General Work Stream Meeting

Data Security Work Stream

Date: March 4
Time: 13:30 UTC
Meeting Chair(s): David Bernick

	Agenda Item	Speaker	Time
1.0	Welcome	David	5 min
2.0	Polish up "Breach Response"	All	40 min
3.0	Summary of Federated Analytics Meeting - DSWS Action Items	JP, David, Francesco	10 min
4.0	"Malfeasance rules" Product - Gauge interested, interested DPs?	All	30 min
5.0	A.O.B.		5 min

Minutes:

Breach Response

DB: Brushing up the document, are there any interested 'owners' of the doc? Been used by EGA but no one is driving it.
OH: Same with AGHA, used but not driven.
DB: Needs to be usable by a broader group
OH: Breach response - bring it to the GHIF.
DB: Mailing out to the mailing list to see if there are any volunteers, otherwise we will own it.

DB: Boiler plate right now. Every country and state has their own requirements so we don't want to be over prescriptive.
KE: Would be useful to have an SOP that people could adopt.
DB: Great guidance, add to appendix.
KE: We have a series of SOPs dealing with different kinds of data. Important to have these for audits.
DB: Does anyone have breach response docs that they would be willing to share? Carry over to mailing list and continue there.
OH: Have run drills following breach response, but happened some time ago. Will check with AGHA.

DB: Polish off as a deliverable this year.

Recap of Federated Analytics and Cloud Security

JPH: Presented the state of affairs in technology for privacy-conscious data sharing. Sending analysis to data, protecting partial aggregates. Shown results that demonstrate it is operational at scale. Presentation from Marilyn on the way they are dealing with MVP and Alzheimer's cohorts - secure shared working space.
Identified a few driver projects to make further progress in terms of how we want to update some of the Cloud APIs. Please get in touch if interested in this topic.

Malfeasance Rules

DB: Key ideas are that we don't have a good way of understanding if someone is doing bad things, exceeding mandates in terms of research. How to detect policy problems. Policies without some method of detection aren't as useful.
[Github page] Example of detection. Categorize what's happening, what's detected. Having a corpus of rules may be in conjunction with the DUO group. What do we as a group do about it? Standards, new AI structures? Open for discussion and something we should have an answer to.
OH: Trying to understand practicality. It's an honor system. How would this change the honor system.
DB: Would change to honor system with something over it. Interested in coming up with a different system
SV: Snort (network) for logging?
DB: Yes, possibly. It's usually effective until it's not. People can write arbitrary tools and do something with the data.
JPH: If credentials are compromised, someone who is ill-intentioned has access to the system. How to separate legitimate requests vs request with ill-intent, very difficult to develop. Should be something to detect non-sophisticated malfeasance.
OH: Looking for access patterns, but often done by hackers.
DB: It's a balancing act, want us to put in a policy that things can be audited for bad behaviors, and if bad things happen here's the consequences. Without a way of detecting that, it's not valuable.
OH: This would only work in cases where data cannot move.
DB: Came up with this from experience in Cloud. Remote access to data, data you're not hosting. No systematic approach to wrapping heads around this as a conceptual problem. As we have more containerized workflows that bring outside tools, having a way of detecting whether these tools are manipulating data.

SV: Offload to SIEM?
DB: SIEM without rules or rulesets becomes a logging repo. Could do this, but then what happens.

PM: Detecting certain behaviours that put someone at higher risk. Are you referring to building some type of UBA engine?
DB: What are the behaviours, and what can we add to that to help. Maybe calling out the honor system as a little flawed with regards to some kinds of data.
PM: Even on a policy level, how do you define something as risky.
DB: Working with DUO group would be great. Eg. this research doesn't match the ontology term.
PM: Maybe come up with a scenario, and put together the pieces. What are we trying to detect, what is bad behaviour. Then look into what logs we need.
OH: Will be interesting whether users accept this.
DB: Maybe just for certain systems.
PM: Are we looking to keep good guys good, or someone malicious using a good guys credentials?
DB: Bit of both. There's bad guys, but also good guys exceeding stuff. For example, people downloading a full data because they're annoyed at using the Jupyter notebook.
OH: Worried about second guessing why people are doing research in a particular way, ie. Breast cancer researcher looking at particular genes that aren't brca1/2.
DB: Had to be applied very carefully to certain datasets, with agreement from users on how they will use that data.

DB: Whole point is that we have protections from downloading it all. Have locked down biobanks, but a lot of initiatives are in the middle with policy and rules but not strict.
PM: Are we trying to police people even though the policy allows access, or looking for people trying to get around controls.
DB: Shouldn't be universally applied. Should have an opinion (or our varying opinion) down on paper. Maybe milestone 1 is how to GA4GHers think about these different directions and can we provide any direction. Technical solutions are hard and social solutions are hard.
PM: Interested in taking a stab at this. Put an outline together on what we should do.

SS: Trying to understand what FedRAMP is doing.
DB: We let users run arbitrary code against data they have legitimate access to. There are a series of system communications control that say you shouldn't allow egress of sensitive data. Allow you to hand wave your way through some of it.

Send out an email to the mailing list to come up with some milestones of this thing. General opinions, different scenarios. Coming up with a framework to rate sensitivity of datasets.


GA4GH
CONNECT

Attendees:

1. David Bernick 2. Jean-Pierre Hubaux 3. Lindsay Smith 4. Justina Chung 5. Francesco Marino 6. Alexander Senf 7. Jamie Cuticchia 8. Kurt Rodarmer 9. Rob Davies 10. Oliver Hofmann 11. Dylan Spalding 12. Alef Janguas 13. David Jackson 14. Heidi Sofia 15. Keith Elliston 16. Jaime Delgado 17. Max Barkley 18. Frederic Haziza 19. Subhashini Jagu 20. Sai Lakshmi Subramanian	21. Plamen Martinov 22. Tim B 23. Guy Kelman 24. James Bonfield (Sanger) 25. Shu (Sue) Hui Chen (NHLBI) 26. David Lloyd 27. Rodrigo Theodoro 28. Susheel Varma (HDR UK) 29. Jordi Rambla 30. Mamana Mbiyavanga 31. Susan Fairley