Insert Name of Deliverable: **Authentication and Authorization Infrastructure (AAI)**		
□ In Progress/Planned (included on current Strategic Roadmap) ✱ New Version (approved deliverables where a new version is planned) □ New Deliverable (not on the current Strategic Roadmap)		
Description	The GA4GH Authentication and Authorization Infrastructure (AAI) Profile is a technical profile for managing and authenticating the identity of users, and for authorizing access requests for data and services offered through the Driver Projects. The GA4GH AAI Profile is based on the IETF OAuth 2.0 standard, and the OpenID Connect identity layer based on OAuth 2.0, and incorporates the researcher identity vocabulary and data-use ontology developed by the Data Use and Researcher Identity (DURI) work stream.	
Source Driver Projects	ELIXIR, EGA, CanDIG	
Other communities that will contribute to development	CINECA	
Proposed Implementations (at least 2)	Planned Server Implementations 1. 2. Planned Client Implementations 1. 2.	
Expected Completion Date		
Milestone (add additional rows if required)	**Description**	**Expected Completion Date (by quarter)**
Milestone 1	Continued updates to align with the expansion of the Passport spec ● Assurance Levels/Identity Proofing	
Milestone	FASP advancement plus Security guidance	
Milestone 3		
Projected Submission Date		

Insert Name of Deliverable: **Risk Assessment Methodology for Software Stacks**		
□ In Progress/Planned (included on current Strategic Roadmap) □ New Version (approved deliverables where a new version is planned) ✱ New Deliverable (not on the current Strategic Roadmap)		
Description	In the community of genomics, many groups lack training in security assessments and the followup of security best practices. This deliverable will be multiple parts: 1. A how-to on assessing risk aligned to a known framework. This will include both formal alignment as well as “colloquial” alignment so that a typical software developer can use it to assess their product. 2. Methodology/algorithm for groups to self-service risk assessment beyond what’s in the DSIP or the Approval documents. . 3. If U24 funding is approved, starting up a group that would actually do assessments and provide for a stream of open source tooling to increase the automation of these assessments. a. This would be a group that would also train other groups and encourage using self-service tools	
Source Driver Projects	Human Cell Atlas is already using a variant of this and would like to continue to pursue. But in general, any project that lacks its own security oversight that aligns to a known standard would be a good driver. TopMed is another.	
Other communities that will contribute to development	Same as Security Infrastructure, as it’s the same audience. Matchmaker exchange wants rules around Data Classification as well as “how to add a new node” and assessment/audit would be good ways to do that. Biodata Catalyst, 7 Bridges	
Proposed Implementations (at least 2)	Planned Server Implementations 1. 2. Planned Client Implementations 1. 2.	
Expected Completion Date		
Milestone **(add additional rows if**	**Description**	**Expected Completion Date (by quarter)**

required)		
Milestone 1	A matrix of “must have” security controls for both builders of software as well as operators. Work with REWS	Q2 2020
Milestone 2	A procedure for which a developer can risk assess, threat model and vulnerability assess their own implementations of GA4GH specifications.	Contingent on Grant
Milestone 3	If funded via u24, automated tools for self-assessment and guidance on operation.	Same as above
Projected Submission Date		

Insert Name of Deliverable: **Rulesets for detecting suspicious behavior**		
❑ In Progress/Planned (included on current Strategic Roadmap) ❑ New Version (approved deliverables where a new version is planned) ✱ New Deliverable (not on the current Strategic Roadmap)		
Description	In conjunction with the DUO group and applications like DUOS, we will build a set of rules that can be used to detect behavior that goes against a user's allowed use of data. There's a well-known Web Application Firewall project called mod_security. Part of that are The Core Rules which spell out a meta-language for detecting certain known attacks. While not entirely comprehensive and perfect, it gives a baseline for understanding what "anomalous" behavior might look like. We want to do something similar for Data Use in life sciences environments.	
Source Driver Projects	NIH, TopMed	
Other communities that will contribute to development	DUO, Passports, REFEDS	
Proposed Implementations (at least 2)	Planned Server Implementations 1. 2. Planned Client Implementations 1. 2.	
Expected Completion Date		
Milestone (add additional rows if required)	**Description**	**Expected Completion Date (by quarter)**
Milestone 1	Core set of rules defined for "low-hanging fruit" of data use malfeasance. Expressed in a simple language. Literature search, develop metalanguage. See https://github.com/Neo23x0/sigma for a good start.	Q1 - brainstorming

Milestone 2	Implementation of these rules that can be running in a code environment -- whether through Code Injection or like a “docker sidecar” or some other methodology. Similar to WAF/RASP technologies that exist today.	
Milestone 3	Connecting this with Passports: follow up on malfeasance	Q4 2021
Projected Submission Date		

Insert Name of Deliverable: **Federated Cohort Exploration and Analysis**

❑ In Progress/Planned (included on current Strategic Roadmap)
❑ New Version (approved deliverables where a new version is planned)
✱ New Deliverable (not on the current Strategic Roadmap)

Description	This can be due either to privacy regulations or to the large volumes of data involved. In such settings, it is nevertheless desirable to be able to take advantage of the overall available data, typically to perform Machine Learning (ML) operations (both training and inference), without moving the data. To realize this task in the best possible conditions of security and performance, the most appropriate approach combines two well-known cryptographic techniques, namely secure multi-party computation (SMC) and homomorphic encryption (HE). Such an approach provides strong security guarantees, but comes with limitations in terms of the sophistication of the operations that can be carried out.
Source Driver Projects	SPHN, VICC? ELIXIR?, Driver Projects related to Large Scale Genomics? Genomics England + Genomics Australia? ENA/EVA/EGA?, NCBI
Other communities that will contribute to development	The iDash community (http://www.humangenomeprivacy.org/) is very active on this topic, and so is the homomorphic encryption standardization group (https://homomorphicencryption.org/), Cloud WS (DRS), BAH (homomorphic encryption)
Proposed Implementations (at least 2)	Planned Server Implementations 1. Cohort exploration server 2. Planned Client Implementations 1. Cohort exploration client 2.
Expected Completion Date	

Milestone (add additional rows if required)	**Description**	**Expected Completion Date (by quarter)**
Milestone 1	Use case definition • Data Quality Checking	Q1

Milestone 2	Design and Implementation of proof of concept & Update with DP needs	Q3
Milestone 3	Evaluation	Q4
Projected Submission Date		2021

Insert Name of Deliverable: **Cloud Security and Privacy**		
❑ In Progress/Planned (included on current Strategic Roadmap) ❑ New Version (approved deliverables where a new version is planned) ✱ New Deliverable (not on the current Strategic Roadmap)		
Description	The Cloud has gained the attention of many GA4GH projects, and it is becoming increasingly used for large-scale distributed computing services. The Cloud Work Stream emerged to focus on API standards to make it easier to send the algorithms to the data in such environments, and run full workflows on the cloud. The use of Cloud services (and outsourced services in general), poses multiple legal, ethical and technological challenges in terms of data transfer and processing, for which GA4GH should develop appropriate specific guidelines and recommendations for a secure and privacy-conscious use of Cloud services. An example issue to be addressed is the recommended policy for cryptographic key management. This effort requires an agreement and joint collaboration between the DSWS, REWS and Cloud WS. This will overlap and intersect in the Risk Assessment proposal above and might be a subset of that effort.	
Source Driver Projects	All driver projects dependent on the Cloud WS and/or making use of/producing cloud services	
Other communities that will contribute to development		
Proposed Implementations (at least 2)	Report (no implementations)	
Expected Completion Date	2021	
Milestone (add additional rows if required)	**Description**	**Expected Completion Date (by quarter)**
Milestone 1	Analysis of Cloud-related frameworks (regulatory, ethical, security technology)	Q2

Milestone 2	Evaluation of technology and organizational approaches matching regulatory/ethical requirements	Q3
Milestone 3	Recommendations/Guidelines aligned with the GA4GH Data Security Toolkit	Q1 2021
Projected Submission Date		Q2 2021

Insert Name of Deliverable: **BlockChains for Query Recording**	
❑ In Progress/Planned (included on current Strategic Roadmap) ❑ New Version (approved deliverables where a new version is planned) ✱ New Deliverable (not on the current Strategic Roadmap)	
Description	**The first thing to do here is to confirm the demand within GA4GH. If this is confirmed, we should refine the objectives sketched hereunder and secure the related resources.** A query interface enables a researcher to interrogate a database that is potentially distributed among several stakeholders, e.g. a distributed cohort. Through a set of well-designed queries, a malicious researcher (or someone who has stolen a researcher’s credentials) can perpetrate inference attacks such as re-identification attacks. It is therefore crucial to keep track of the queries, in case an enquiry has to be carried out.. Blockchain designates a recent Computer Science technique that allows different stakeholders to keep track of transactions and store them in an immutable distributed ledger, with a full copy at each of the stakeholders. As such, a blockchain is an appealing solution to the problem mentioned above. It is to be noted that the blockchains that will be used here are “permissioned” (or closed) as opposed to the “permissionless” type of blockchain that typically underpin cryptocurrencies such as Bitcoin.
Source Driver Projects	SPHN, MatchMaker ? ENA/EVA/EGA ? ClinGen?, NCBI, NHGRI, ELIXIR?
Other communities that will contribute to development	The iDash community is somewhat active on this topic.
Proposed Implementations (at least 2)	Planned Server Implementations 1. Permissioned blockchain 2. Planned Client Implementations 1. Client of the blockchain 2.
Expected Completion Date	

Milestone (add additional rows if required)	Description	Expected Completion Date (by quarter)
Milestone 1	Requirements analysis	Q2 2021?
Milestone 2		
Milestone 3		
Projected Submission Date		Q4 2022