GA4GH Product Approval Processes

Executive Summary

At the GA4GH Strategic Planning meeting in Hinxton, UK in May 2017, the community identified a need to have a set process for GA4GH to certify the specifications which are its products. An approval process involving the Foundational Work Streams and Steering Committee peers was suggested. This document covers the process a product will flow through to be a blessed GA4GH standard. A specification moves through five stages; proposed, submitted for approval, under review, approved and retired. Specifications must have a "Product Proposal Form" outlining their scope, expected impact and those affected to be reviewed by GA4GH before product work begins. There is an additional submission process before a product is approved. Products should be hosted publically with adequate documentation.

Introduction

At the GA4GH Strategic Planning meeting in Hinxton, UK in May 2017, the community identified a need for a best practices guide to help steer the development and adoption of GA4GH product. This document establishes a number of best practices, that when adhered to, would allow a products to be adopted into GA4GH's ecosystem. It does not prescribe an in-depth and strict set of do's and don'ts. Each product comes with its own community and working practices. To enforce the same working practice over all products would be counterproductive and only increase the administrative burden of initiating new work or bringing external products into GA4GH.

Product Lifecycle

A product can exist in one of the following states

1. **Proposed** - GA4GH has been notified of the intention to develop this product
2. **Submitted for Approval** - a product has been submitted for approval to GA4GH. It will be undergoing review by GA4GH committees outlined below. The product may possibly undergoing refinements during this phase.
3. **Approved** - a product is accepted into GA4GH
4. **Retired** - a product is no longer deemed suitable for GA4GH

Product Proposal

Products should be proposed and developed in response to the needs of the GA4GH Driver Projects or in response to a known need within the wider community for standards.

The initiation of new products is likely to occur when new Driver Projects are introduced to the Work Stream and work on underway products is completed. The need for a new product may also coincide with a documented community requirement. Initial prototyping and possible exploratory work may take place wherever those developing feel is appropriate.

The Work Stream Leads should bear in mind that the aim will be to produce **a small number of new specifications**, with wide adoption, each year. Priorities should be reflected by a request for a desired product coming from multiple sources, be they Driver Projects or others active in the Work Stream activities. The Driver Projects who have explicitly requested a product will be the source Driver Projects for that Product.

If the Work Stream Leads feel they would like to formally turn an initial prototype into a product to develop to completion, a Product Proposal Form should be filled in. This will be used to:

1. Notify the other Technical Work Stream leads of the nature of the product being developed and identify potential areas of overlap.
2. The Secretariat will prepare a copy of the Data Security Questionnaire for the proposed product. This should be filled in by the relevant Work Stream members. Once filled in this can be passed on to the Data Security Work Stream who can then review it.
3. The Regulatory and Ethics Work Streams will be notified
4. Notify the secretariat of resource requirements

There may be some amendments made to the initial Product Proposal. These could be in response to issues raised by Data Security Foundational Work Stream, or by other Work Stream leads identifying a possible area of co-operation or scope enhancements.

Product Approval Process

The product will be ***Submitted*** by completing the GA4GH Product Approval Submission Form (available here) and then emailing GA4GH Secretariat at secretariat@ga4gh.org to notify the Secretariat that the form has been completed.

The submitted product will be reviewed by the following bodies:
1. The Data Security Foundational Work Stream
2. The Regulatory and Ethics Foundational Work Stream (REWS)
3. A specially convened Product Review Committee.

The Steering Committee will be notified by the Secretariat that the product has been submitted.

A questionnaire will be provided featuring questions set by the REWS. This questionnaire will be provided by the Secretariat and should be filled in and then passed as a link in the GA4GH Product Approval Submission form.

The Product Review Committee will consist of three members as detailed below nominated by the submitter.

1. A Work Stream leader from one other Technical Work Stream
2. A member of a third, different, Technical Work Stream
3. A representative of one of the source Driver Projects for the product, who has been involved with the Product Development

The choice of members will be approved by the Engineering Committee, subject to any changes suggested by the Secretariat. The review committee nominates a representative to communicate back to the Secretariat (this is the technical work stream leader should no nominations be made). The review committee may give a response of "Accept", "Reject" or "Changes Requested". All three members must agree unanimously for the committee to give a positive assessment. The committee should give this response 1 month after submission of the specification to be reviewed. Guidelines for PRC Members are available here.

It is recommended that the relevant Foundational Work Streams are contacted with any Security, Regulatory, or Ethical concerns they may have during the development process. The reviews made at this stage are in addition to those performed during the **Proposed** phase.

A positive assessment is confirmed when the Secretariat receive confirmation from the representative of each of the reviewing bodies that the review body passes the product. If a review body does not pass the product, requested upgrades will be communicated to the submitting Work Stream Managers. An upgraded product can be sent to the reviewers directly. This cycle can be repeated until the product passes the review body requirements.

Once positive assessments have been made by all three bodies, the product will be sent to the GA4GH Steering Committee for Approval. One Work Stream Lead from will present the product to the Steering Committee. The Work Stream leader in charge of the PRC may also be called upon to explain the PRC processes. after which it will be deemed **Approved**.
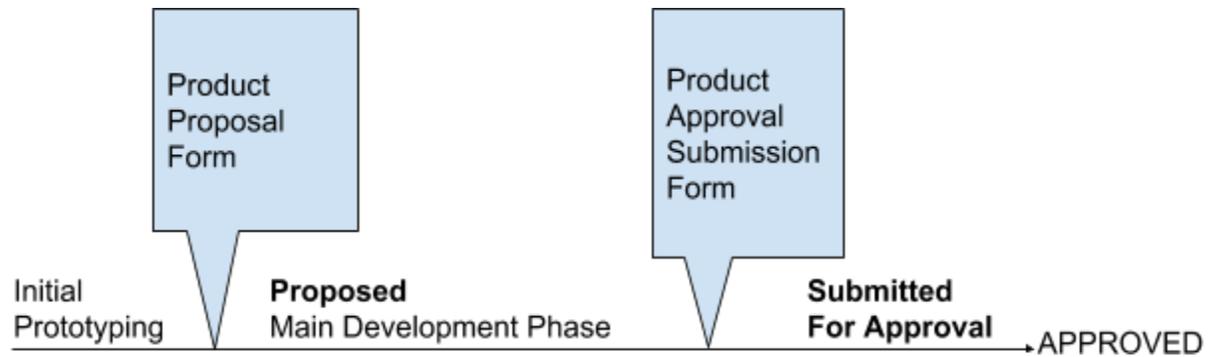
Versioning

Specifications should follow the semantic versioning pattern of MAJOR.MINOR.PATCH e.g. 2.1.1 as detailed at http://semver.org/. This is to communicate the severity of changes in a product to downstream users. MAJOR changes are considered breaking e.g. changing a URL or renaming a field in a schema. MINOR changes are considered as changes in functionality but not breaking e.g. adding a new field to a schema. PATCH changes should be reserved for changes that do not change the contracts built within a product e.g. expanding the size of a field in a schema.

Exceptions may be made where the product intentionally fits in to an alternative versioning scheme dictated by the products target community. In this case a clear schema for the versioning must be available and a link to the versioning schema provided in the Product Approval Form, and if possible, the specification itself.

Approving New Versions

Minor and Patch updates to the product may take place without the Product Review Committee being re-convened. Major versions will cause incompatible changes to previous versions of the specification. The approval of a new major version requires a new Product Approval **submission**, this time using the provided GA4GH Product Update Form. Once this process starts the product will be reviewed by the same bodies as for a first time release.

Publish Location

The product should be available in a public configuration management system capable of tracking requests for changes, authorship, and history. Submitting the specification to GitHub for tracking suffices. The repository should have at least three people assigned by Work Stream leads capable of resolving product change requests subscribed and watching. Any issue or pull request should expect a response within a short time-frame[1]. Each person should be a GA4GH

---

[1] 2 weeks is assumed a maximum response time with public holidays allowing for some extensions to this

contributor, who agrees to the GA4GH IP policy (once finalised), and GA4GH Standards for Professional Conduct.

If a GitHub repository is used as the publish location, the GitHub user ga4gh-vc should be added as an owner or manager of the repository in question. This user is managed by the Secretariat and will help ensure long term oversight of the repository in question. A Secretariat managed user with similar rights should be setup if a different version control system is used.

GA4GH Paladins

GA4GH paladins are a set of nominated people within the GA4GH community who ensure the smooth running of the product according to the above best practices. Sets of paladins will be assigned a product/designation according to their background and knowledge. They will also be given commit ability to the product. This will only be used in extreme circumstances if product owners cannot be contacted. These people will be selected from the Engineering Committee members.

Review

Products will be subject to an review by GA4GH Steering Committee after the period of one year, if required. Typically the Steering Committee will review a specification after 5 years.

Retirement Process

Work Stream Leads wishing to retire a product from GA4GH can submit a request using the GA4GH Product Retirement Form detailing the reasons why. Once agreed, the product may be withdrawn. Any attribution or mention of GA4GH will be removed if appropriate. Products will be updated to point to their replacement if they have been superseded by a newer GA4GH Approved Product. GA4GH may choose to fork a product if the Secretariat deem a need to.

Existing Standards

The following GA4GH specifications were in existence before these Approval Procedures were put in place, so may deviate from the requirements laid out in this section.
1. CRAM file format
2. SAM/BAM file format
3. VCF/BCF file format
4. Genomics API
5. htsget[2]

---

[2] htsget was reviewed by the Security FWS before it became a standard

Specification Best Practice Recommendations

Below are the recommendations made by GA4GH to ensure a robust specification is developed.

Document Structure

Each specification should be laid out in a document detailing the following:

- A clear author list of the document (may be within the document or handled by another mechanism)
- A date of publication and a version
- An executive summary summarising the major points of the document (if applicable)
- Introduction setting the document's scope
- Standards incorporated
- Detailed layout of the specification including schemas, APIs, data formats employed
- Describe any domain specific terms within the document
- A brief version history for documents that have received substantial updates

Document Format

The document should be presented in a human readable format without the need to install a paid for third-party application or viewing outside of the browser. HTML output and PDF are acceptable formats so long as the source markup format used to generate the document is available e.g. Markdown, RST, HTML, LaTeX. In the case of Markdown and RST, GitHub provides automatic HTML generation features. We recommend the use of GitHub flavored Markdown (https://github.github.com/gfm/) and to limit the use of dynamic code within documentation.

Standard Endpoints

GA4GH API specifications should use the following common end-points to convey information

/service-info

The precise nature is to be determined

/error

The precise nature is to be determined

Implementations

Each specification should have at least two associated implementations. These do not need to be officially maintained reference implementations. For a client-server model there should be at least two servers implementing successfully with two clients. These are viewed as best efforts and viewed as a way to ensure a robust product is developed.

These implementations are not expected to be written or managed by Work Streams themselves. They are to be developed by the Driver Projects and the community. Ideally they would operate in the real-world environment, with real-world the data, in which the developed standards aim to be used. Work Streams would facilitate interoperability testing between these implementations. This might result in updating specifications during the development process if the implementations highlight unforeseen circumstances.

Requirement Level Indications

GA4GH specifications should meet [RFC 2119 on the use of key words to indicate requirement levels](#).

User Requirements

GA4GH specifications should meet the requirements of the source Driver Projects which requested them, or have had some feedback from the community in which it is intended to be used. [RFC 7282](#) is recommended for development teams to consider to guide the decision making process.

Citable

The submission should include a plan for publication to a journal or similar entity to allow for a citable reference to the product. Journals used for publication must be open access and not behind a firewall. In addition GitHub repositories can be registered with Zenodo to mint DOIs (https://guides.github.com/activities/citable-code/). This is not a requirement of GA4GH to do for all standards but is a way to generate a citable entity.

Interoperability

If the specification is required to interact with areas covered by other GA4GH standards, it must comply with those standards.

Security Contact

To enable a long-term contact for security flaws an email address of security-notification@ga4gh.org has been set up. This email will be monitored by secretariat and security members to allow for an incoming response to be directed to appropriate parties. Please feel free to use this email address in your specifications or websites if required.

Editor(s)	Description	Date	Version
Andy Yates/ Rishi Nag	First unified copy	2017-07-27	1
Rishi Nag	Comment responses and fixes, updates to new version procedures	2017-08-02	2
Andy Yates	Expanding on the life-cycles of a product and description of software best practices.	2017-09-20	3
Rishi Nag	Existing specifications listed, new product ideas section added	2017-10-19	4
Andy Yates, Thomas Keane, Dixie Baker, Melissa Haendel, Ewan Birney	Extracted software into new document, expanded summary and cleaning up comments with input from GA4GH Work Stream members	2018-01-15	5
Rishi Nag, Andy Yates	Updates to Product Review Committee, GA4GH Secretariat oversight to repos, Fast Track section removed	2018-04-11	6
Rishi Nag	Changes from Steering Committee Meeting 2018-05 incorporated	2018-06-19	7
Rishi Nag, Adrian Thorogood	Updates on Security, PRC steps	2018-08-19	8