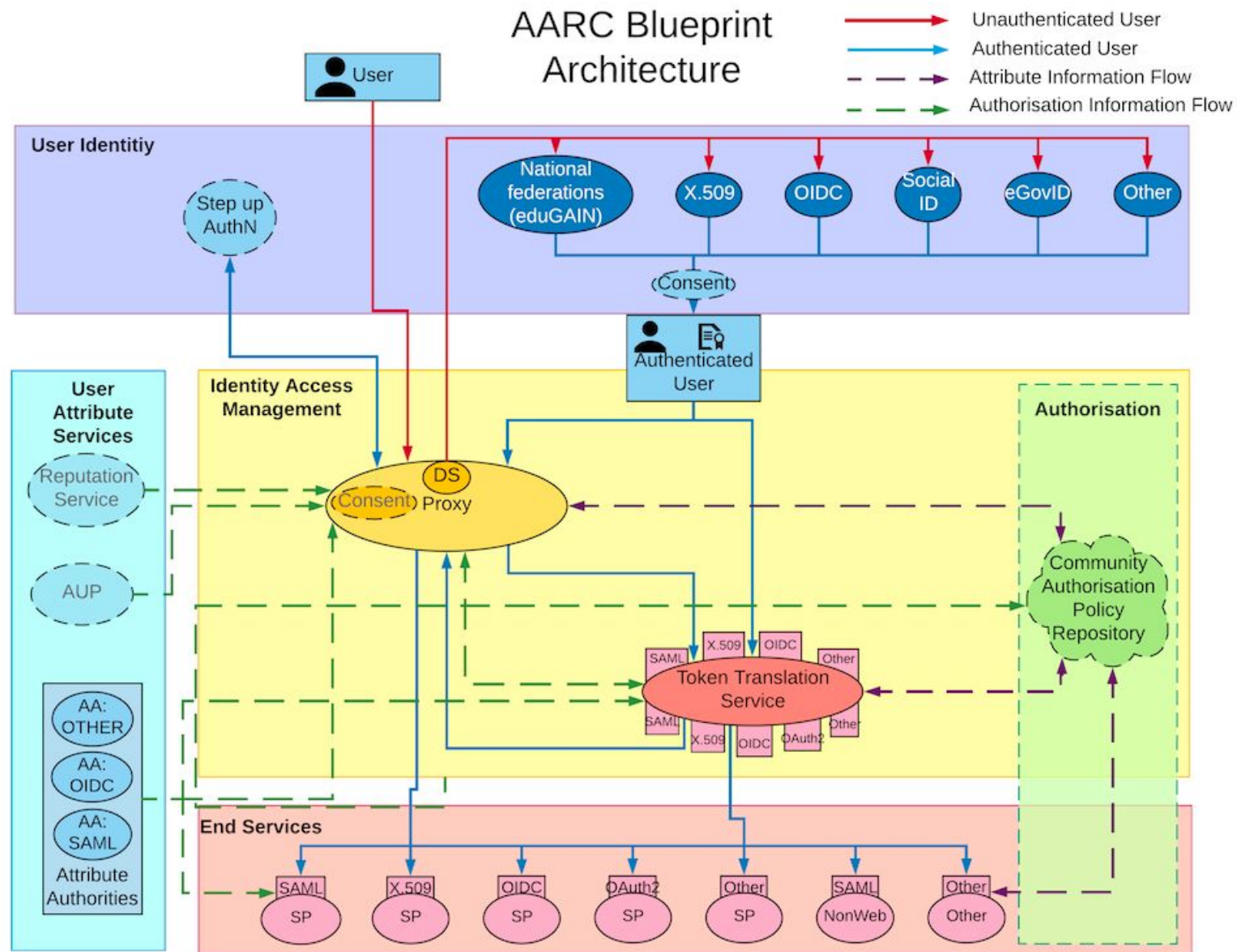
The Proxy Architecture allows Research Communities to satisfy many FIM4Rv1 requirements themselves, but comes with a large overhead

https://aarc-project.eu/architecture/

Outstanding Challenges (1)

- Usability & User Experience
  - Considerable time spent in bilateral debugging
  - Discovery services unintuitive
- Data Privacy and EU GDPR
  - Better data access and privacy expectations need to be balanced, E.g. ELIXIR Human Data resources are potentially liable for breaches
  - Attribute release by risk averse IdPs is already problematic, this may be aggravated
- Level of Assurance
  - Frameworks defined but not well adopted or propagated

Outstanding Challenges (2)

- Interfederation and many federations do not offer an adequate level of operational support and security for Research use cases
  - eduGAIN has become recognised as a brand and those wishing to "join" experience a steep learning curve and inconsistent federation practices
- Although several generic AAI infrastructures are evolving, their respective advantages and availability are unclear

Research Community Use Case Highlights

Our old friends...

World-wide LHC Computing Grid Project (WLCG)/CERN

- 13K physicists, analyse Large Hadron Collider (LHC) data
- Long history of federated X.509 through IGTF
- Non-X.509 increasingly interesting including eduGAIN and ORCID
- "The definition and adoption of Trust Frameworks, including Sirtfi, has built confidence in the potential for such identities to demonstrate an acceptable level of assurance"
- Several WLCG web services have already been integrated with CERN's IdP-SP Proxy
- Working group has been established that will address non-web use cases
- "An increased level of speed and flexibility to enable the adoption of community frameworks by identity federation participants. The inclusion of additional sources of authoritative information beyond federation registries is likely to support this requirement."

"Life Sciences"

- A common Authentication and Authorisation Infrastructure (AAI) to allow SSO across BioMedical Science European Strategy Forum Research Infrastructure (BMS ESFRI) identified as strategic target in the community
- A concept of a Life Science AAI has been developed based on existing AAI implementations (e.g. ELIXIR AAI, BBMRI AAI and Instruct ARIA)
- Will be tested using the service portfolio for the research infrastructures participating in the biological and medical science ESFRI cluster (CORBEL) and in the upcoming Life Sciences cluster in the European Open Science Cloud
- AAI requirements developed with CORBEL, will be implemented in the AARC2
- SPs relying on ELIXIR AAI benefit from a centralised user identity and access management services
  - Growth rate for users and services is rapid
  - Supported protocols are SAML2 and OIDC

EU Neutron and Photon Facilities

- UmbrellaID
  - single-sign-on system managed collaboratively by 16 photon and neutron sources around Europe
- Users of the umbrella ID is steadily growing at around 20% per year.
- Umbrella participates in eduGAIN and projects such as GÉANT's eduTEAMS and the EC-funded AARC and AARC2
- A crucial part in this system is non-web-based access...realized with the Moonshot technology

- Digital Research Infrastructure for the Arts and Humanities
  - "support virtual research environments in the humanities, and thus furthers the digital humanities"
- 5 years in production as mesh federation
- DARIAH guest IdP operated by DARIAH-DE
- Shift to proxy based on AARC Blueprint Architecture
  - Easier integration path and lower barriers for SPs
  - interoperability with eInfrastructures like EGI, EUDAT and in future EOSC

CLARIN Service Provider Fed

- Common Language Resources and Technology Infrastructure (CLARIN)
- Participating national identity federations (IDFs) grown to 18 in mid 2017
  - Enables SSO for users from 1577 organisations to CLARIN SPs
- "The attribute release issue is not as big as it is sometimes pictured"
  - "The problem varies widely among the national federations, with Germany being the most problematic case…"
- Need for homeless IDP still exists today
  - CLARIN ERIC is operating this homeless IDP for the CLARIN SPF
  - Best-effort basis and accounts have to be verified on an individual basis
  - Technical effort required to operate is quite low
  - Administrative verification and end-user support is much higher

- "Operated...around the world for the purpose of disseminating model data and observations relating to the Earth system"
- Biggest challenge had been support for non-browser and bulk data download
  - First solution based on SSL client auth and short-lived user X.509 certificates
  - problematic and complex for many users
- Pivot towards all data publicly available
  - Less dissemination, aggregate at computing centers
  - Manage access to computing rather than data
- Oauth2 credential translation to X.509 for bridging to legacy infrastructure
  - Hoping to migrate to OIDC for SSO
  - "number of identity providers...reduced recognising that only a smaller set of core institutions are able to commit to the long-term support of such services"

Our new friends…

Cherenkov Telescope Array (CTA)

- Gamma-Ray Astronomy
- CTA Consortium > 1000 scientists and engineers in more than 160 institutions from 27 countries
- SAML-based AAI infrastructure
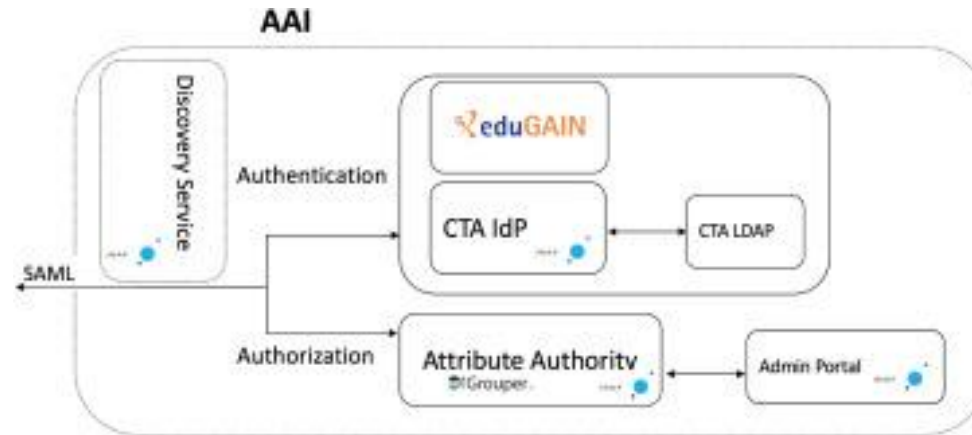
gw-astronomy.org

- Collaboration hub for gravitational-wave and multi-messenger astronomy (MMA)
- Federated identity collaboration registry (COmanage), wiki, and email list server
- Used to manage collaboration around the August 17, 2017 kilonova event (binary neutron star inspiral with gravitational-wave and electromagnetic counterpart)
- Mesh configuration but beginning transition to proxy architecture
  - Both SAML (SATOSA) and OIDC (CILogon)

European Southern Observatory
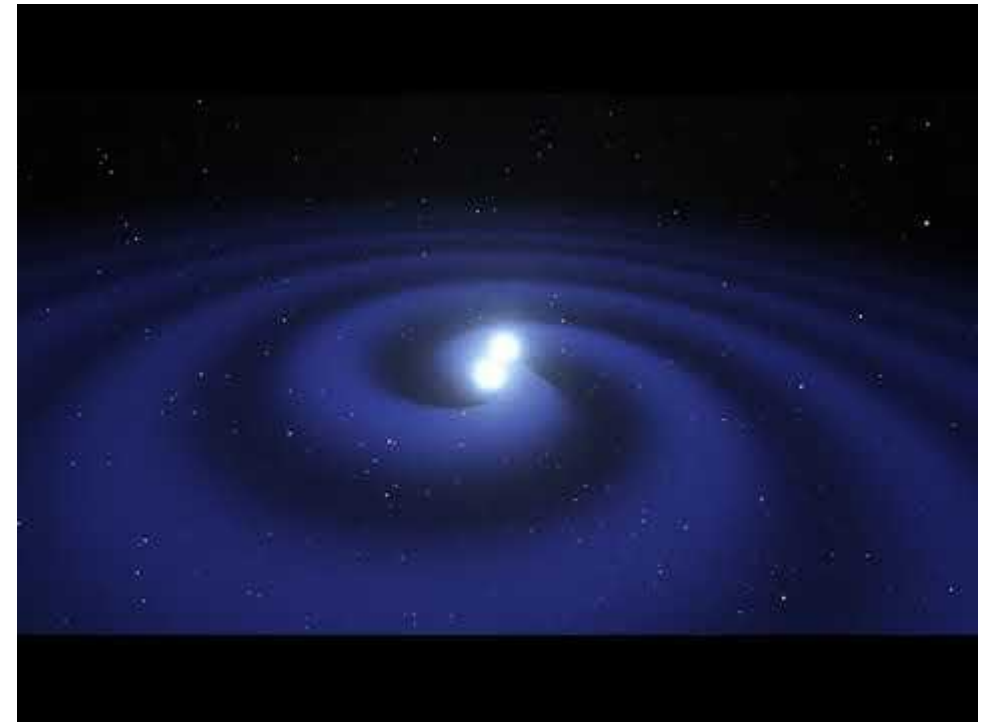
NIAID VRO Platform

- National Institute of Allergy and Infectious Diseases Virtual Research Organization
- US National Institutes of Health (NIH)
- Support International Centers for Excellence in Research (ICERs)
  - Bamako, Mali; Entebbe, Uganda; and Chennai, India
  - Collaboration hubs for global research communities investigating the pathogens endemic in those regions
- Federated through eduGAIN but face challenges:
  - Many users with no access to federated credential, multiple guest IdP required
  - IdPs send targeted identifier so move to AARC proxy architecture
  - Wide spectrum of downstream services including commercial, eg. Slack, GitHub

Murchison Widefield Array (MWA)

- Low frequency radio telescope in Western Australia
- International collaboration
  - Approximately 400 members
  - Australia, Canada, New Zealand, the United States, and more recently China
- Recently deployed proxy-centric federated infrastructure
- SP metadata published in the Australian Access Federation (AAF) by Curtin University and exported to eduGAIN
- Primary issue is federated access for Chinese scientists
  - Work to help push federation in China
  - Need for guest IdP (not social) that does not use Google Captcha

Helix Nebula Science Cloud Pre-Procurement Project

- €5.3 million Pre-Commercial Procurement tender for EU hybrid cloud platform
  - support deployment of HPC and big-data capabilities for scientific research
- Challenged to enable access via SAML/eduGAIN and become SP in ELIXIR AAI
- WLCG is key use case, represented through CERN as a procurer
- Issues:
  - R&S bundle use not clear
  - Lack of central automated testing for SPs against IdPs
  - Lack of central point of contact for operational support
  - Lack of structured documentation for SP onboarding, service best practices
  - No native command-line option, need to be connected to complex AAI that follows the AARC blueprint, but is not suitable or feasible for some use cases

Our future friends…

Earth Observation (ESA)

- "Bring the numerous and largely disparate EO datasets into a federated layer of...platforms"
- "Bring the user to the data"
- "Ensure these approaches are in-line with the AARC Blueprint Architecture and the technical practises in eduGAIN"
- "Plans to establish a Commercial Operator Identity Hub which shall provide federated identities to all EO users whose home organisations are not able to federate in any way"
- ESA as intergovernmental organisation has found it difficult to coherently manage SP federation through eduGAIN since no NREN can act as a central channel for all ESA sites located in various european countries

EISCAT_3D

- Ionospheric and Atmospheric Science
- Distributed phased array radars will enable comprehensive 3D vector observations of atmosphere and ionosphere above Northern Fenno-Scandinavia
- Nordic countries (FI,NO,SE), UK, Japan, China, France, Ukraine and South Korea
- EISCAT data used world-wide through collaboration with EISCAT scientists
- "EISCAT does not anticipate the budget nor possess the expertise to manage credentials or virtual organizations"
- "Users must not be exposed to low-level credential handling (x.509 certificate/key or browser certificates etc)"
- "...easiest mechanism for EISCAT_3D users to authenticate and be authorized is to use their current university/institutional login... must be minimally compatible with the national federations (SWAMID,Feide,Haka) of the participating countries and, considering the participation from outside the Nordics, eduGAIN"

FAIR

- Facility for Antiproton and Ion Research
- Currently under construction
- Shareholders: DE, FI, FR, IN, PL, RO, RU, SE, SI, UK
- 50 countries involved in FAIR science program
- About 2500 users
- To date scientific communities of FAIR accesses the computing infrastructure through local assigned credentials
- Some services make use of X.509 certificates issued by IGTF CAs
- Prototypes of SAML federation, based on the DFN-AAI, are in testing

INFN Gran Sasso National Laboratory (LNGS)

- Largest underground laboratory devoted to neutrino and astroparticle physics
- To date the scientific community of LNGS accesses the computing infrastructure (web- and non-web services) through local assigned credentials
- Use of federated identities is requirement coming scientific community
- "Physical access to the LUNA-MV infrastructure could benefit from federated authentication system, if participant's home Institution can provide in a secure way, also metadata/attributes stating the values of enabling factors (eg. validity of radioprotection sheet, etc) and their corresponding Level of Assurance, as defined in a commonly approved framework."

INFN LABORATORI NAZIONALI DEL GRAN SASSO

Square Kilometre Array (SKA)

- Largest Radio Telescope ever designed
- Dishes and Antennas spread over two continents (Australia and South Africa) with the Headquarters located in Europe (Jodrell Bank)
- User access point will be the SKA Observatory user portal
- SKA requires the authentication system should be flexible, standards-based and allow identities to be "joined" to other external accounts
- Prototype for the Authentication and Authorization system has been implemented
  - Includes federated users, both for visualization and access to SKA resources as well as for deep computation and management of the cloud infrastructures
- Provide the users with the major authentication mechanisms (SAML, OIDC, certificates, Kerberos, whatever) at their choice

Virtual Atomic and Molecular Data Centre

- 41 heterogeneous atomic and molecular databases in an interoperable way
- VAMDC provides middleware, called Node Software, for transforming an autonomous database into a VAMDC node
- Each node understands queries and provides result formatted using the XSAMS standard (XML schema for Atoms, Molecules and Solids
- VAMDC is distributed setting, with single point of access available but no central management of the data or the infrastructure
- Do not want to manage passwords, registration, lifecycle of users
- Find federated identity and SSO attractive for user experience
- Waiting for consensus of the community on which authentication, authorization, and accounting solution to pursue

Common Requirements

Requirements Collection

- Began during 11th FIM4R Workshop in September 2017 in Montreal
- 28 individuals, presentations by research communities and cyber infrastructures
- Pooled thoughts on the outstanding requirements
    - Integrated into matrix
    - Established initial categories
    - Produced initial mapping of requirements to constituencies needed to address them
    - Identified requirements for which something entirely new needed
- Process continued at I2 TechX in October 2017 in San Francisco
- 12th FIM4R meeting in Vienna in February 2018 performed final review & identified recommendations considered to be most important

Requirements Categories

- Identity Lifecycle & Linking (2)
- Discoverability & Usability (3)
- Authorisation and (De-)/Provisioning (8)
- Attribute Release (4)
- Security Incident Response (4)
- Research Community Proxies (4)
- Assurance & Multi Factor Authentication (2)
- Interoperability (3)
- Non Web (4)
- On-Boarding & Support (5)
- Sustaining Critical Infrastructure (3)

Total: 42

FIM Ecosystem Constituents

Identified specific constituencies within the FIM ecosystem who, it is believed, can best address each of the requirements that were gathered
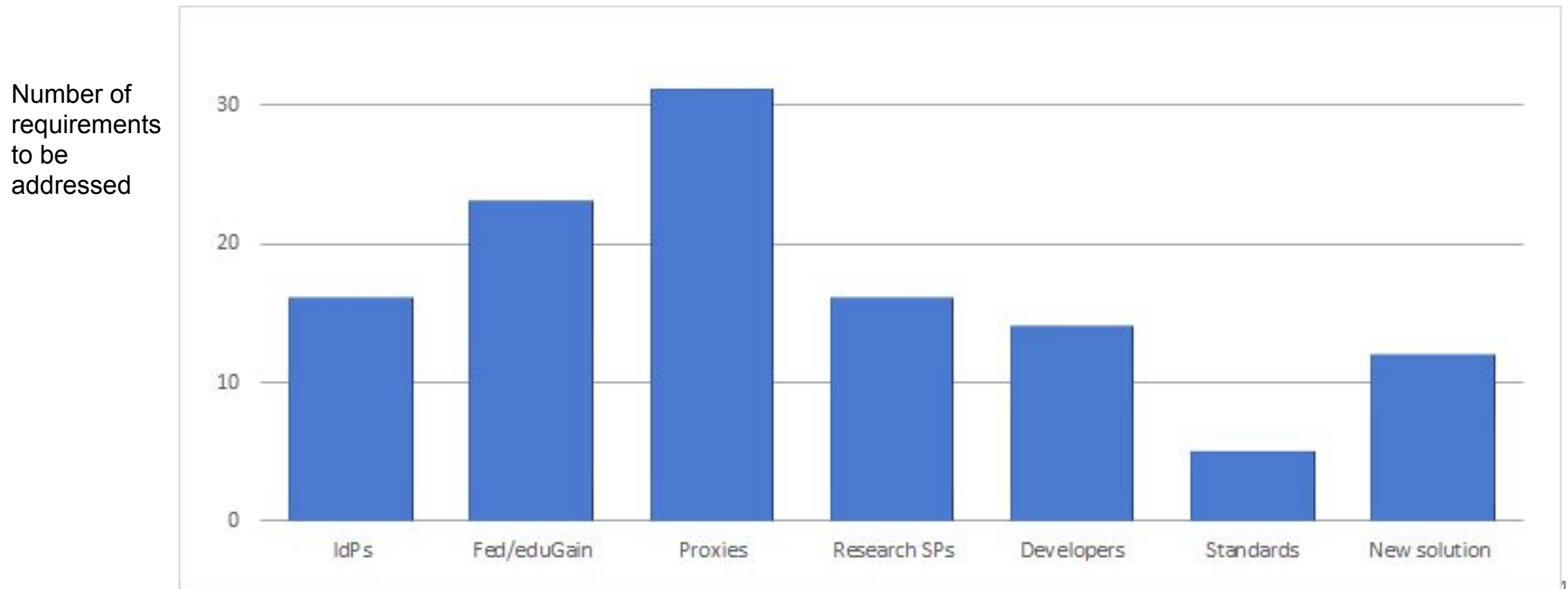
Relative Importance of the Requirements

- February 2018 FIM4R participants asked to indicate which were of particular importance to them, as one indicator of breadth of impact that implementation of each requirement might be expected to have
- Bar chart shows the number of different research communities so indicating, by each requirement
- This "vote" reflects a point in time for those research communities present at that meeting, representing less than half of those taking part in data gathered for this paper
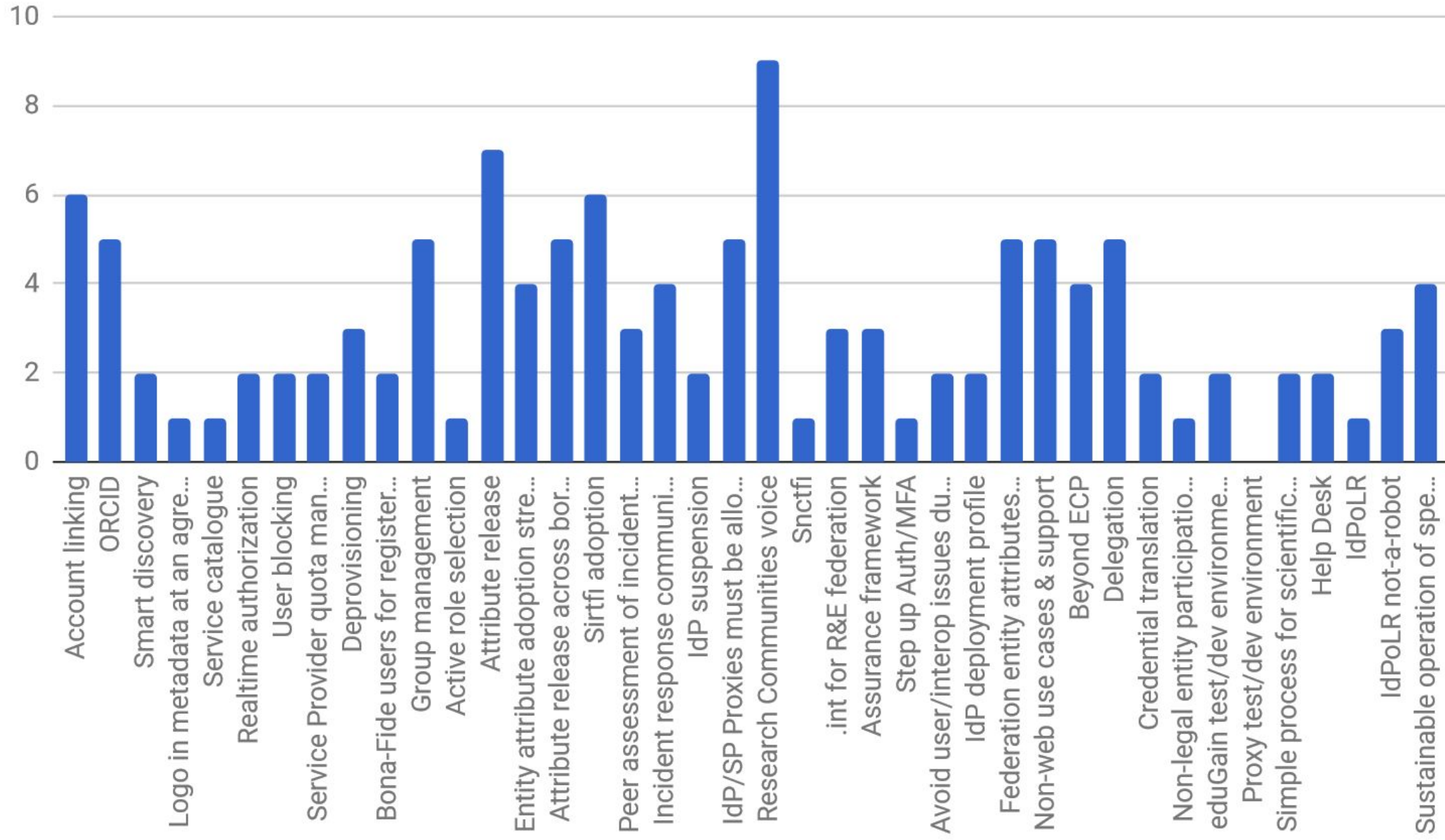
Relative Importance of the Requirements

- Top two reflect keen interest in research communities having direct representation in federation governance and in release of R&S attributes by identity providers
- Sirtfi adoption, likewise reflects how critical identity and federation operations are to research communities
- ORCID reflects increasing uptake of ORCID both as an identifier that researchers are now expected to have and as an identity provider for researchers who may have insufficient, or absent, identity provider support by their home organisation

Recommendations

Vision of future FIM

- Based on the requirements collected, we know where we want to go:
  - an interoperable infrastructure that:
    - solves current issues
    - is as invisible to the researcher as possible
- All depends upon the substantial participation and cooperation of several groups
- In our recommendations we specifically address each of these groups
- Each recommendation is expected to have broad positive impact on the viability and value of FIM for research collaborations

Nine Stakeholder Groups to address

- General Stakeholders
  - Network coordinators and operators: GÉANT (Europe), Internet2 (US)
  - Research funding bodies
  - REFEDS (Research and Education FEDerations group)
- Identity federation operators
  - Researcher's Home organisations
  - National R&E federations
  - eduGAIN operators providing the Interfederation
- Research
  - Research e-infrastructures in general
  - Research community proxies in particular
  - Research communities

Network coordinators

Funding bodies

REFEDS

Home organisations

R&E federations

eduGAIN operators

Research Infrastructures

Community proxies

Research Community
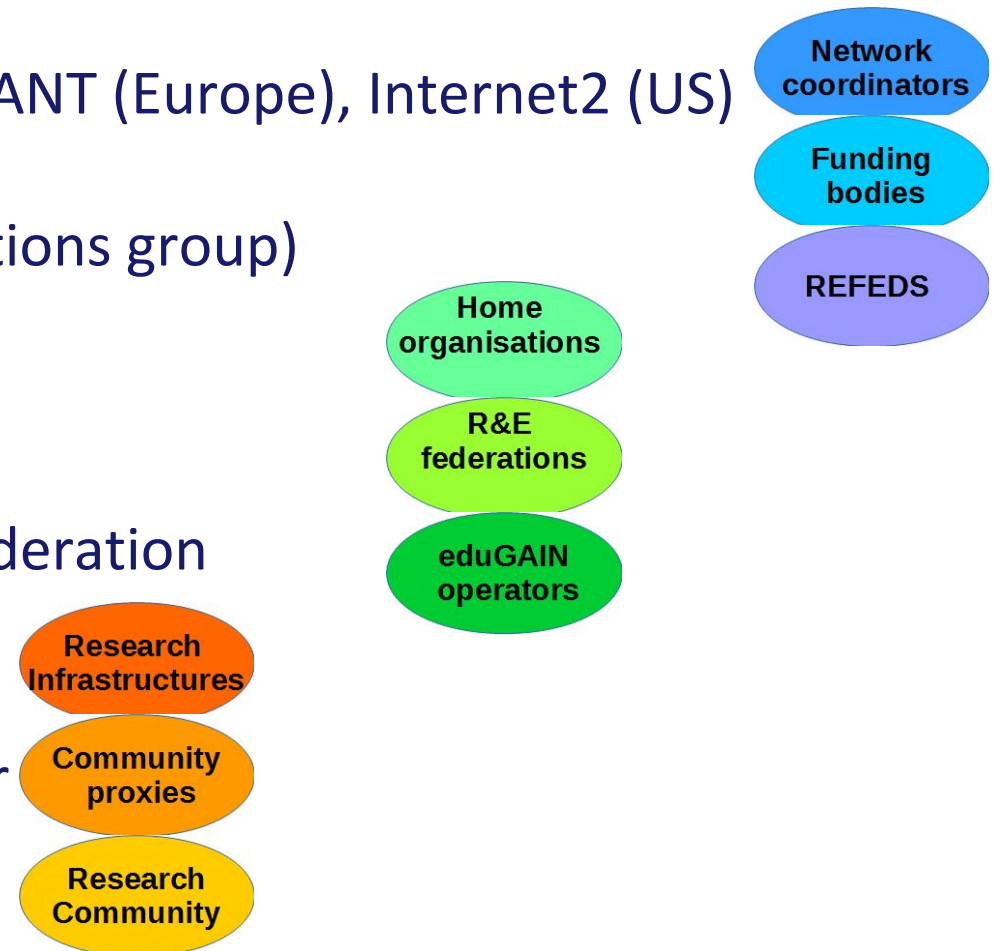
Overview on recommendations

Governance & Sustainability

Research representation, funding for sustainable operation, ongoing coordination

Baseline of User Experience

Attribute release, remove interoperability barriers, non-legal status, user mobility

Security Incident Response Readiness

For federations, interfederation and organisations

Harmonisation of Proxy Operations & Practices

Reuse generic services, follow best practices for interoperability

Sensitive Research User Experience

Support multifactor authentication and publish Assurance Profiles

Governance and Coordination

- Increase research representation in FIM governance
  Network coordinators   R&E federations
  – plan and prioritise with input from all key stakeholders
  – Having sufficient research representation in their governance and advisory functions helps ensures the alignment of actions with the mission to provide collaborative access to research
- Sustain operation of critical FIM services
  Network coordinators   Funding bodies   R&E federations   Research Infrastructures
  – It is essential that FIM services be operated sustainably, reliably and with user support fit for all research use cases
  – Testing environments, help desks, accessible documentation, IdP of Last Resort, token translation
  – diversity of the research communities should be reflected in the AAI offerings; we do not see a single solution as a sustainable future

Governance and Coordination

- Provide avenues for ongoing coordination
  - Collaboration among parties across the FIM ecosystem (federations, research e-infrastructures, research communities), should be on a continuing basis
  - A forum for exchanging AAI experiences should continue to be owned, supported and attended by research communities

Network coordinators

Funding bodies

R&E federations

Baseline of Research User Experience

- Release Research & Scholarship attributes
    - Some research communities rely on their underlying proxies to obtain basic user attributes directly from users when users' home organisations do not supply them, but it would enhance user experience if this would not be needed
    - All home organisations should participate in the R&S Entity Category
    - Research service providers and proxies should support the Data Privacy Code of Conduct by implementing its recommendations and asserting a corresponding entity tag in their federation metadata.
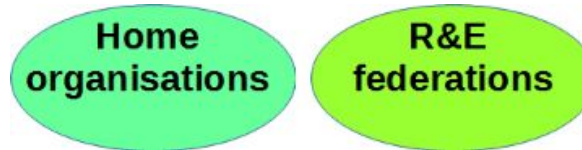
Baseline of Research User Experience
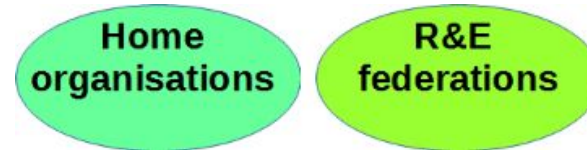
- Provide usability essentials
  - It is important to make the user experience as intuitive as possible, especially within the discovery process
  - R&E federation members should ensure that all of their entity metadata includes basic aides to good user experience:
    - Identity and service provider logos help users find their way
    - error URLs help them to get the right person's attention

Baseline of Research User Experience

- Remove interoperability barriers in eduGAIN metadata processes
  - Different R&E federations implement different policies:
    - determining which of their entities to export to eduGain
    - filter some entities from the eduGain aggregate before publishing the result to their members
    - Every such filtering excludes users from the federation experience
  - R&E federation operators should harmonize their eduGain export/import practices
    - eduGain itself can addresses risks presented by entities sourced elsewhere
    - All R&E federations should support the ability of researchers and scholars at their member identity providers to access research services they need for their work.
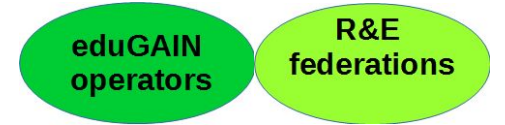
Baseline of Research User Experience

- Admit research organisations to federation

R&E federations

- – Some research e-infrastructure operators and research communities are not legal entities and are not represented by other legal entities
- – Some research organisations are intrinsically and essentially transnational, not aligned with membership of any specific national R&E federation
- – One or more R&E federations, or perhaps eduGAIN, should provide reasonable processes to include such cases into FIM and widely promulgate them across R&E federation operators

Baseline of Research User Experience

- Enable researcher mobility
  - Researchers change home organisations more often than they change their research domains or collaborations with colleagues
  - Research community proxies should support account linking techniques that enable a researcher to continue their appropriate access to research resources across such transitions.
  - ORCID IDs should also be leveraged for this purpose.

Community proxies

Security Incident Response Readiness

- Organisations participating in R&E 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.
  - These plans should be tested periodically.

Harmonisation of Proxy Operations

- Follow the proxy model and related AARC guidelines
  - The AARC BPA reflects the many successful proxy innovations produced by the research e-infrastructure community
  - Continued innovation, such as further development of federated solutions for non-web use cases, should build on this model
  - AARC Guidelines on operational practices should be followed by proxy operators

Community proxies

Harmonisation of Proxy Operations

Research Infrastructures — Community proxies — Research Community

- Re-use shared AAI and related services
  - Research communities and research e-infrastructure operators are strongly encouraged to use shared AAI services and related components, such as token translation and group management.
  - AAI services generally made available to research communities by research e-infrastructures should be assessed for adoption before duplicate what is already available
  - All large-scale e-infrastructures are encouraged to make AAI services generally available to research communities
  - Research communities are strongly encouraged to facilitate this re-use by collaborating to govern and sustain operation of a minimal set of enabling proxies and e-infrastructures sufficient to support their research activities

Sensitive Research User Experience

- Identity provider organisations are encouraged to provide strong authentication credentials to their researchers
  - implement the REFEDS MFA Profile to enable research service providers
  - proxies to signal when a user needs strong authentication to continue their activity
  - Identity assurance frameworks such as the REFEDS Assurance Framework should continue to be developed to respond to these needs

Next Steps

Next Steps

- Final draft published on fim4r.org

- Aiming to publish on Zenodo in the coming weeks

  – Full list of Authors, along with their ORCIDs

- The whitepaper will be widely circulated to relevant stakeholders, including:

  – Federation and interfederation governance

  – Technology Providers

  – The FIM Interest Group of the Research Data Alliance (RDA), with the aim to publish a derived document as an RDA Document

Over to you…

- Where should we be circulating this paper?

- Do you already see areas where FIM4R recommendations can be applied?

Acknowledgements

Acknowledgements

Visit fim4r.org

Questions?

Many thanks for joining the session