Agenda and Notes - 2016-08-10

Per-Entity Metadata Working Group - 2016-08-10
Agenda and Notes

[EtherPad used to create these notes: sqE3fhfdjL.etherpad]

===>> Note the new PIN and meeting URL <<===
Dial in from a Phone:
Dial one of the following numbers:
+1.408.740.7256
+1.888.240.2560
+1.408.317.9253
195646158 #
Meeting URL (for VOIP and video): https://bluejeans.com/195646158
Wiki space: https://spaces.at.internet2.edu/x/T4PmBQ

Attendees

- Scott Koranda, LIGO
- Michael Domingues, University of Iowa
- Nick Roy, InCommon
- Ian Young
- Paul Engle, Rice U
- Scott Cantor, tOSU
- Tommy Doan, Southern Methodist University
- John Kazmerzak, University of Iowa
- Chris Phillips, CANARIE
- Tom Mitchell, GENI
- Walter Hoehn, Memphis
- IJ Kim, Internet2
- Tom Scavo, InCommon/Internet2
- Kevin Morooney, InCommon/Internet2
- Phil Pishioneri, Penn State
- Ann West InCommon/Internet2
- Steve Carmody, Brown

Agenda and Notes

1. NOTE WELL: All Internet2 Activities are governed by the Internet2 Intellectual Property Framework. - http://www.internet2.edu/policies/intellectual-property-framework/
2. NOTE WELL: The call is being recorded.
3. Agenda bash
4. Client caching - Do we have consensus?
    a. Client caching should not distract from building a highly available service (the "just like DNS argument")
        i. We have consensus
    b. The existing Shibboleth SP and IdP MDQ caching is sufficient for most campuses for a MDQ service operating at 4 nines
        i. What are the details about how the caching works? Tom Scavo thinking of starting conversation on the dev list.
            1. IdP functionality is less tested? Probably.
            2. Not making a plea...fact finding mission.
        ii. We should be targeting 5 nines, not 4 nines. 1 minute of downtime per week is not acceptable.
            1. 5 9's allowed downtime: 5.26 min/yr, 25.9 sec/month, 6.05sec/week
        iii. With 5 nines Scott C feels that the Shib caching in memory is acceptable, 4 nines would require disk caching by the client
            1. Scott C has some concerns about the TIER Docker integration work and its intersection, 5 nines helps to alleviate that...
        iv. Scott C discusses discovery feed would be JSON just for discovery initially, TLS protected web site
            1. Fallback that does not assume SPs are loading the big full aggregate
            2. Chris C, need to make sure that the user experience ala discovery continues to be the same, keep the security principles we have today resulting from what we have today with SAML and the aggregate
            3. Do we know enough about SimpleSAMLphp?
                a. Like the Shibboleth IdP, support for MDQ in SSP is new and untested
                b. Scott K will reach out to developers and try to get some input on caching behavior
                c. Caching behavior is unknown
                d. Documentation is lacking
        v. Volunteer to poke at Ping, Microsoft again?
            1. Nick Roy will follow up
        vi. also have an ongoing conversation with Ellucian (don't you mean WSO2? :) maybe invite them / make them aware of this activity)
    c. Campuses requiring higher levels of availability can invest in their own caching layer or service
        i. That should not be advertised as necessary but for the rare campus that needs to make such an investment
            1. What if any is the interaction here with TIER and what they deliver?
    d. What do we tell large campuses that rely on InCommon metadata "internally" about risks if they lose internet connectivity to the "outside"? (long cache durations in absentia of local MDQ svr?)
        i. Is that a reasonable risk anymore or is network connectivity redundant enough these days?
        ii. Nick shares that campuses using Duo have seen this issue, though actually with Google analytics stopping the loading...
        iii. If 5 nines not good enough then have to run something locally (software claiming 5 9's on a network requires the network to run at 5 9's or better OR we attempt to mitigate that lower down in the stack risk - CP)
        iv. Need to have something in the report about this.

v. How is this different than other cloud services (like Google analytics)?
vi. Will an SLA become necessary for InCommon?
    1. Current is "best effort" as documented in the FOPP
    2. Known as an issue. Was highlighted in earlier review. Working on. Needs to happen.

5. "It should not be overly difficult or costly for a federation to run the entire per-entity infrastructure for its members"
    a. Is the consensus that the costs of the service itself (UK estimates £200/month on Azure CDN) are not overly costly?
    b. What about the cost to Ops? Does InCommon Ops have the necessary expertise and personnel with the necessary skills?
        i. TSG and IJ has insights? Not a problem in terms of signing and deploying and running on CDN.
        ii. No current experience with CDN, but do have experience delivering services from cloud (web service, some development servers)
        iii. Current aggregate distribution is in I2 data centers
        iv. Network is qualitatively different service, but it is there (and run by IU)
        v. Report from Leif on SAMLbits? Chris to ping him.
        vi. Does InCommon have an operations model currently that fits in the proposed devops model? What are those costs to operations?
        vii. Assessing a total cost of operating the service here.
        viii. This is different than running the federation manager.
        ix. We need to provide enough so that Ops can do the TCO analysis.
        x. Getting advantages from leveraging the I2 network. Also have data egress fees waived due to Net+ arrangements. Fact of the current AWS arrangement today. Egrees fees generally small compared to other costs.
6. Tom Scavo on current InCommon metadata aggregate process
    a. https://docs.google.com/drawings/d/134iWL9Ue_LC-hZqOL3i8YLlU3B83A6X-_tJK1dJXkQM/edit
    b. Pre-eduGAIN ingestion MDQ beta sources metadata from the same locations as in diagram, but post eduGAIN now draws on the preview aggregate (preview because beta service)
    c. Could Ops gain experience by moving current aggregate delivery (or a fraction) into cloud?
    d. What is the value of moving the distribution of aggregates into the cloud? I'm not seeing any value in that, for either Ops or deployers. The current distribution method Just Works.
    e. Action Item: What are the costs and risks associated with splitting the aggregate? (TomS)
7. Splitting the current aggregate
    a. What splits realistic?
    b. Does any split require the "preview->main->fallback" triple?
    c. Can an aggregate containing only IdPs help bridge through MDQ and the discovery issue?
    d. How might advertising a split aggregate impact MDQ adoption?
8. Turning identified risks for per-entity metadata service into requirements - https://spaces.at.internet2.edu/x/WIEABg
    a. Security
    b. Availability <-- need more precision per last call -- retrieval of aggregate is different than editing/producing it at the same availability criteria (will reduce cost/effort)
        i. Is the consensus "4 nines"? (downtime of 52.6 min/yr, 4.32 min/month, 1.01min/week)
    c. Responsiveness/Capacity
        i. Are we able to start writing down "Consumer MUST be able to retrieve MDQ ack/nack of a record in X ms"?
        ii. Should we build from Tom Mitchell's first analysis?
    d. Instrumentation (of the service for analysis of use by clients)
    e. Costs
    f. Organizational (what if any are new requirements on how InCommon staff responds to issues)