Global Alliance for Genomics & Health

ARCHIVED

Notes for 2025 here: ▶️

GA4GH Variant Representation Meeting Minutes 2025

GA4GH Variant Representation

Minutes and Actions 2023-2024

Others: 2017 | 2018 | 2019 | 2020 | 2021 | 2022| 2025

These are the minutes for the Variant Representation (VR) meetings, a subgroup of the GA4GH Genomic Knowledge Standards Work Stream. For further information, please visit the GA4GH Work Stream page (will link to the page on the website when it is ready) at https://ga4gh.org.

Meeting Protocols

- Please note that by participating in meetings, attendees agree to adhere to the GA4GH Standards of Professional Conduct
- Meetings may be recorded for note-taking purposes. Recordings will be deleted within three months of the meeting taking place.
- Variation Representation is a subgroup of the Genomic Knowledge Standards Work Stream.
- GitHub Project board is here.

Table of Contents

Meeting Protocols
Table of Contents
- 2024-02-03
- 2024-10-28 PRC and SequenceReference
- 2024-09-30
- 2024-08-26: VRS 2.0 community Q&A session
- 2024-08-12
- 2024-07-29
- 2024-06-24
- 2024-06-10
- 2024-05-06

2024-04-25
2024-03-18
2024-02-19
2024-01-29: Connect meeting planning
2023-01-15: Canceled: Public Holiday
2023-12-04: VRS 2.0 alpha to beta approval process
2023-11-20
2023-10-16: Canceled
2023-09-04: Labor Day
2023-07-24: Review of VRS 2.0a changes
2023-07-10
2023-06-19: CANCELED (Public holiday)
2023-06-12: CANCELED (Due to the unavailability of the leads)
2023-05-22: VRS 2.0 working meeting
2023-05-15: VRS 2.0 working meeting
2023-05-01: VRS 2.0 working/planning meeting
2023-04-24: CANCELED
2023-04-17: Connect preview
2023-04-03: CNV,VRS 1.3 & SV-SIG
2023-03-27: CANCELED (Due to unavailability of the leads)
2023-03-06: Copy number,MNV & Connect
2023-02-27: CNV & VRS 2.0
2023-02-06: Pipeline/Tool Rundown,Roadmap &CNV
2023-01-23: Recap and 23Q1 Roadmap Plan(Canceled - MFW)
2023-01-16:Sync for Genes Phase 5
2023-##-##: TEMPLATE

2024-02-03

Please see **new** meeting minutes:

GA4GH Variant Representation Meeting Minutes 2025

2024-10-28 PRC and SequenceReference

Hosting: Alex Wagner, Larry Babb
Attendees “Name (Affiliation)”: Beatrice Amos, Daniel Puthawala, Kori Kuzma, Senkei Umehara, Wes Goar, Robert Freimuth, Liam Mulhall, Catherine Procknow, Arthur Ko, Melissa Cline, Tristan, Alan Rubin,Olusola Afuwape

	Agenda Item	Speaker/Leader
1.	Product review updates	Alex
2.	Trial use classes	Alex
3.	SequenceReference new feature discussion	Alex

Link to meeting recording:
https://us02web.zoom.us/rec/share/xCICbCAiG7AsT1io224kRDNQ3MrhQnnn7Na-B076_ByvHzfjPEulh13iVDDUbzNY.flPUdo2acWWbrTWG?startTime=1730148921000

Meeting summary:

PRC updates:
First, the ongoing PRC process for VRS 2.0 was highlighted, with registered implementations having a month to review VRS classes and provide feedback. Thanks everyone for your feedback, especially Bob, for his extensive input!
Please see the VRS project board on GitHub where you can track tasks under the GA4GH organization. Items on the board are tagged for easy identification, particularly those emerging from product review, such as item #565.
Larry's new class hierarchy diagram - https://github.com/ga4gh/vrs/issues/56, - please review and comment on it.
The VRS team will continue PRC discussions on Thursday, focusing on documentation rather than model changes. Input on these updates was encouraged to help refine the VRS documentation and organization.

New feature discussion:
Please can everyone weigh in on proposed updates.

1) The first feature is adding a "molecule type" to the sequence reference class to specify if a sequence originates from a genomic, transcript, or protein source. Alex expressed uncertainty over using "genomic" and considered alternatives like "chromosome" or "template" to better reflect the category.
2) The second feature, "sequence reference literals," arose from work on the MAVE project. This addition would allow the explicit sequence used for digest generation to be represented directly within the sequence reference object, particularly useful for shorter sequences relevant to specific experimental datasets. This small change, utilizing the existing sequence string class, adds expressiveness to the sequence reference class.

Alan supports the use of "chromosome," "transcript," and "protein" as molecule types by referencing the naming conventions in RefSeq, where "NC" designates chromosomes and "NM"

indicates transcripts. This alignment with established terminology serves as a strong justification for retaining the current nomenclature.

Bob expressed concerns regarding the relationship between the value sets in the enum for molecule type and the residue alphabet fields. If the sequence being referenced does not directly align with these fields, it could lead to inconsistencies.
Alex clarified that the residue alphabet field was introduced to interpret the characters in a literal string, such as distinguishing between amino acids and nucleic acids.
Bob then asked about potential modeling constraints to validate the alignment of these two enums, suggesting that this topic had not yet been discussed.
Alex acknowledged the introduction of the molecule type concept during the meeting and mentioned that while he doesn't want to create validation logic unless there's a clear implementation need, it remains a valid concern for future discussion.

Melissa raised concerns about the terms "genomic" versus "chromosome," questioning whether using "chromosome" could limit future applications, especially regarding the Human pangenome project.
Alex suggested that "template" could be a more functional term to describe sequences.
Catherine expressed difficulty with the word "molecule," preferring a term like "molecular context" to distinguish between actual measurements and conceptual knowledge of variants.
Alan said that "template" might not be suitable, as it could restrict the classification of elements, particularly non-coding ones.
Senkei inquired about the description field's potential for arbitrary strings, highlighting concerns about how it relates to the reference type's validity.
Alex clarified that while the sequence itself is defined by a constrained string format, the description does not enforce any particular validation and exists merely as a JSON schema descriptor.

Larry talked about the importance of defining terms clearly. He supported using "molecular context" instead of "molecule."
There was a consensus that "genomic" could be an intuitive term to use, although concerns about its specificity were raised. The conversation explored the classification of elements like mitochondrial DNA as chromosomes and whether terms like "transcript" or "protein" adequately reflect genomic data.
Everyone supports Alan’s suggestion of ‘borrowing’ established nomenclature from RefSeq would be beneficial, as it would provide a straightforward mapping to commonly used accession systems without reinventing existing terminology.

<u>Class diagram visualizations:</u>
Larry initiated a discussion on whether to add class diagram visualizations to the documentation, seeking feedback on its value and prioritization. Although some feedback from team members, including Bob and Jules, showed interest in this feature, he was cautious about the resources required. Larry proposed creating diagrams that show relationships and

associations between classes, similar to how FHIR displays UML views, as these could provide clearer context and ease of understanding.

Bob supported this idea, noting he relies on graphical representations to understand structural relationships in documentation. He suggested adding structural diagrams to show class interactions and inheritance explicitly, which could enhance usability.
Catherine questioned whether the diagrams would depict direct inheritance or just associations between classes, to which Larry clarified that they would also show interconnected references, such as how an allele links to sequence locations and references.

Alan voiced strong support for simplified, high-level visualizations without excessive detail, likening them to a "you are here" map. He said such visual aids helped make the VA spec accessible, particularly a table that clarified class inheritance. Alan also suggested generating diagrams dynamically to reduce maintenance. While Larry liked the idea, he mentioned technical limitations and proposed starting with static diagrams.

Alex: diagrams would be useful, but they should be prioritized lower given current documentation demands. Perhaps initially focus on essential components and possibly adding these views as enhancements later.

Bob suggested that Larry consider creating targeted diagrams corresponding to specific examples, showcasing how various classes work together in a straightforward way. Larry will create mock-ups for further feedback.

GKS Commons" to "core":
He then introduced the idea of renaming "GKS Commons" to "core" for clarity, aligning with recent structural changes within the project. Larry explained that GKS Commons was initially a repo for shared superclasses and data types, with VA retaining core knowledge model components. However, as these components have since been moved to SEPIO, it made sense to rebrand GKS Commons to reflect its current function more accurately.
Alex expressed strong support for this renaming, noting that previous constraints around the "core" label no longer applied.

Larry also raised the topic of how to manage a subset of entities under "domain entities," which was created for essential yet minimally organized classes like genes and conditions. He asked if this subset should remain under core or become a separate package, emphasizing that delaying this decision might complicate future changes.

Folder Structure Changes:
Larry suggested renaming “Core IM” to “Base” and “GKS Commons” to “GKS Core” to better represent foundational elements. Alex supported these changes, suggesting that "Base" would more clearly indicate that these elements are foundational to other components.
Larry noted that in FHIR, “foundational” is used similarly in organizing documentation, agreeing that "Base" would be a suitable term.

What is the best way to label “Domain Entities,”? Larry described these as concepts like genes and diseases, which map to established authorities (e.g., HGNC for genes, MONDO for diseases). Ultimately, it’s decided to keep “Domain Entities” as it is to avoid complicating existing schema structures.

Updates on the review process:
Alex announced a shift to a model with a minimum one-week review period for all features, with extended time for complex requests if necessary. He emphasized that the goal is to prevent unnecessary delays while ensuring everyone’s input is considered. After this period, pull requests will either proceed or return to discussion if major issues arise.

2024-09-30

Hosting: Alex Wagner, Larry Babb
Attendees “Name (Affiliation)”: Beatrice Amos, Daniel Puthawala, Kori Kuzma, Senkei Umehara, Beth Pitel, Wes Goar, Kyle Ferriter, Terry O’Neill, Andrew Patterson, Mrinal Thomas, Robert Freimuth, Liam Mulhall, Olusola Afuwape, Salem Bajjali

	Actions Arising
	1. Schedule a meeting with the Phenopackets team to address low-level alignment issues. 2. Larry to initiate the PRC process for VA and Cat-VRS standards with Daniel and Matt. 3. Add class hierarchy figures and examples to the documentation before submission to the steering committee. 4. Explore the classification of sequence objects and clarify the impact on schema validation and digest computation. 5. Leads to discuss potential loosened constraints on RefGet accessions and alternative referencing systems - guided by feedback gathered via GitHub and Slack. 6. Align documentation with current implementations, ensuring it supports multiple accession types. 7. Consider introducing an optional field for additional identifiers in VRS2.1. 8. Gather additional feedback through Slack and GitHub during the community review period.

	Agenda Item	Speaker/Leader
1.	Welcome and reminders: - October 9th: GA4GH Initial VRS2.0 PRC meeting 2024 - ASHG GKS Unconference meeting Nov 4, 2024 - register now!	Bea 2mins
2.	Connect summary and review.	

	2 GKS sessions: 1. **Mad GKS 2.0: the Implementation Warrior:** ○ Summary and key takeaways 2. **Mad GKS: beyond maturity:** ○ Summary and key takeaways 3. Any other relevant outcomes from other Connect sessions?	
3.	Review of VRS 2.0 implementer feedback	
4.	Review and status of previous action items: - A concise summary for the latest release notes is still pending. Action is needed to create and distribute this. - (Update older examples in the VA documentation - there was a call for volunteers to self-assign this task and contribute to improving the spec.)	
5.	AOB	

Link to meeting recording:
https://us02web.zoom.us/rec/share/jZylinuHYV1SJE-soC6cr_2_Aiykwnt0KzeYMS6C7ovwEQVFVBuDKg0Hl_NwpeFX.ZQSmjkIYD9CzAcP1?startTime=1727729612000

Meeting summary:

GA4GH Connect recap:
Key takeaways discussed. Larry emphasized that while alignment of global standards is important (e.g., between FHIR, Phenopackets, and other data structures), the team's priority is to focus on **delivering current projects** rather than pausing to consolidate standards. **There is value in alignment but resources are limited and will make it difficult to pivot right now.** Alex supported this stance, noting that during Connect, there was consensus on moving forward with implementations while working on alignment in parallel without it being a blocking factor. A meeting with the Phenopackets team was also mentioned as an upcoming opportunity to address low-level alignment issues.

Larry will be working with Daniel and Matt to initiate the PRC process for these standards, aiming to release them soon. Regarding maturity beyond implementation, Larry noted that they reviewed the maturity model, reflected on responses from the first two trial surveys, and discussed how this feedback will guide future development, aiming for progress early next year. Community feedback is very important in shaping this trajectory.

If you haven't already, please complete the registered implementations form [link].

Going over the feedback form results:
Larry noted that the feedback primarily consisted of general statements about interest or disinterest in various features of VRS, with only one significant issue raised regarding the need

for attributes that was missing from the specifications, necessitating the use of extensions. Alex then discussed the trial use ballot and highlighted resources available for accessing the ballot and related discussions.

Larry asks about the status of the comment period for implementers, which Alex confirmed to be closed, although a broader public review would occur later as part of the GA4GH product approval process.

Improving the documentation:
Suggestions included the addition of class hierarchy figures and links to examples from concept model pages, both of which would be valuable and feasible to include before submission to the steering committee.

Sequence class within VRS2.0, and its support for RefGet accession numbers:
The idea was to enhance descriptive metadata for sequences while maintaining the RefGet identifier as a primary reference. Should certain entities, such as variations and sequences be classified as domain entities?
The implications of these classifications and their relationship to value objects within VRS was explored - there are differences in how these concepts relate to mappings across vocabularies. The definitions and relationships are complex and further discussion is needed to clarify these classifications and their impact on VRS functionality.

There was discussion around the need for an identifier in a class structure, specifically focusing on the differences between "Id" and "identifiers." Bob requested clarification on the abstract classes and inheritance structures that Alex had discussed. He specifically asked for clarity regarding the inclusion of certain attributes - Id, alternative label, description, and accession - in the digest.

Should the "sequence" class only reference refget accessions? If so, we need clearer documentation to avoid ambiguity about its use. Alex clarified that while the class can reference other sequences, it requires a digest for proper utilization, emphasizing that if the underlying objects lack digests, no dependent digests can be computed. There was consensus about potentially loosening constraints on the use of sequence objects that do not include refget accessions. There was also consideration of the impact of missing refget accessions on schema validation and digest computation, with a focus on ensuring robust error handling in the coding model.

Maybe we need a flexible approach to referencing sequences, particularly as requirements around keys may loosen. How do we encourage consistent use of standardized referencing approaches, such as adopting the RefGet standard alongside the VRS system? Bob expressed concerns that the current structure limits support for implementers who may wish to use alternatives to RefGet. He proposed creating an abstract class above the current one to accommodate various accession systems, allowing users to specify the system and accession

separately. Alex argued that RefGet has become a de facto standard, suggesting it's easier to loosen constraints later rather than tighten them now. Bob reiterated the importance of not alienating potential users who might not adopt RefGet.

Bob suggests that leads should think about aligning the documentation with the current implementations, and ensure that it accurately reflects the intent to support multiple accession types.
Mrinal adds that having a consistent identifier across different systems is useful, but that clinical labs might lack the technical resources to implement more complex solutions. What about the idea of introducing an optional field for additional identifiers - this would be a more inclusive solution that would accommodate various implementations while maintaining clarity in documentation. Leads will further explore these ideas and ensure that any proposed changes will address the community's concerns. Or maybe think about incorporation into VRS2.1.

Larry says that while some solutions will be beneficial, they must prioritize those that are crucial for implementations. The best solution would be one which could extend the existing model without breaking changes, especially with the upcoming release.
This discussion will be continued over Slack. Additional feedback and comments, particularly trial-related issues, will be gathered through GitHub during the community review period.

Great dive into FHIR, and leads will have a plan to address remaining action items in future discussions.

2024-08-26: VRS 2.0 community Q&A session

Hosting: Alex Wagner
Attendees "Name (Affiliation)": Kori Kuzma, Wes Goar, Kyle Ferriter, Terry O'Neill, Andrew Patterson, Mrinal Thomas
Apologies: Larry Babb (vacation), Beatrice Amos (UK Bank Holiday)

	Agenda Item	Speaker/Leader
1.	Welcome and Reminders: - GA4GH Plenary/Connect meeting in September: Register: https://broadinstitute.swoogo.com/ga4gh-12th-plenary/ - Melbourne booth planning: GKS Plenary Booth - Roadmap renewal doc - ✅	
2.	VRS 2.0 community Q&A session! Please review the VRS 2.0 ballot release documentation.	

3.	AOB	

Link to meeting recording:
https://us02web.zoom.us/rec/share/sS-cAIIUun8Rd7HpyCfg9ZrlyWLEJ80b7Z5H6uK_hOsB8lZXn9YtbNEvSOqPJfiS.MfCkpDbmJ5Hmz-Pm

🔈 **The 2.0.0-ballot.2024-08.1 release has been published! Full documentation is available here. If you represent a registered implementation, please review and submit comments using this form.** You can provide feedback on the data classes, documentation, conventions, or any other aspect of the specification.

Meeting summary:
Community Q&A session:

Q: When will the VA spec be released for use in the Gk pilot data?
A: Community meetings will restart after the Plenary and that progress was being made internally.

Q: Is there a summary or an easier way to understand the changes between the two published versions of the spec for those who aren't constantly involved, rather than manually comparing the differences themselves.
A: the closest resource available is the summary provided during the GA4GH Connect event in April 2024. See GitHub repo, summary of updates available in the April release under the "tags/releases" section.
The documentation for these updates has been compiled into the trial use ballot.

Action: a concise summary for the latest release notes is still pending.

Q: What is the general process for translating VRS IDs back into detailed data, such as locations and positions, particularly when merging observation data from labs and sending it back? E.g. use case: a small implementation within Australian genomics is focused on rare diseases. It's clear how to retrieve data from an ID generated but what is the standard practice or agreed-upon methods others use to reverse-engineer VRS IDs into detailed information?
A: Two common approaches to using VRS identifiers:

a) In one scenario, when running a live service, a VRS ID might compactly represent a variant linked to a specific disease. If more detail is needed, the service could provide an endpoint to retrieve the full VRS object.
b) In the second approach, when creating a dataset for broader use, full-form objects with their corresponding IDs might be stored in a variation records file. These IDs would then be referenced in larger documents, helping to keep data manageable and preventing overly large JSON files.

These approaches allow for flexible use of VRS identifiers in both persistent data storage and live services.

The importance of updating older examples in the docs highlighted - if anyone is interested to self-assign the task and contribute to improving the spec, please let him know.

Q: Link to the maturity model document requested:
A: GKS Technical Specification Maintenance document. This outlines the maturity model, product maintenance, and development processes.

Please engage during this 4-week review period and reach out with any questions.

2024-08-12

Hosting: Larry Babb, Alex Wagner
Attendees "Name (Affiliation)": Liam Muhall, Oscar Pastor, Kori Kuzma, Tristan Nelson, Terry O'Neill, Wes Goar, Mrinal Thomas, Bob Freimuth, Salem Bajjali, Melissa Cline
Apologies: Beatrice Amos

	Agenda Item	Speaker/Leader
1.	Welcome and Reminders: - GA4GH Plenary/Connect meeting in September: Register: https://broadinstitute.swoogo.com/ga4gh-12th-plenary/	Alex
2.	(CGC feedback, also from cytogenomics community? How does this feedback into the VRS roadmap?)	Alex
3.	VRS 2.0 and documentation updates	Alex
4.	ASHG 2024 Hackathon	Larry

Link to meeting recording:
https://us02web.zoom.us/rec/share/8KoXxfmPC5Ze9KxFqevtaxU3PWeaoXAWq8o5vc0yHyCh6hsAFJhWnAEcOp47Bv9h.fGb9eMWDuraZqUAZ

Meeting summary:
Welcome👋:
Oscar Pastor (Technical University of Valencia) expressed interest in actively participating in the group. Oscar mentioned a project in Valencia focused on managing genomic data for clinical use, which he hopes to align with the group's standards.
Liam Mulhall(Stanford) is involved in an HLA project and the ACMG v4 guideline model.

🗒️GKS product implementer form:
Please can new participants register their implementations for better communication and collaboration as the standards advance.

Documentation:
The VRS team is focused on updating documentation for trial use - would like an early review of the major version release. Please add feedback to vrs.ga4gh.org.
Alex went over the documentation updates for the upcoming 2.x version of the VRS model. He provided an overview of the key changes from version 1.3, highlighting new classes such as "adjacency" and "terminus molecule," which allow for better representation of molecular variations and structural rearrangements. He also discussed the "Cis phase block," which replaces the previous "haplotype" concept due to ambiguity concerns.
🔔Please review the updated documentation at vrs.ga4gh.org and provide feedback during the four-week review period.

Review process for implementers:
Alex explained the **maturity model** used for advancing specifications through different stages: draft, trial use, and normative - this will ensure stable and well-tested implementations. Community involvement in reviewing and endorsing the documentation to ensure its robustness before moving to more stable phases is important.

Larry introduced a new review process for implementers to evaluate the readiness of various classes and data types for trial use in the VRS 2.0 model. Implementers are asked to rate their commitment to implementing specific classes, ranging from having no plans to actively using them in production. This process aims to gauge the level of engagement and identify any major issues or concerns that might hinder implementation.

The review form allows implementers to abstain from voting on classes outside their domain or to flag any issues that prevent implementation. The form also breaks down categories like location, molecular, and systemic variation, and includes essential core data types from the newly established GKS commons repository. The team emphasized that trial use doesn't mean the classes are finalized; rather, only minor tweaks will be made moving forward, and major structural changes are less likely at this stage.

Versioning:
Alex explained the versioning process and stability expectations for the upcoming VRS 2.0 release. The team uses **semantic versioning**, with different stages of maturity indicating the likelihood of changes. At the draft stage, nothing is stable, and changes can happen rapidly without broad input. Once the draft progresses to the trial use stage, changes will occur less frequently, and any updates will involve community consultation, although they could still be significant. A minor version update will be made for new features or breaking changes at the trial use level, while reaching the normative stage means the specification is highly stable, and any backward-incompatible changes would trigger a major version release. **The goal is to**

involve implementers throughout this process, ensuring transparency and input at every stage.

Larry said the start of a four-week review period is for implementers to assess and provide feedback on the VRS documentation. The goal is to gather input through a form, which will be processed at the end of the period, allowing for the promotion of elements to trial use if they meet the criteria.

Liam asked about the relevance of him filling out the review form, given his focus on the VA spec. Larry responded that while it might be premature for HLA-related input, it could be useful for him to register his interests and be involved in the ACMG-related aspects. Larry will assist Liam with the form.

"Chromosome location" discussion:
Alex revisited a previous discussion about the "chromosome location" class in VRS 1.3, which relates to cytogenomics data. This class helps specify locations on molecules using methods like G-banding, often used in cancer diagnosis. There are no active implementations, but interest was shown in the past in developing data standards for cytogenomics to improve data sharing and reproducibility. While VRS 2.0 won't include these updates, future releases may address them once implementations are registered and tested.

Upcoming hackathon in Denver, Colorado, preceding the American Society of Human Genetics (ASHG) conference. The hackathon, led by Kori Kuzma, will focus on GKS and VRS projects, potentially including cytogenetics. It aims to provide a platform for pilots in VA and the core IM, as well as VRS 2.0.

2024-07-29

Hosting: Larry Babb, Alex Wagner
Attendees "Name (Affiliation)": Daniel Puthawala, Kori Kuzma, Terry O'Neill, Mrinal Thomas, Salem Bajjali, Melissa Cline, Kyle Ferriter, Wes Goar
Apologies: Beatrice Amos (vacation)

	Agenda Item	Speaker/Leader
1.	Welcome and Reminders: - GA4GH Plenary/Connect meeting in September: Register: https://broadinstitute.swoogo.com/ga4gh-12th-plenary/ - Turn on GitHub notifications in VRS repo - VRS Project Board	Alex

2.	Updates on product approval process for VRS 2.0 - ✅Reconvene PRC - ✅Schedule meetings with REWS & DS - 🗓️Schedule meetings with PRC (BA) - Progress documentation	Alex
3.	Active Issues: 1. Updates to JSON Schema $id fields 2. Discussion #461: derivative molecules/sequences	Alex

Link to meeting recording:
https://us02web.zoom.us/rec/share/hsGGdZpo86LHfVVKZrYj7_JuVvD8QWVUWx8R6m-D7lw5R9B4PA30_msNSmmgvDWN.uC0GxKkI1XA5VxQi

Meeting summary:
Reminders: Please turn on GitHub notifications for the VRS repo to ensure you stay updated on ongoing work. The VRS project board includes a significant issue of updated documentation subdivided into tasks for each data class.

Active issues: the recent TASC session decision topic will ensure schemas have resolvable URIs. The VRS repo now includes URIs in the JSON Schema id fields, allowing schemas to be referenced and resolved using standard JSON Schema processing tools. These URIs are now live and can be accessed through a browser, applicable to VRS, Cat-VRS, VA spec, and GKS common libraries.

Daniel: Is this something that TASC gave like blank? Get approval for? Or is this something that we'll have to go through as part of the product review process for catverse as well and other specs?

The approval of establishing a convention for GA4GH schema persistent URLs: The approved convention involves using the W3C id.org resolver with GA4GH organisation schemas, product versions, and resource-specific paths. This structure, now approved by the TASC, can be utilised across any GA4GH product. For specific cases like the VRS and Cat-VRS specs, which have straightforward directory structures, the URLs will resolve to the raw JSON schema documents via GitHub. For GKS Common and VA specs, which have more complex internal structures, a slightly different pattern was devised to ensure accurate resolution. The next step is to document this agreed resolution for the PRC approval process, ensuring that future product reviews check for compliance with these URL conventions.

Derivative molecules/sequences:
The proposal is to adopt the term "derivative molecules" to align with the concept of molecular variation, alongside classes like allele and adjacency. This terminology could also facilitate integration with non-sequence models representing variations, such as karyotyping and

cytogenetics. The upcoming VICC Hackathon at CGC will further explore these topics. Outstanding tasks include adding circularity and specifying range interpretations in the model.

Phase blocks and alleles in derivative molecules. Larry summarised his recent thoughts on the incorporation of alleles. A single allele cannot exist outside a phase block. This aligns with the idea that an allele should coexist with other components like block or adjacency within the phase block. Using phase blocks for star alleles, which are complex and may involve multiple SNPs (single nucleotide polymorphisms) or even no SNPs, depending on the reference.

Star alleles are more suitable for inclusion in the Cat-VRS framework, as they relate to knowledge statements and profiled sequence locations. The implementation should capture these complexities. Mrinal supports this approach. The consensus was to **include adjacency within phase blocks in derivative sequences**, and this would be added to the list of issues to tackle for derivative molecules, moving towards implementation.

Documentation process and feature implementer interest forms for the GA4GH VRS project: Larry suggested collecting similar data for Cat-VRS and VA specs to get a holistic view of the data sets needed. Alex emphasized the importance of updating documentation and ensuring class implementations in the reference VRS Python before advancing to trial use. Mrinal will work on adjacency and derivative sequence documentation.

Discussion on the terminology of "derivative molecules" versus "derivative sequences," with Alex explaining that the broader term "molecule" accommodates future cytogenetic mapping. Important to future-proof the framework by using inclusive terminology.

2024-06-24

Hosting: Larry Babb, Alex Wagner
Attendees "Name (Affiliation)": Alan Rubin, Melissa Cline, Aleksander Milosavljevic, Beatrice Amos, Bradford Powell, Catherine Procknow, Daniel Puthawala, Emma Tudini, Kori Kuzma, Mrinal Thomas, Salem Bajjali, Kyle Ferriter, Tristan Nelson, Wes Goar, Robert Freimuth, Tim Hefferon

	Agenda Item	Speaker/Leader
	Welcome and Reminders: - Register for Connect Plenary meeting - Connect session agendas - deadline due June 26th - Connect session proposal form - Review roadmap milestones - *5 VRS meetings remaining before Connect*!	

	- Product approval process for VRS2.0 - expected SSC presentation Jan 2025:
	  - ✅Reconvene PRC
	  - 🗓️Schedule meetings with REWS & DS
	  - 🗓️Schedule meetings with PRC	

	VRS Discussions - Limited Activity - discussion on ways we can encourage more engagement	Larry
	Maturity Model Process next step for registered implementers	Larry / Alex
	(update) VRS 2.0 PRC - what's the process ?	Larry
	Connect Proposals	Larry / Alex

Link to meeting recording:
https://us02web.zoom.us/rec/share/hDqf0kXCOnrKLPb7pqG6plUV1ULFnqh8f-P98yRrMKZglJ86GKp6jAE3fwIahKWX.izbGPyPb5UL4BM2E

Meeting summary:

github.com/ga4gh/vrs/discussions/461

github.com/ga4gh/vrs/discussions/493

Turn on GitHub notifications in VRS repo (if you want)

VRS engagement:
Larry talked about the need for consistency in meetings, using the VRS platform for discussions, and encouraging more activity and participation on the GitHub page. Everyone is encouraged to subscribe to updates and contribute more actively to the discussions. Alex highlighted a proposal regarding modeling sequences and derivative sequences, urging feedback and involvement from the group.

Kori suggested creating groups or teams within GitHub to tag and notify relevant members about updates, emphasizing the importance of opt-in participation.

Maturity model process and the next steps for registered implementers:
The team reviewed the current list of implementers and emphasized the importance of independent project teams to align with GA4GH standards. We need implementers to have a significant influence on prioritizing work to ensure broader adoption and effective implementation of standards.

Alan raised a point about distinguishing between groups that want to use and influence the standard versus those that want to use the reference implementation. He said that his team

lacks the capacity to implement the standard themselves and relies on the reference implementation.
Alex responded by highlighting that the key factor defining an implementation group is their intent to produce or consume data formatted using the specifications. He gave examples of different types of users and their needs, emphasizing the importance of understanding these needs to engage them effectively in relevant conversations about the standards.

👀 For GA4GH secretariat:
Alan emphasized the need for clear communication when discussing projects related to GA4GH, cautioning against using terms like *Driver projects*, which might exclude non-driver projects from feeling involved.
Alex agreed and explained that the term "product implementers" is being used to include a broader community, allowing various groups to engage based on their capacity and interest. The goal of understanding the level of effort and interest from different groups to better align the development of new features with the available resources and needs of the community.

Larry clarified that each product implementer should be able to clearly articulate their work with the project, including how they are using or developing it, even if they do not have public source code. Alex said yep, everyone on the call likely represents an implementation and should be involved in shaping the standards.

In chat (Alan): *How does this interact with (or not) with the bigger-picture GA4GH Implementation Forum?*

The discussion touched on the issue of having multiple implementations from a single team, questioning if this would mean multiple votes or influence. Larry showed an example with Bob, who submitted two implementations: one internal and one public. Alex noted that the focus is on ensuring project representation and relevant specifications being linked to the project. The plan is to move to the next steps in the maturity model process, and will send out a new form to be sent out later.

Next steps in the GKS product development and maintenance process:
Alex talked about reaching out to registered implementers to gather input on features relevant to them for VRS 2.0. This includes selecting data classes and providing feedback as the specification progresses from draft to trial use. The goal is twofold:

1) ensuring that the scoping aligns with community needs and
2) identifying features where there may be insufficient interest or implementation capability.

Alex emphasized the importance of engaging implementers who can actively trial and provide feedback on proposed features. This approach aims to avoid developing features without real-world implementation interest or support. The survey will allow implementers to specify their interests and level of commitment, whether in terms of providing feedback or contributing developer effort. The aim is to foster a **collaborative**, **community-driven approach** to advancing specifications within GKS products.

Bob: How should individuals handle the challenge of focusing their development efforts on specific features they're interested in, especially when they've indicated interest in multiple features but can't commit to all of them?

Alex clarified that the current form includes multiple features simultaneously, but moving forward, they plan to address features one at a time. Since all listed features already have committed development efforts, individuals interested in contributing should simply indicate their general interest on the form. Specific contributions and further details will be coordinated based on responses submitted.

Larry: how does the registration process relate to moving features from draft to trial use? Alex explained that when individuals submit the form indicating their interest in specific features like allele, copy number change, and sequence location, their implementation and contact details are added to the group associated with those features. This ensures they receive communication about the evaluation process for advancing features from draft to trial use.

The decision to move a feature to trial use isn't solely based on the number of implementations but on the readiness determined by the product developers. Once a feature is ready for evaluation, all registered parties are notified. They are given an opportunity to review, comment, and suggest changes before the feature progresses. This process involves sending out emails or forms where implementers can provide feedback within a designated time frame.

Does having multiple implementers for a feature like sequence location would automatically trigger its transition to trial use. Alex clarified that the decision is made by the product leads based on readiness, and all interested parties are involved in the evaluation phase.

Alex n Larry agreed to streamline the process to ensure clarity and effectiveness in gathering feedback on each feature slated for advancement.

In prep for the upcoming Connect conference, Larry outlined a strategic agenda aimed at maximizing the limited session windows available. He and Alex collaborated on drafting proposals, focusing on three key topics:

1) a session dedicated to the GKS Product Roadmap and Maturity Model, intending to elucidate the roadmap. This session would also go into the evolving maturity model and its relationship with the PRC, clarifying how product maturity and approval processes align.
2) a GKS Implementation Forum, building on successful past sessions. This forum would highlight various implementations within GKS, encouraging attendees to review pre-recorded videos from different implementation groups beforehand. The aim is to facilitate in-depth discussions rather than extensive presentations, fostering interactive engagement and deeper insights.

3) a Roadmap Refinement Process session. This session would capture fresh insights and discussions from the conference, providing an opportunity to refine and update the roadmap based on the latest feedback and developments.

Deadline is on Wednesday - please share feedback.

2024-06-10

Hosting: Larry Babb,
Attendees "Name (Affiliation)": Bradford Powell, Brett Johnson, Catherine Procknow, Daniel Puthawala, Madhu Ouseph, Kori Kuzma, Mrinal Thomas, Salem Bajjali, Terry O'Neill, Tristan Nelson, Wes Goar
Apologies: Beatrice Amos (ELIXIR meeting), Alex Wagner (vacation)

	Agenda Item	Speaker/Leader	Time

	Welcome and Reminders:
	- Register for Connect Plenary meeting
	- Connect session agendas - deadline due June 19th
	  - Connect session proposal form
	- Review roadmap milestones - *6 VRS meetings remaining before Connect*!
	- Product approval process for VRS2.0:
	  - ✅Reconvene PRC
	  - 🗓️Schedule meetings with REWS & DS
	  - 🗓️Schedule meetings with PRC
		Larry	2

	Clarification of what defines an implementation project for the purposes of our maturity and product development process and priorities. This will also be an opportunity for a quick status update from any implementation projects underway or in the planning phase.	All	
	An update/review on implementor projects, the registration process (step 1) and the second step in the process to begin identifying Trial Use elements of the VRS and the Commons classes they depend on.	All	
	Discuss topic proposals for the upcoming Connect due by June 19		

Link to meeting recording:
https://us02web.zoom.us/rec/share/Qi8ijqaVGeacuDA7V1u7St0kIDXxCet3tNKXMmOOXcJh3HW1q3btA9pBdS878qQy.dwfhIkPRjy2gm0XK

Meeting summary:

GKS Technical Specification Maintenance
GKS Product Implementer Form

- kferrite@broadinstitute.org to add gnomAD
- Melissa Cline to add BRCA Exchange
- Others: Add your products/implementations!

The primary goal is to discuss **advancing towards the trial use of VRS and other GKS products**, acknowledging that instituting new plans and processes takes time.

Larry explained the concept of an implementation project, focusing on three main GKS products:

- VRS ,
- Cat-VRS (categorical variant representation specification), and the
- Variant Annotation (VA) specification.

Sequence annotation (SA) was mentioned but excluded for today's discussion. Larry is working with Terry, Kyle, and Tristan to transform Clinvar data into a GA4GH standardized representation using the VA, Cat-VRS, and VRS specifications. He emphasized the urgency of moving these projects to at least a trial use phase to encourage adoption and tool development.

The product technical specification and maintenance doc, written by Alex, outlines the GA4GH maturity advancement process. This process is designed to standardize specifications before they undergo the GA4GH product review for finalization.

The team has been working on the VRS product specifications, currently in the 2.X branch, which reference a shared schema from the GKS common repository. This common schema serves as the foundational building block for VRS, Cat-VRS, and VA specifications. It's **everyone's collective responsibility** for maturing the common schema entities and abstractions, which are integral to all related specifications.

Implementations discussion:

The Implementers Form was shared - this aims to gather information about current implementation projects - those that will influence and drive project priorities and maturity. Implementers are crucial as they commit resources to real projects, unlike academics or interested individuals who do not have the means to implement. The form helps in understanding who is using the GKS specifications, the scope of these implementations, and their influence on prioritizing development aspects, such as which types of variants to focus on first.

Wes mentions several ongoing projects involving GKS standards. Key projects include a grant-funded collaboration with the CIViC team, aiming to implement GKS standards within CIViC to enable direct ingestion of GKS data into their platform. Another project, the MetaKB project, focuses on knowledge base harmonization using GKS standards, specifically VRS and VA, to represent therapeutic responses, which are crucial for semantic treatment.

The **VarCat (Variation Categorizer) project**, currently internal but planned to be open-sourced, aims to be a comprehensive tool for variant assessment and classification in somatic diseases. This project employs guidelines from ClinGen CGC and AMP/ASCO/CAP for oncogenicity and clinical actionability and uses VRS and VA standards to represent therapeutic responses and prognostic values. Varcat is initially focused on pediatric cancer due to a lack of publicly available semantic variation classifications for this demographic, the tool is designed for broader applicability across all cancers. The overarching goal of these projects is to enhance data structure and richness, particularly for pediatric cancer, and to submit GKS data to ClinVar, ultimately supporting both internal and external use.

Larry discussed the integration of other projects using GKS standards. The focus is on developing tools in a stepwise approach, moving from draft to trial use and eventually to normative standards based on multiple successful implementations.

The **maturity of the development process relies on implementers validating and using these classes**, proving their effectiveness before solidifying them as standards. Alex's team has multiple projects under their implementation, including the VarCat project, which uses GKS standards for variant assessment and classification. Another team is working on converting ClinVar data into GKS format, ensuring no data is lost and making it semantically useful. This implementation is separate from others, such as Nithu Shah's team, which works on the Allele Registry and Linked Data Hub, using the VRS specification independently.

It’s important to register implementers to track progress and coordinate efforts. Each implementation team, despite potentially working on the same driver project, operates independently and contributes uniquely to the development and validation process.

Mrinal gave an update on the implementation project aiming to use GKS and VA specifications to connect variants entering **Epic** through various sources like VCF files, HL7 v2, or FHIR, with knowledge bases such as ClinVar and CIViC. They have successfully imported the sample data set into their system and matched it against GKS variants created from VCF files. The project is currently hindered by the inability to process certain ClinVar records that have incomplete information, specifically those missing required keys. However, all other records in the data set were successfully matched against VCF variants.

Wes says the cancer community is particularly interested in obtaining data for functional study evidence and clinical validation according to the Oncogenicity guidelines. The goal is to pilot the

integration of this data into a structured format that can be ingested and utilized within clinical pipelines, ultimately enabling its use in clinical applications.

Eugene Clark from **Mass General Brigham** did a large project involving a large genomic data hub. They are redoing their entire genomic data repository and have been adding VRS identifiers to thousands of whole genome sequencing VCF files. This process aims to facilitate easier mapping of variants to knowledge bases. Eugene presented preliminary work at the last plenary, and recordings are available for those interested.

The **GnomAD** project involves a cohort allele frequency draft model, which also needs to be registered.

The next step in the process involves a second Google form that will be sent to registered implementers. This form, expected within a month, allows implementers to specify which classes within each product they are interested in using for their implementations. This level of specificity enables implementers to prioritize classes they are actively working on, while possibly deferring newer classes like adjacency until they are ready. Implementers will also indicate if they have implemented these classes and if they believe they are ready for trial use. Once two or more implementers confirm readiness for trial use, a discussion will ensue to finalize their promotion to trial use status. This process encourages open dialogue where concerns can be addressed and decisions made collaboratively before moving towards normative status and product review. If there are objections to readiness for trial use, they will be discussed thoroughly to ensure consensus.

Connect meeting in Melbourne discussion

Kori proposed prioritizing the implementation of Cat-VRS, VRS, and VA spec. She highlighted the current dependency on Alex for updating schema files, which sometimes causes delays. She'd like more team involvement in this updating process to expedite progress.

There was a discussion about the effectiveness of previous Connect sessions and a proposal to make **future sessions more hands-on and technical**. It was suggested to replicate the success of a recent 3-hour implementers session where various projects showcased their use of VRS and discussed their progress and challenges. Wes emphasized the importance of practical demonstrations and resolving technical challenges collaboratively during such sessions. There was interest in showcasing projects like VarCat and integrating their outputs into Epic, but not quite yet ready for the upcoming Connect session in Melbourne.

Larry suggests a proposal to conduct a **practical "how-to" session** at the upcoming Connect event. We could use existing tools like VRS python tools, the normalizer, and potentially others to demonstrate the process of taking a VCF file, VRS'ifying it, and integrating it with databases such as ClinVar, GNomad, or the Link Data Hub. Wes suggested leveraging previous work done in Jupiter notebooks, specifically referencing a project involving real semantic VCF data normalized against patient data from Clinvar, etc. This project could serve as a model for a

public workshop or session to engage newcomers and provide hands-on experience with genomic data integration.

Mrinal indicated a tentative plan to attend the event and will provide updates on their progress, particularly focusing on their approach and challenges rather than showcasing visual demos.

2024-05-06

Hosting: Larry Babb, Alex Wagner,
Attendees "Name (Affiliation)": Bob Dolin, Daniel Puthawala, James Stevenson, Katie Stahl, Kyle Ferriter, Melissa Cline, Mkpouto Pius, Bob Freimuth, Salem Bajjali, Terry O'Neil, Tristan Nelson, Tim Hefferon, Wes Goar, Kori Kuzma

Apologies: Beatrice Amos (UK public holiday)

	Agenda Item	Speaker/Leader	Time
	Welcome and Announcements: - Ascona Connect - thank you to all the participants and contributors	Alex / Larry	2
	Ascona Connect review: - Any feedback / thoughts? - Please review meeting notes for accuracy and completeness, and add any key takeaways - Begin to think about milestones to target before Melbourne? - Goals and roadmap for VRS leading up to Connect, updates to the SV models, and TASC approvals.		
	Follow up on previous action items: - Review of classes from VRS2.0 that are implemented in VRS-Python (Kyle) - Review of project dashboard to simplify tracking changes and developments across branches. - Discussion of VRS recipes		
	Updates to the VRS 2.x documentation for community review		

Link to meeting recording:
https://us02web.zoom.us/rec/share/Hl1T00BATpjuvMTJVXB1O2ZteQHhMmfFxWcSCZZ6NZYSZCvK2_w-RcroMNrtqWmN.eLAC1qs-pY1CGrXN

Meeting summary:

(copied from chat re: molecule vs sequence)
Robert Freimuth to Everyone
suggested: variation IN the sequence corresponding to a molecule
ftr: I think the differentiation between sequence and molecule is very helpful
IMO: sequences can be associated with molecules. molecules can have biological relationships with each other. variation in sequences can have biological impact.

Here's the google doc Roadmap planning session at Connect Roadmap Planning

Ascona Connect Summary:
Larry provides a rapid summary of 5 sessions related to GKS:
Here are the recordings for reference:

GKS: You only implement twice	https://us02web.zoom.us/rec/share/5FZ1uO2t6sPUW4Gr3cE9N7OaNyK2Jmgvah Q2egMA4oTBWhXJawV8OzNVh596uDw.a4YbvFoCKC-Ep2X7
GKS: releases are forever	https://us02web.zoom.us/rec/share/ow4rOSXcU1S47fTpjeRbLpxUuzNF8i4LjH3M X4fNRKa5Flg9ndUTJbnvlXNpOia1.ZwRWZgiVREGNC7Vj
GKS: casino Cat-VRS	https://us02web.zoom.us/rec/share/dvatuDlx9ySUw9nRzhZ3XJeSCxplu3c33Pg_ MVKuQOutHIe0VW9YrsbPo0lAfweU.nM2XfV97qAJzOhKY
GKS - The Feature is not enough	https://us02web.zoom.us/rec/share/popZD0VINxxr-7xW4H5utwj3u3ArNIg-Ri8vw P4ekdpen43gaIBEqKNV03yQN9AP.rcxXNa16EaQipBAi
GKS - From the Roadmap with Love	https://us02web.zoom.us/rec/share/gmfEJt25MdmK0Ze-tltiCCULTDu0QH_5yQb me-8avxM56Jf9xTVul6VRvsRprAb_.QGlpv4ey5i0ZKq8n

Larry emphasizes the need for prioritizing implementers' needs, stabilizing the system, and advancing to VRS 2.0. He'll be focussing on aligning projects like ClinVar with the VA Spec in a 1.0 format and urges others to do the same for smoother data interchangeability.

Alex agrees with Larry the **importance of implementation-focused approaches** and the significance of prioritizing implementation-driven design, aligning with GA4GH and GKS goals.

Upcoming milestones and goals ahead of the Melbourne event:
Alex mentions feedback from Connect regarding implementation priorities and proposes two key points:

1. the need for more public documentation on VRS and GKS products and
2. simplifying adoption processes, such as through a "VRS in a box" concept.

The plan is to **present VRS 2.0 to the steering committee at Plenary** and outline plans for feedback and involvement from the community in the GA4GH product development process. He emphasizes the importance of **making the implementation process easier**, echoing sentiments from others who want a **more user-friendly experience**.

FHIR & GA4GH:
Bob Dolin raises questions about the role of pilots in integrating GA4GH with EHR systems, particularly in relation to FHIR. What about the respective roles of GA4GH and FHIR in clinical workflows? Is a demonstration of their interface necessary?

Larry says FHIR is important, but notes challenges in its implementation, particularly in the US. He sees GA4GH's focus more on knowledge management and research, with integration into EHR systems being a downstream effort. Alex adds that different standards serve different purposes, with FHIR focused on EHR data representation, while GA4GH standards facilitate genomic knowledge management. Bob emphasizes the opportunity to influence how data is represented in EHR systems before a potential content freeze in 2025, highlighting the need to align clinical data with public knowledge bases within this timeframe.

Bob D agrees with Larry's points about the complexities in the field, which prompt the need for piloting to navigate through such complexities effectively. There are challenges posed by the coexistence of GA4GH and FHIR standards in the US healthcare system and suggest that these complexities make piloting essential. He acknowledges the conceptual differences between GA4GH and HL7 but emphasizes that practical implementation often reveals valuable insights. He proposes promoting end-to-end pilots as a way to gain practical knowledge and address time constraints mentioned by Bob F.

Project board:
Alex introduces a project board designed to monitor ongoing feature development within VRS. It tracks issues, pull requests, and planned work, complementing discussions on the VRS discussion forum. Please contact Alex or Larry for permissions to contribute if needed.

Upcoming work on a VRS roadmap:
This will outline planned features and their development timelines. Please add comments on the project board.

Circular molecules as derivatives:
Discussion on the modeling considerations within the VRS framework, specifically addressing the representation of circular molecules as derivatives. Alex explains a simple solution involving adding a "circular" attribute to the derivative molecule class. Larry unsure about the approach, citing potential complications in computational identification and conceptual clarity. Discussion about the distinctions between sequences and molecules, with Alex advocating for the usage of "molecular" to define the physical object on which sequences reside. Discussion over terminology and conceptual clarity. The value of the "circular" flag on sequences is questioned and an alternative naming convention proposed. TBC.

Clarifying Terminology: Derivative Sequences vs. Molecules:
Tim points out a distinction between sequence and molecule regarding the representation of ring chromosomes. While the sequence of a ring chromosome can be defined with reference assembly coordinates, the molecule it currently resides on is unique because of its circular nature and lack of association with a typical chromosome assembly.
Larry expresses uncertainty about the need to introduce concepts like molecule and chromosome into the model, suggesting that VRS may be content with existing in "sequence land."
Alex explains the concept of molecular variation in VRS 1.3 and suggests continuity in the representation of variation concepts, but Larry questions whether molecular variation truly refers to variation on a sequence within a molecule in VRS 2.x.

Tim highlights the relevance of sequence variation in discussing balanced translocations and derivative chromosomes. He suggests that while new derivative chromosomes can be described in terms of sequence variation, the discussion about ring chromosomes may not be directly influenced by this aspect.

Larry suggests that the resulting product of such genetic events is a derivative sequence, as it is derived from the original sequence. How does this derivative sequence transition into being considered a molecule, especially considering that balanced translocations typically involve only a portion of the chromosome rather than the entire chromosome?

Alex acknowledges that derivative sequences may not always represent entire molecule and emphasizes the need for technical precision in terminology. Maybe a broader community discussion on the distinction between derivative sequences and molecules is needed? He'll set up a discussion thread to gather input from the community.

Technical precision in VRS's terminology is important here - Ale-rry (Alex n Larry) propose a collaborative effort to ensure clarity and consistency moving forward.

Bob (in chat): IMO: sequences can be associated with molecules. molecules can have biological relationships with each other. variation in sequences can have biological impact.

Bob: Maintaining a clear distinction between molecules and sequences is paramount in understanding phenomena at the molecular level. While these entities are interconnected, they serve different functions and possess unique attributes. This differentiation is not only crucial for observing biological phenomena beyond mere sequence data but also aids in comprehending the intricate relationships between various biological entities. Leveraging this distinction moving forward offers valuable insights and ensures precision in research and clinical contexts. Preserving this distinction remains vital.

VRS recipes
This aim is to compile common use cases and demonstrate how to utilize VRS effectively in various scenarios. Examples include representing CNV array results, genotypes at a locus, and HGVS expressions. Please contribute specific use cases relevant to your fields.

There's a plan to tag commits across the four GA4GH repos with a snapshot from Connect April 2024, ensuring versioned releases for users' reference.

2024-04-25

Hosting: Larry Babb, Alex Wagner
Attendees "Name (Affiliation)": Beatrice Amos, Kori Kuzma, Catherine Procknow, Wes Goar, Salem Bajjali, Daniel Puthawala,Kyle Ferriter, Daniel Puthawala, Terry O'Neill, Bob Freimuth, Mrinal Thomas, James Stevenson, Karyn Onyeneho

	Agenda Item	Speaker/Leader	Time
	Welcome and Announcements	Beatrice / Alex	2

	Restructuring of schemas, use of resolvable pURLs for VRS schemas		
	Draft PR for follow-up on #461 (PR to be created)		
	Documentation discussion (Discussion to be created)		
	AOB		

Link to meeting recording:
https://us02web.zoom.us/rec/share/MxDIMGt5XoiQUGzCsVn-AZYaG5A230lRi0WutIec6e4t8pVaofUryJd5xXoV650N.WLyOQ9NXx2MzMUQ4

⁉️Review day/time of future VRS meetings?

Meeting summary:

Restructuring of schemas, use of resolvable pURLs for VRS schemas:

Alex referred back to a presentation from a previous plenary in 2022 [find the link]. He mentioned his recent efforts to update and implement these concepts, particularly focusing on the W3ID persistent URLs proposed earlier.
This restructuring initiative extends beyond a single project, covering various GKS products. The goal is to achieve consistency across different areas, including the common library, VA spec, categorical variation, and potentially sequence annotation.
Plan is to bring these changes to GA4GH TASC. The aim is to establish consistent structures for persistence URLs.

Alex gave some background info. He explained the new structure implemented during the reorganization of repos.He demo'ed the implementation of this structure within the GA4GH VRS repo.

✅ Action item: Alex will export the document as an image and upload it to GitHub for accessibility. He mentioned addressing concerns raised about the difficulty of tracing properties, especially those inherited from other files.

From Slack:
First topic of discussion: https://github.com/ga4gh/TASC/issues/50
And the illustration of the use of the DerSeq and CPB classes in the discussion thread from the second VR call topic:
https://github.com/ga4gh/vrs/discussions/461#discussioncomment-8888676

Alex went into the class dependencies and organization within the repo -notion of dependent or aggregate classes. Is it necessary to create a singular JSON or YAML file containing all definitions for convenience and ease of access? Brief mention of namespaces and the idea of

consolidating all relevant information into a single comprehensive file or organising them per repository.

Kori (in chat): We might also want to consider “mega” for when we consider generating pedantic models from schemas (rather than manual) in vrs-python. Less clicking!

Discussion on how to structure and reference schema definitions effectively: the idea of using a single merged definitions file vs individual artifacts for referencing schema definitions. What is the process of constructing valid URLs for schema documents? How are URL redirections and updates for schema documents hosted on platforms like GitHub managed? Need to provide clear and accessible schema definitions while ensuring ease of maintenance and future updates.

Draft PR for follow-up on #461

Kyle, “why block” is answered here:
https://github.com/ga4gh/vrs/discussions/461#discussioncomment-8888676

Modeling derivative sequences and cis-phase blocks: review of the draft PR and discussion of the advantages of using cis-phase blocks over alternative methods.

2024-03-18

Hosting: Larry Babb, Alex Wagner
Attendees “Name (Affiliation)”: Beatrice Amos, Kori Kuzma, Tristan Nelson, Catherine Procknow, Wes Goar, Salem Bajjali, Daniel Puthawala,Kyle Ferriter, Daniel Puthawala, Terry O’Neill

	Agenda Item	Speaker/Leader	Time
	Welcome and Announcements	Beatrice / Alex	5
	Upcoming Discussion Overview: Sequence Reference serialization (#465)	Alex	2-3
	Upcoming Discussion Overview: Allele ordering in SV Haplotype (#461)	Alex	2-3

Link to meeting recording:
https://us02web.zoom.us/rec/share/t6WOg9YS1OMd9cBLd1F0ozrXpWtqlodkS77r1_HFiEABJLZE15xQ15xlmu2TJVgv.RjzsIlf3EK-8SBfW

Meeting summary:
The meeting focuses on two pinned discussions (up there in agenda) on the GA4GH VRS discussions board.

Sequence reference serialization: the devt from VRS 1X's sequence location class to VRS 2X's sequence reference class and the challenges of incorporating annotations.
Alex mentions various approaches - preference is leaning towards treating sequence reference as a standard value object. Also to think about is the handling of non-GA4GH identifiable classes used in value objects, exemplified by the genotype member class. Alex explained the digest serialization and sequence reference topic.

Draft SV model:
This includes a revised haplotype and adjacency model. Community feedback concern was on the potential misleading information due to the ordering of haplotypes and the complexity it introduces in computation. Also highlighted were issues with the computational definition of haplotypes and their deviation from traditional inheritance concepts. To address these concerns, Alex proposed 4 changes, including renaming the haplotype class to phase block and introducing derivative sequence, adjacency, and sequence terminal classes to add clarity and align with community concerns. Please add to Github for further discussion and refinement.

Sequence terminus class discussion. Alex clarified its purpose, explaining that it signifies the start and end points of a derivative chromosome sequence. The terminus class is necessary for defining the end points of a sequence, particularly when describing adjacencies.

Kyle asked about representing an inversion of the CD region without specifying its extension to the sequence ends, to which it was clarified that the default case allows for this representation without explicitly defining the sequence termini.

Larry asked about the class model, saying the sequence terminus should be separate from sequence location due to differences in their attributes.

The semantics of the terminus class and its distinctions from adjacency mentioned. Addressed questions round potential renaming of the haplotype class, the choice between derivative sequence and ordered haplotype as the containing structure, and the consensus regarding the adjacency structure, which was deemed solid and in need of minimal adjustment.

Foundational concepts in structural variant representation: derivative sequence could become the foundational structural variant concept, but adjacency remains fundamental, particularly in describing gene fusions (check this). Alex likened adjacency to gene fusion events and is the fundamental unit for most structural variant representations.

The renaming of haplotype will reduce confusion. (Wes - Discussion about building off these concepts into a genotype structure is deferred for a separate discussion because of its complexity. TBC).
Will move forward by creating a draft model based on the discussed concepts, which need to be tested and refined before advancing its maturity.

2024-02-19

Hosting: Larry Babb, Alex Wagner
Attendees "Name (Affiliation)": Beatrice Amos, Kori Kuzma, Andreas Prlic,Anastasia, Aleksandar Milosavljevic, Andy Yates, Beth Pitel, Daniel Puthawala, Kyle Ferriter, Rodrigo Martin, Salem Bajjali, Wes Goar

	Agenda Item	Speaker/Leader	Time
	Welcome and Announcements - Steering committee meeting feedback	Beatrice / Alex	5
	Connect session review	Larry	10
	Variants as Alleles or SVs (#453)	Alex	10-20
	Variant call metadata and CNVs (#445)	Alex	10-20

Link to meeting
recording:https://us02web.zoom.us/rec/share/H21aWxwj90Q3aBD5P4BdFnztDYj4gxprHXgbu6_0Zi3Wctxt_oj6nyzUkZFTKdcD.SdcvvKFZUh4foeoC

Meeting summary:
Larry discussed the proposed sessions for the upcoming Connect meeting:
GKS Connect Sessions 2024

Session A: will focus on the release process for GA4GH products, with an emphasis on separating the maturity and release process from product delivery.
Session B: will be about implementation and showcase tools and challenges faced by early implementers.
Session C: will focus on categorical variations, addressing modeling work and establishing baseline classes.

Variants as Alleles or SVs (#453)
Alex talked about policy or implementation guidance regarding the use of adjacency or legal representations in the context of structural variations, particularly in VRS2.0 development. The

issue arises when deciding whether to represent a sequence variant as a small variant or a structural variant. Some groups have used arbitrary length cutoffs to distinguish between the two representations, but there's a preference to avoid such cutoffs.

Aleksandar highlighted distinctions between structural variants ascertained by different methods, like arrays vs sequencing, leading to different semantics. In older versions of VRS, molecular and systemic variations were disambiguated, but the current focus is on how to represent sequences composed of multiple references in VRS 2.0.

Need to clarify the purpose and necessity of implementation guidance regarding the use of adjacency or legal representations for structural variations. Larry asked about the value of this guidance and whether there's harm in not providing it.

Having two different data structures for representing the same resultant sequence state could lead to potential confusion. What about the practical aspects of how various callers operate independently and how downstream implementers may interpret and apply the guidance?

Larry in chat:
Alex - maybe this should be deferred until we take a look at our “cookbook” documentation idea which would provide guidance within different breakdowns of variation.
This is Andreas “brainchild” … We could start with the types as defined by pacbio here https://www.pacb.com/products-and-services/applications/whole-genome-sequencing/variant-detection/

Beth Pitel in chat:
HGVS nomenclature only combines variants that are in the same amino acid or in adjacent base pairs. Personally, I don't think we should use anything based on this just thought nomenclature standards might be interesting to compare.

Important to provide implementation guidance through specific examples and scenarios. There is value in linking guidance to practical applications, such as through notebooks demonstrating common use cases. Beth raised a point about the selection of the 50-base pair threshold by ClinVar and suggested considering the blatability of variant sequences as a criterion for differentiating between alleles and structural variants.
Also some challenges related to repetitive DNA regions and the educational value of creating a resource for learning about genetic variation.

Andreas in chat: For our cookbook, one of our examples should be how to represent FMR1 repeat expansion variants: https://medlineplus.gov/genetics/gene/fmr1/

Larry in chat:
Do we need to establish a GKS “policy committee” that has the right group of folks that would be able/willing to tackle these guidance recommendations?

I think we wear a lot of hats on these calls, but I think we should separate the engineering/design/modeling focus from the polic/guidance focus. At least for me these are two areas that require different skill sets and experience. It seems hard to do both at the same time and it feels as if the policy and guidance is downstream of the engineering/modeling.
The details at least should be broken out into special sub group discussions (maybe)?

Determining thresholds for categorizing alleles versus structural variants, with considerations for sequence length is complex. Alex shared insights from the Allele Registry's approach to representing structural variants, focusing on abstractions such as positions and attributes. Daniel keen to incorporate these concepts into the work of the Categorical Variation Group. Alex summarized - generally all were against concrete threshold proposals and a preference for developing a "cookbook" of representations linked to the Verse 2.0 data model. Please contribute your view on GitHub.

Variant call metadata and CNVs (#445)
The issue raised by Mrinal about representing average copy numbers across regions, which are essential for interpreting data from bulk tumor samples. The current copy number count class is insufficient for this purpose, so need a separate specification or area within the specification to handle average copy numbers. The analogy was drawn to variant allele frequency in VCF files, which are crucial for filtering strategies in cancer research but are not currently captured in VRS.

Beth is chat:
We use this kind of metric to estimate mosaic gains/losses in chromosomal microarray where we have large chromosome portions that we report as a single item.

Alex proposed collaborating with other work streams, such as the one focusing on QC metrics, and engaging experts like Matt Brush and Daniel Cameron, who have experience with variant representation formats like VCF. A suggestion was made to potentially establish a new group or integrate the discussion into existing efforts to ensure comprehensive coverage of variant call attributes. Action item: Alex(?) will reach out to relevant individuals and groups to initiate further discussions and potential standardization efforts.

2024-01-29: Connect meeting planning

Hosting: Larry Babb, Alex Wagner
Attendees “Name (Affiliation)”: Beatrice Amos, Kori Kuzma, Alan Rubin, Tierra F, Daniel Cameron, Beth Pitel, Andreas Prlić, Melissa Cline, Terro O’Neill, Bob Freimuth, Kyle Ferriter, Salem B, Mrinal Thomas, Daniel Puthawala, James Stevenson.

	Actions Arising	Assigned To	Deadline
	Please review upcoming GA4GH Connect meeting planning document	Leads	Feb 16th to submit the sessions
	governance process for GKS.common	Bob and leads	

	Agenda Item	Speaker/Leader	Time
	Welcome and Announcements	Beatrice / Alex	5
	Progress update: Introduction to the GKS Commons	Alex	15 + Q&A
	Progress update: Maturity model in VRS 2.0	Alex	15 + Q&A
	Connect meeting planning	Alex / Larry	
	Upcoming Discussions: #453 and #445	Alex	5
	AOB		

Link to meeting recording:
https://us02web.zoom.us/rec/share/dEp7N2RFTIX0vgpbXNxUxWtBjNEQeMa-ZFrJlAtG4rO47TIe-KNz4ObvRzNAnbzL.dm2unVf7s0bNgVoJ

Meeting summary:
Please can leads think about the upcoming GA4GH Connect meeting. Please share session proposals, and review the roadmap's deliverables (doc will be made available by email).
Alex discusses GKS Commons repository, and explains the integration across GA4GH products. Updates on the maturity model and a review of branching and merging strategies were discussed.
GKS Commons repository - Alex discusses core structures, emphasizing the importance of a common space for collaborative maintenance and development. Alex talks through the recent move of the core source work to the GKS Commons repository, including the existence of a maturity aspect. DanielC raised a question about extensions, its schema, and potential conventions. A proposal was made for a Connect session to address the placement of annotations for genomic rearrangements, specifically focusing on determining the appropriate specification for such data within the GA4GH framework.

Bob in chat:
“**These are exactly the types of questions we need to address at the WS level. We need a cross-product governance mechanism for common objects like this. We don't have one (yet), but I hope we can create one soon. I'll think about it, and welcome anyone that is interested in doing the same.**”

The **need for a governance process to manage elements within the workspace is needed**, particularly those originating from specific specifications that may require broader use. Bob supports the idea of implementing guidelines and governance for cohesive maintenance. He acknowledges the absence of all necessary guardrails but emphasizes the importance of establishing them.
We all agreed on the need for a mechanism to **identify and promote elements for wider use**. Bob suggests Daniel bring examples of structural variant extensions to a meeting for discussion and recommends starting with an issue ticket. What about the structural variant work's place within the GA4GH framework? Does it belong in the VRS space or elsewhere?
We need to discuss and resolve these issues in upcoming meetings, possibly leading to a dedicated session at GA4GH Connect?
Need to encourage participants to provide feedback and examples related to structural variant work to ensure the model aligns with their use cases. BA- what about one last comms drive?

Action item:
Beatrice: could you please add "governance process for GKS.common" as an action item?

Alex talks about how the documentation for GKS Commons will be disseminated. Alex says the suggestion that stakeholders need to be engaged in the discussion to determine how to manage documentation for GKS Commons is a good one. The technical process for schema development is in place, but details about documentation and maintenance are yet to be worked out across the GKS teams.

Maturity model and versioning strategy for GA4GH specifications, particularly VRS: Alex provides an overview of the previous discussions around maturity levels and the complexity of managing different branches for each minor version. **The decision is made to simplify the model by having a continuous 2.x branch**, ensuring **backward compatibility** and using GitHub releases for alpha, beta, and release candidates. This change is seen as more intuitive for users familiar with GitHub workflows.

Connect planning:
Connect meeting and scoping for the beta release of the VRS specification. Larry says we need to focus on key topics during the 2-day event and suggests keeping momentum on structural variant work. **Larry plans to deliver a beta version at Connect** and says we need documentation, schemas, and updated tooling.

Larry discusses engaging with pilot projects, including groups like BRACA Exchange, Mass General Brigham, and Epic, to ensure that VRS meets their use cases. **Goal is to have a stable identifier for diverse alleles by September 2024**. Sequence location, allele, adjacency, and haplotype will be included. What about genotype - should it be part of the beta release - complexities and lack of pilot testing.

Andreas suggests releasing what is currently available with confidence and flagging features that lack confidence. Larry and Daniel agree that haplotype should be included in the beta release to engage the structural variant community. Andreas notes that InVitae may be interested in structural variation and is less likely to adopt new features until they become more stable. Agreement to include haplotype in the beta release to engage the structural variant community.

Larry proposes focusing on including haplotype in the beta release and leaving genotype as a potential addition later. The timeline for achieving this target is set at around 2 months before the Connect meeting. There is a need to prioritize and showcase real-world applications of VRS in various contexts, including implementations with VA spec, ClinVar, and other genomic databases.
Alex suggests including a technical forum or a parallel session for developers to engage in a hackathon-style environment. BA - create a list of poss attendees, both virtual and in-person, to understand their technical and strategic interests?

The overall goal at Connect is to present a cohesive and comprehensive picture of the progress made with VRS and its integration with other genomics standards.

Please can everyone check the agenda and review GitHub issues that will be discussed in future VRS calls. Alex says the intention is to make VRS calls a regular mode of operation to keep track of ongoing and upcoming tasks.

2023-01-15: Canceled: Public Holiday

2023-12-04: VRS 2.0 alpha to beta approval process

Hosting: Larry Babb
Attendees "Name (Affiliation)": Beatrice Amos, Tristan Nelson, Daniel Puthawala, Kori Kuzma, Wes Goar, Mihai Todor, Katie Stahl, Catherine Procknow, James Stevenson, Vivek Nuthalapati, Salem B, Bob Freimuth, Tim Hefferton, Natassa, Kyle Ferriter, Andy Yates, Mrinal Thomas

	Actions Arising	Assigned To
	https://vr-spec--455.org.readthedocs.build/en/455/index.html - review and submit comments on the proposed classes and maturity model process.	All
	Repo - needs further cleaning and organization - Daniel P.	

	Bob: We need a visual representation or a live dashboard to simplify tracking changes and developments across branches. Maybe create a simpler table or representation to facilitate understanding and tracking of ongoing changes.	Kori?
	Ensure the classes from VRS2.0 alpha are implemented in VRS-Python	Kyle, Wes and other VRS-Python developers

	Agenda Item	Speaker/Leader
	Review the github repo reorganization	Larry
	Review the policy for implementing a data class maturity model and VRS versioning	Larry
	Discuss the VRS 2.0-beta classes and documentation scope and plans to target Connect in April	Larry
	AOB	all

Link to meeting recording:
https://us02web.zoom.us/rec/share/pUSVdB63qmQTmMyIXrEglnomQLBzVIlm4sXJLrrA267iZ6LG1zcb32dDNIdMwjJH.D8bieSTk4NeQkdG6

Minutes:
https://vr-spec--455.org.readthedocs.build/en/455/index.html

Larry demos changes made in SV-VRS-2.0 preparing for the Connect meeting in April and discussing the scoping down of classes and entities within the VRS to reach the beta stage by that time.
Goal is to **update documentation, pilot implementations**, and get approval from the group.

Please attend the SV meeting on Thursday for further updates. The GitHub repos have been reorganized to support a maturity model plan, and specific classes like allele, haplotype, and copy number change are being considered for the beta stage. Genotype and chromosome location are undergoing discussions and potential adjustments. Larry and Alex's primary focus currently is on sequence location and related fundamental classes, aiming to move them to the beta stage. Hope is to refine the GitHub repos based on practical use cases and real data pilots before finalizing certain classes for the beta release. Larry suggests for the group to go through doc and **raise issues in GitHub**. Will be reviewed by the team.
Larry's intention is to establish a clear process for maturity levels and versioning and refine the documentation collaboratively.

Bob: We need a visual representation or a live dashboard to simplify tracking changes and developments across branches. Challenge is in maintaining it.
Maybe create a simpler table or representation to facilitate understanding and tracking of ongoing changes. TBC. Kori?

Larry goes over 3 repos:
CAT-VRS will replace VRS-atile which allows more flexibility in adding attributes and creating categorical variants. Repo - needs further cleaning and organization - Daniel P. VA Spec is a comprehensive knowledge statement model, encompassing more than mere variant annotation. It includes a broader framework and various statements like variant pathogenicity and therapeutic responses.

Larry talks about reorganizing the VRS Python documentation into different types: base abstract types, general purpose types, and special purpose types. The base abstract types hierarchy includes various classes inheriting from one another. Need to provide clearer documentation.

The general purpose types consist of complex data types usable in multiple scenarios, with incomplete documentation yet. Hopes for beta inclusion. Larry us organising the classes in the new version of the spec, breaking them out from the long page with all the classes in the old version. Alex reviewing, but others urged to jump in too.

2023-11-20

Hosting: Larry Babb, Alex Wagner
Attendees "Name (Affiliation)": Kori Kuzma, Alan Rubin, Vivek Nuthalapati, Daniel Puthawala, Terro O'Neill, Catherine Procknow, Kyle Ferriter, James Stevenson, Alan Rubin, Salem B, Beatrice Amos, Mrinal Thomas, Amanda Spurdle, Bob Freimuth

	Agenda Item	Speaker/Leader	Time
	Announcements		5
	SV with VRS 2.0 alpha	Alex, SV team	15
	Upcoming work: issue review & community feedback	Alex	10
	Updates on Cat Var SG and Cat-VRS	Daniel	15
	AOB	Larry	

Link to meeting recording: (Will be shared before Weds 22nd Nov)

Link to slides presented: alex.wagner@nationwidechildrens.org

Meeting summary:

Agenda will focus on discussing the work done with the SV team and aligning their progress with the VRS 2 alpha specification.
Alex summarises different branches of work, particularly the VRS 2 Alpha and SV alignment work, merging the SV team's efforts into the VRS 2 Alpha branch. There is ongoing work affecting things like haplotype and data class names. Ongoing discussions in ad hoc calls.

Please join next SV call for more detailed discussions - Dec 7th. Add to calendar.

SV summary:
Alex demos detail about changes planned for the VRS 2 alpha model to accommodate SVs (structural variations). Also various alterations, such as transitioning to an IRI-based architecture for references (not CURIEs), introducing new concepts like length expressions and adjacencies (break ends), and their significance in capturing ambiguous sequence length expressions and junctions between sequences. Important to understanding adjacencies in comprehending the new haplotype structure, which is now an ordered set of different types of molecular variations.

There is a significant **change in the haplotype structure**: transitioning from an unordered collection to an ordered set, this will facilitate capturing and understanding information computationally. Importance of ordering for representing concepts from the SV group effectively within the haplotype structure highlighted. Very complex and there is a need for precision in articulating these changes.
Alex introduced a test definitions file to provide examples of these models in application, aiming to support better understanding and implementation. Yay - Kori will be happy! Mrinal's example slide deck presented during an SV call shown, intending to illustrate these changes with visual aids.

Alex emphasises the usefulness of ordered attributes within the SV model to enable a concept called "sequence transitivity." This concept allows for the representation of molecular variations compactly within the Haplotype structure.
Alex shows more complex structures, like a complex haplotype example involving multiple sequences, adjacencies, and alleles. The emphasis is on the compactness and expressiveness of the new Haplotype structure compared to simpler examples, demonstrating how sequence transitivity enables concise representations.
Daniel - keep the balloons coz we all love 'em!

How do we represent inversions within the sequence model: expressing inversions in terms of sequence locations and the transition from interval notation to adjacency. The representation involves expressing either the start or the end of a sequence terminal, not both. E.g. if

describing a break at the beginning of a chromosome, one might leave the attributes empty to indicate the very start of the sequence.
Larry adds that by simplifying the representation of inversions, stating that in version 2.0, the start can be greater than the end, indicating a range of sequence inverted in the reverse complement.
Inversions: there is a shift from interval notation to adjacency notation and a simplification of expressing inversions within the sequence model.

Amanda discusses challenges related to integrating cyber-genetics work with sequencing work within an Australian diagnostic meeting. The concept of ring chromosomes and how they might impact sequencing representation, particularly regarding the start and endpoints mentioned.

Discussion on how chromosome locations model variations and the ongoing discussions about representing reference molecules, whether circular or linear. Amanda raises questions about how to handle ring chromosomes within this model and seeks guidance on whom to engage with regarding these challenges.
Larry suggests raising **GitHub issues** on the VRS platform for discussing and resolving these concerns.
How do we address the assumption about sequence equality between haplotype members and the reference sequence? It's clarified that the assumption of sequence equality shouldn't be made, and the representation should be explicit about differences rather than presuming similarity.
Overall, the transcript captures a comprehensive dialogue around representing complex genetic structures like ring chromosomes within the context of sequencing, along with related considerations like derived sequences, sequence equality, and model representations.

Daniel's Cat Var Summary
TBC

2023-10-16: Canceled

2023-09-04: Labor Day

2023-07-24: Review of VRS 2.0a changes

Hosting: Larry Babb, Alex Wagner
Attendees “Name (Affiliation)”: Alex Wagner, Larry Babb, Wes Goar, Melissa Cline, Mrinal Thomas (Epic), Catherine Procknow (Epic), Vivek Nuthalapati (Epic), Kyle Ferriter, Bob Dolin, Anastasia, James Stevenson, Kori Kuzma, Salem Bajjali, Wes Goar, Bob Freimuth, Beatrice Amos

Apologies: Andreas Prlić

	Agenda Item	Speaker/Leader	Time
	Review of VRS 2.0a changes ● Discussion of changes to the schema to accommodate descriptive fields ● Review of planned changes in VRS Python Initial discussion around forming a GA4GH Categorical Variation specification ● Review concept with examples from ClinGen and VICC ● Discuss concept of “members” and mappings to categorical variation		

Link to meeting recording:
https://us02web.zoom.us/rec/share/Q8WE37Ejmf8BsKVVDdmRuqmGHBbHyY7ev075icvK4FRff1fb4WCrJudYLHmmhm_r.qdWn2CxjQei5ofRK?startTime=1690232115000
Passcode: *n3=DDUt

Minutes:

Agenda - working session on VRS2.0 alpha

✅ BA - to send details to leads about categorical variation study group, (and comms - add interested participants to new group mailing list).

Discussion of some of the changes on https://github.com/ga4gh/vrs (2.0 Alpha branch) to prepare for computed digests on variation classes that now have descriptor elements.

Kyle to add GitHub issue:
What to do if there's a mix of attributes or if a mix of attributes is allowed between identifiable versus non identifiable objects. If you could put both of them for example alternatively into a slot. (But could be both repetitive and error prone to specify at the containing class level, what fields are expected to have digestible identifiable objects.)

Review of plans changes in verse Python:
Moving to a model where our verse classes are no longer just imported and instantiated at runtime. Was used in VRS-Python using the Python JSON schema Objects library. Technical challenges including library insufficiencies, causing downstream conflicts etc.

Not using it anymore, and taking VRS2.0 alpha and implementing as *pedantic* data classes directly in VRS-Python. Has performance and flexibility advantages.

Anyone interested in participating can reach out to Kyle Ferriter(?), Kory Kuzma and they can tag you into the relevant branches, commits and PRs.

Demonstration of ClinVar dataset (LB):
Example ClinVar variant:
https://github.com/ga4gh/va-spec/blob/metakb-cvc/examples/clinvar_ca_example1.jsonc

Demonstration on the variant portion of a Clinvar variant and how it would manifest itself in a JSON structure.

Definition of categorical variation (AW):
What is needed to support categorical variation is a clearly defined semantic representation of assay variation that can be plugged in and used inside of a secondary encompassing spec for the type of subjects seen in genomic knowledge statements. I.e. categorical variation is like a domain of relevant variants that can match to a concept.

2023-07-10

Hosting: Larry Babb, Alex Wagner
Attendees "Name (Affiliation)": Tim Hefferon (NCBI), Samriddhi Singh, Tristan Nelson, Vivek Nuthalapati (Epic), Zhenyu Zhang, Salem, Martin Stoves, Ion Phan, Kyle Ferriter, Kori Kuzma, Kevin Riehle, Daniel Cameron, Bob Dolin (Elimu Informatics), Andy Yates, Ryangguk (Rick) Kim (Oak Bioinformatics)

	Agenda Item	Speaker/Leader	Time
	Welcome, Introductions, Housekeeping - GA4GH Connect 2023: San Francisco - Call for Sessions: Please let beatrice.amos@ga4gh.org know if you have a session in mind or https://forms.gle/qqGRiiz441LgtyQP7 (Connect session proposal form)	Beatrice	5 min
	Use case presentation	Ryangguk Kim, PhD CEO, Oak Bioinformatics, LLC	5-10mins
	Continue to review VRS 2.0 alpha plans with a focus on changes within the sequence expression area of the VRS model Review of VRS 2.0a Locations, Alleles, and Haplotypes	Alex	
	Explore use-cases, and workflows how people want to use V2	Andreas	
	Road map review (keep on agenda)		
	AOB		

Link to meeting recording:
https://us02web.zoom.us/rec/share/WL5JWxfA3fppfA8goZ7knbG1WgT-F89wXDJnu28YHlMaqV7vuOxMQQoHllgrQEB6.tPzVefa76zkv-Tqd?startTime=1689004600000
Passcode: .UIT!eN7

Minutes:

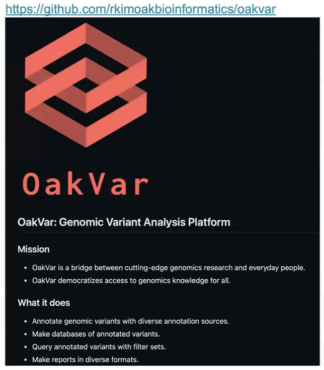
Goal of today's meeting is **review work going into VRS2.0 alpha**. Next week: discuss use cases people would like to see covers by VRS. Please send use cases, and areas of development.

Rick Kim (Oak Bioinformatics, Indonesia) presented use case:

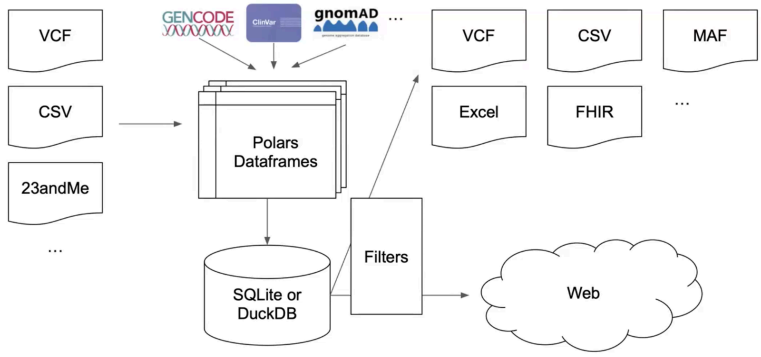
Ryangguk Kim - OakVar is a variant annotation platform and we are implementing a VRS reader to ingest variants written in the VRS format into the annotation pipeline. We have some questions regarding parsing and processing VRS files.

[GitHub - rkimoakbioinformatics/oakvar: Genomic variant analysis platform]

- Variant annotation platform
- Library/CLI tool/web server included
- Extension by Python modules



Annotations are collected, made into data frames, stored in a database, application of filters to produce output files:



Our thoughts so far

- We want OakVar to be able to read variants in VRS format.
  - JSONL files of VRS format?
- Also, thought about a writer module for JSONL files of VRS format.
  - Is a JSONL file of VRS format used / needed?
- VRS is
  - Variants themselves only?
  - Can include annotations (allele frequencies, pathogenicity, etc)?
  - If variants only, other GA4GH formats suitable for delivering annotated variants?

LB: there is a plan to represent population allele frequencies and other types of variation statements in VRS2.0.

Other groups suggested to explore within GKS include the VA group (waiting for VRS2.0 out). Suggests looking at SV group too. BA - to send email to Rick with meeting details.
AW: suggests remain engaged so use case can influence development of specification, as lots of potential to grow and develop at the initial stages.
RK: will collate and summarize use cases from 150 modules and share with the group (Github?).
https://demo.oakvar.com/
LB: use case demonstrated by RK is not uncommon, so VRS2.0 development work will be aligned.
AW: requests everyone with use cases to share, in order to shape and drive continued development of VRS2.0.

VRS2.0:
AW demonstrates what dev is going on in the schema:
https://github.com/ga4gh/vrs/tree/2.0-alpha
Discusses new types of **sequence expression and associated work required for normalization**.
Most dev work will be on new github branch - (some merges still required from main).
Discusses a way of deciding when you use a *ReferenceLengthExpression* and when you use a *LiteralSequenceExpression*, and what are the rules around creating a computed digest when you're using things like ReferenceLengthExpressions.

Sequence location, the length and the repeating unit length provides a much more compact representation of large number repeat units or repeats of very large sequence locations. Also overcomes challenges of very large tandem duplications, for example.

DC:
What's the general approach for choosing a variant representation? Will there be any guidance around not only normalization of the representation there but normalization of the alternate forms of representation you can do?
AW:
Yes. Problem in current VRS and in other formats.e.g. an allele over a repeated sequence can be represented in multiple ways.
One of the trade offs when designing the specifications. We can provide a single way to do everything, but that way is going to be very inelegant and cumbersome. So we need to provide alternate mechanisms. And when we do that, even if we've got good conventions for picking one or another that are very deterministic.They don't always get followed appropriately. And we've decided that the one of the best things we can do is provide **clear guidance**.

LB:
VRS structure has changed, so all to ensure development of subgroups aligned with dev in VRS2.0. E.g need to work on semantics/ terminology, e.g. definition of haplotypes in SV and VRS.

DC: What is the context around the variants? What is the meaning of that context, which isn't clearly defined in VRS or VCF.
SVs break many assumptions and terminologies. Have to redefine it, such that the traditional meaning still holds true in a standard deployed genome. But when there is structural variation, it still makes sense.

LB: we want a single model that can span across sequence and SV. Is there a certain set of classes at the lower level that are appropriate for the kind of details that SV needs, that maybe some SV don't need, but it's tolerable in terms of putting different kinds of low level structures in there to make.

AW:
One of the challenges is that different communities use different definitions of what a concept is.

LB/AW/DC:
Meet to discuss sequence location, break- and break points, period at next SV meeting. BA will add to agenda.

AW:
Thoughts on registering and resolving the **GA4GH namespace**, which is an organizational wide namespace as an organization. Will be discussed in TASC..

Note: HL-7 FHIR - created their own FHIR reference concept and they have since they're building a spec that has to do with building a server and compliance etc, they could do that.
VRS - went with JSON pointer style URI reference data type that JSON scheme offers. Good solution for now.

BD: in FHIR you can point explicitly to another object's ID but it also has this indirect pointing where you can almost point to another resource, like by an Xpath expression. So as long as there's some hook, there's a way to point to it.

2023-06-19: CANCELED (Public holiday)

2023-06-12: CANCELED (Due to the unavailability of the leads)

2023-05-22: VRS 2.0 working meeting

Hosting: Andreas Prlic, Larry Babb, Alex Wagner

Attendees "Name (Affiliation)": Kori Kuzma,Michael Baudis (UZH),Vivek N,Kevin R,Samriddhi S,Salem Bajjali,Hangjia Zhao (UZH),Tristan Nelson,Likhitha ,Katie Stahl,Anastasia,Kyle F,Andy Yates,James Stevenson,Tim H,Wes Goar,Reggan Thomas

	Agenda Item	Speaker/Leader	Time
1.	Working meeting : Continue to review VRS 2.0 alpha plans with a focus on changes within the sequence expression area of the VRS model	Alex / Larry	50 min
2.	A.O.B	Group	10 mins

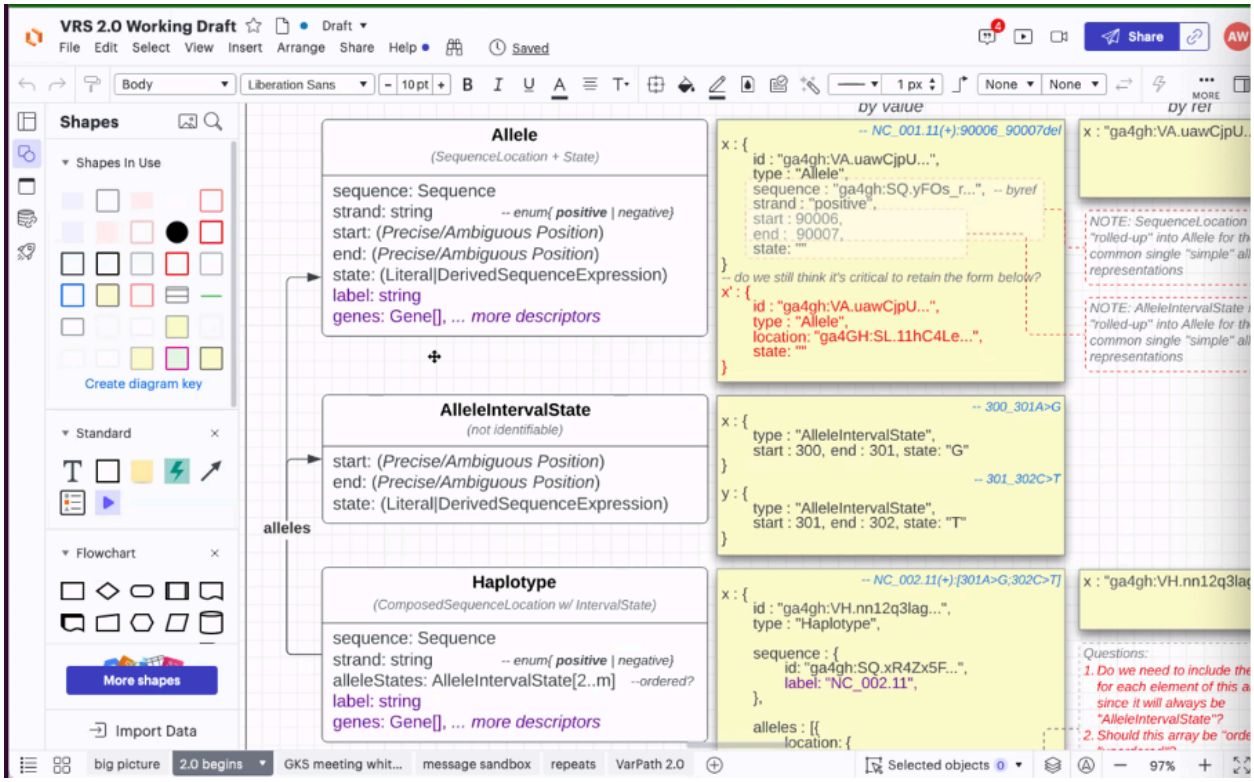
Meeting recording:
https://us02web.zoom.us/rec/share/f1RT8jfxMAls2tfLhCw5bKoBZjrtlnV66ULr0NlPil3YN3glf0QsKN2r0pgd6dRP.TfrRiVbXw_R2e3pc
Passcode: je!A6@pX

Minutes:

AW- **https://vrs.ga4gh.org/en/stable/terms_and_model.html#repeatedsequenceexpression**

Here an example of V1, how repeat-expansion variants could get represented. This is a test example, but that's how I would express "FMR1 has 60 copies of the CGG repeat of clinical interest"
https://github.com/ga4gh/vrs-python/blob/6a70bdea4e5514834fb5fbd4f99fe84926ff2dd8/misc/debugging/sometimes-referrable.py#L23

Walkthrough of VRS 2.0 Working draft.

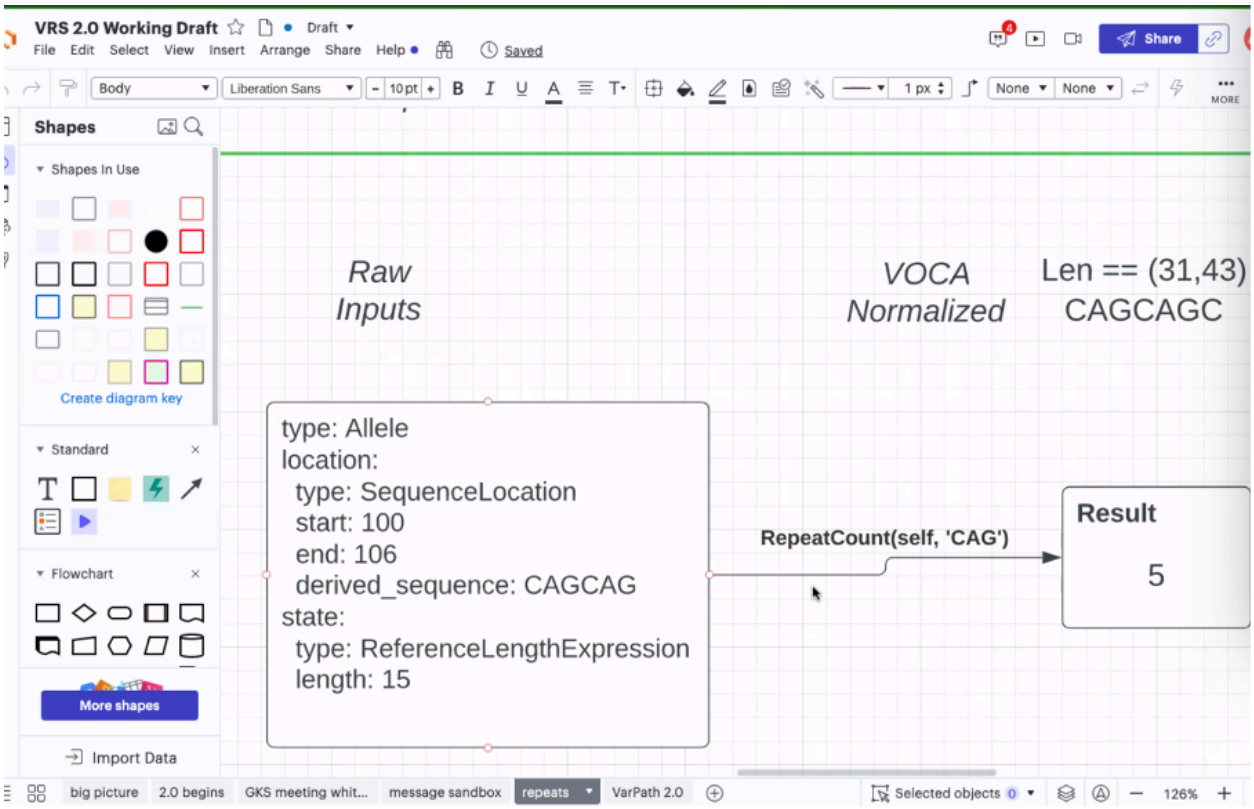


Repeated region expression:
Reference region – repeated CAG

Length of repeated seq are not visible.
CAGX5
CAFCAG
How to use of repeat in V2.0
Instead of repeating a substring – represent as integer.

Works for ambiguous length – Appropriate seq location and length.

Know repeated seq –
Allow for translation without seq repo / external data



AP – Use cases?
LB – agnostic for fragile x repeats.
AW –
TN – The main use is going – having ambiguous
Similarity and not identity.
AW – cnv matching approaches.
MB – Process of representing the repeat. By schema representation.
documentation,
KR – copy number aggregation – using descriptor nomenclature

Looking for reevaluating.
AW - Inter-residue coordinates: same as BED and SPDI
LB – seq position, deletion
Which of the records can be applied?
LB – applied to normalization. All cases except the seq location in the residue point.
Seq location is zero. It won't be applied

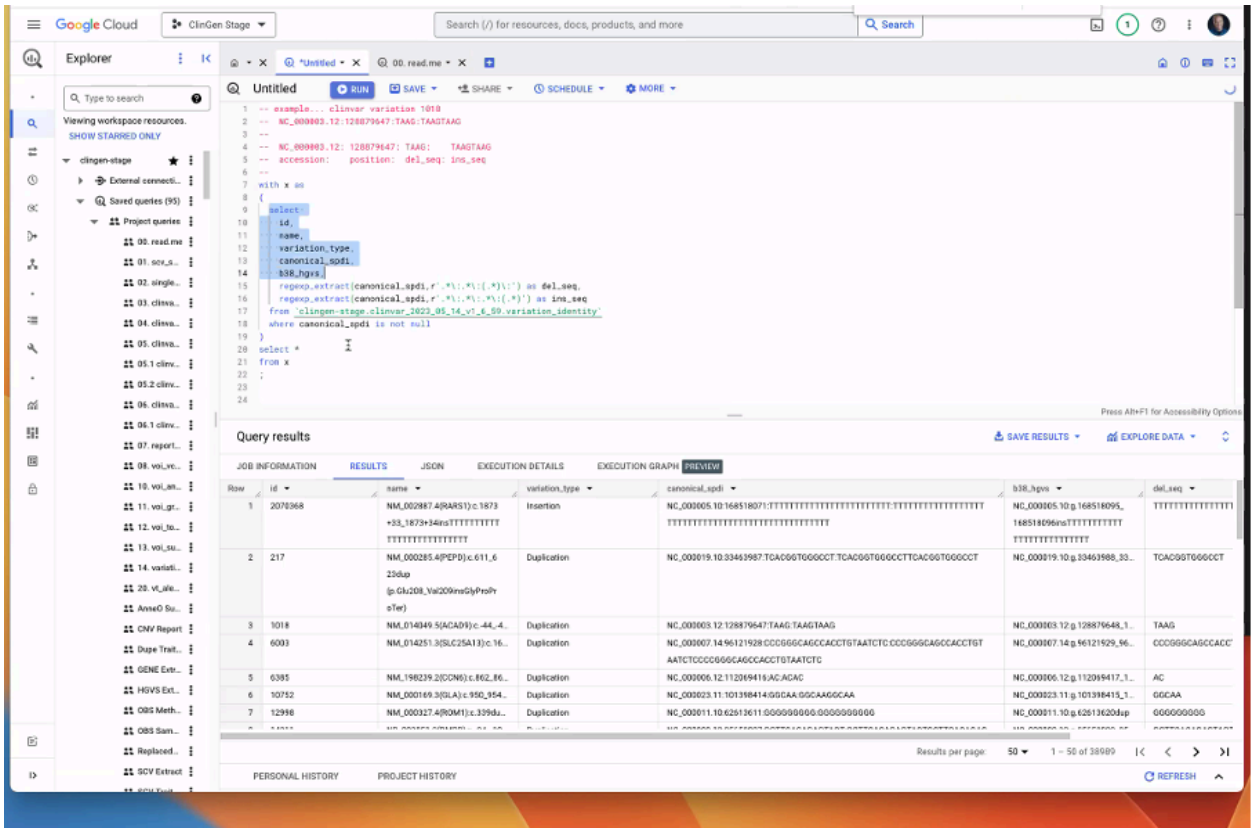
AW - VRS has a seq location. The range of seq location is zero.
LB – how to get relative expression from SPDI?
AW – exclusion by substitution .

LB – when normalization – repeating is not a true repeat.

How we would take this expression.
AW – We got a canonical SPDY here. This represents the case of what is going on . need to investigate. Make a github issue.
AP – Fine with both series.
LB – we have an empty string here now. Needed for allele.
AP – Larger insertion.
AW – need to come up with a role that consistently applies
LB – if we have repeated but extra bases added.

VN - @Larry- If we use a RLE of 0 to represent a deletion, then would we also allow using a RLE of 1 to represent a state when it's the same as the reference sequence at that location?

MB - We have people here working on STRs (w/ unit lengths 1-6 or such) where the definitions are rather computational, so kind of nicely suited to VRS representation (I'd hope).



2023-05-15: VRS 2.0 working meeting

Hosting: Andreas Prlic, Larry Babb, Alex Wagner

Attendees "Name (Affiliation)": Likhitha S, Kori Kuzma,Wes Goar,Kyle Ferriter,Melissa C,Emma T,Anastasia,Salem B,Mrinal T,Alain R,Beatrice A,Reggan Thomas,Zhenyu Z,Amanda S,Catherine P,Vivek N,Daniel C

	Actions Arising	Assigned To	Deadline

	Agenda Item	Speaker/Leader	Time
1	Working meeting - VRS 2.0	Alex/Larry	50 minutes
2.	A.O.B	Group	10 minutes

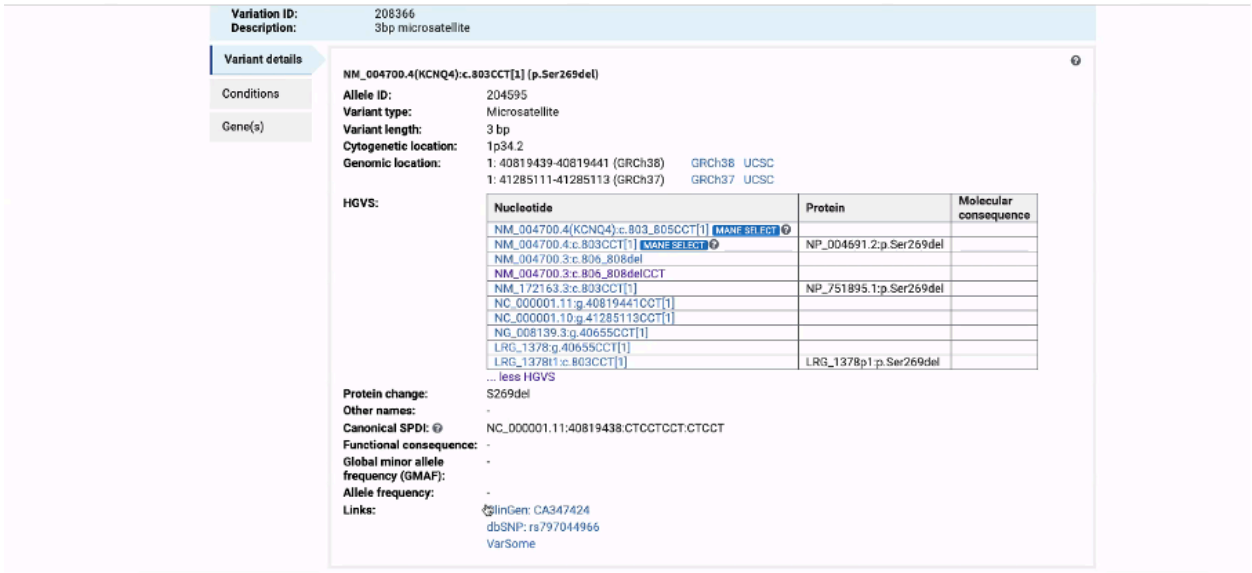
Meeting recording :
https://us02web.zoom.us/rec/share/YngzYN09SknEwkRZqsa2XLz4MasepverGbw_7mFH78tnAsIpXrPgF02c--2UTcD4.TvSd7bIDHIG7tZc5
Passcode: W^XZhJ!3

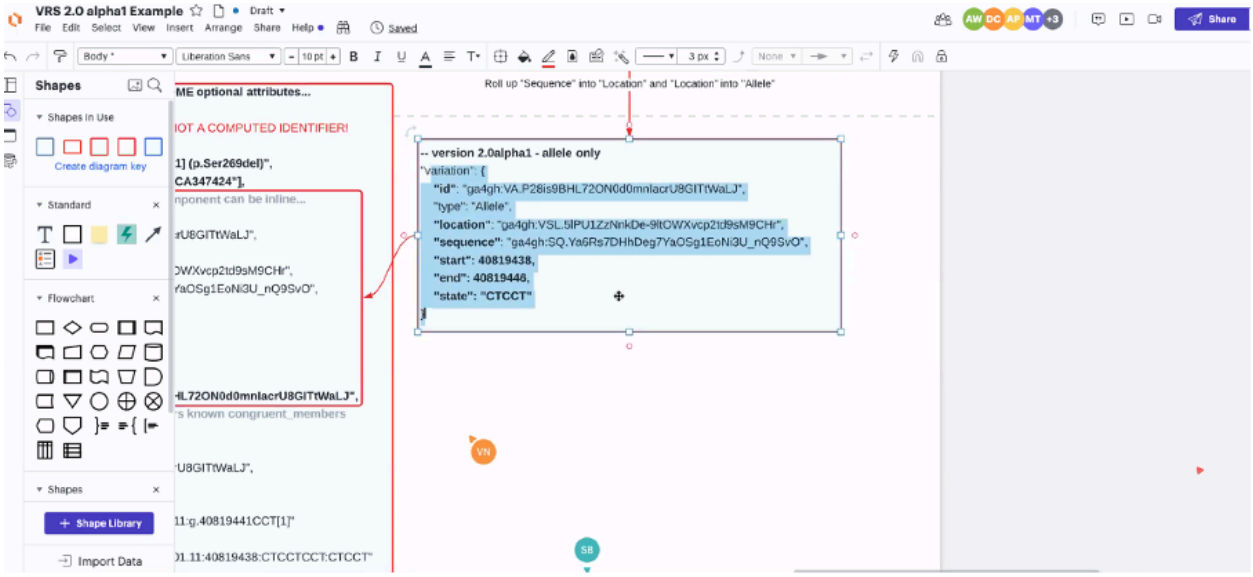
Meeting minutes:

AW – Discussion on VRS 2.0 Alpha 1

- Idea is to discuss the idea how are sandboxing
- Team to take a look and provide and feedback

AW - **https://github.com/ga4gh/vrs/issues**
LB - **https://lucid.app/lucidchart/18646cc7-82f4-4be0-b3da-8cfadbf2e6a9/edit?invitationId=inv_24899ed2-c4f4-4a8e-904a-e1546ebe0fd8**





LB – Balance between pragmatic engineering

LB – Decided that they are not pragmatic for both use cases.

MC – What does the state actually means
LB – HGVS –
MC – reference or alternative state?
LB – Ref state
LB – Calling it as canonical variant. One orgination contextual value.

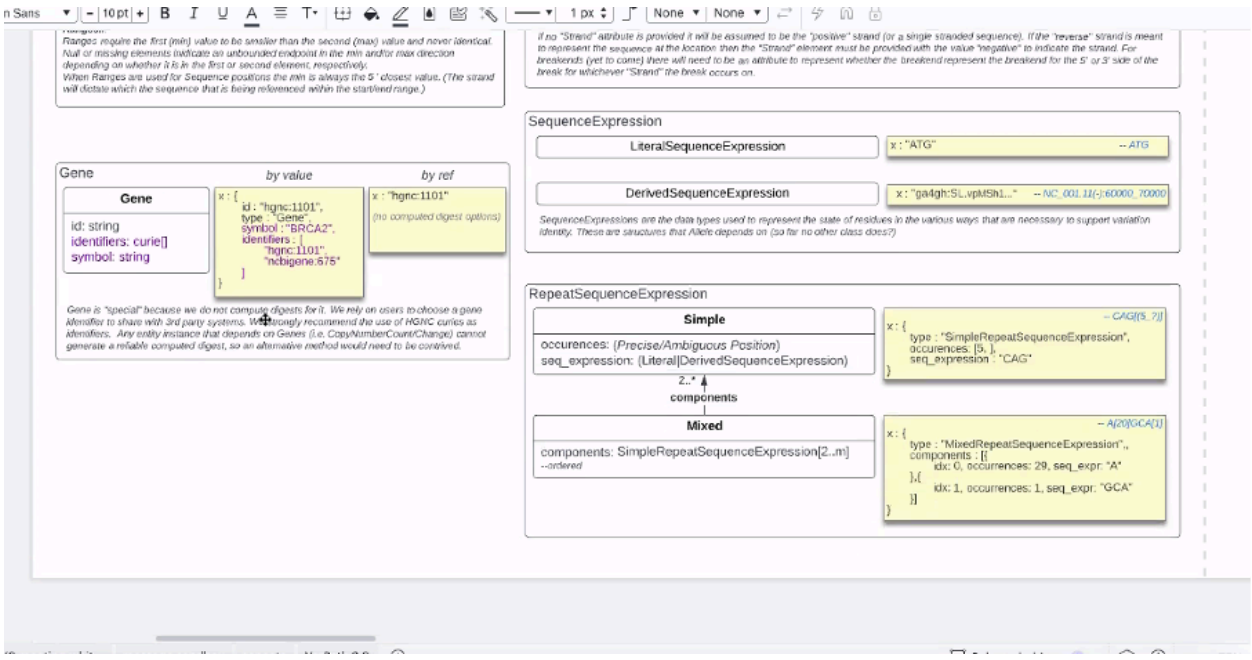
LB – Difference – We are not creating at the top level.

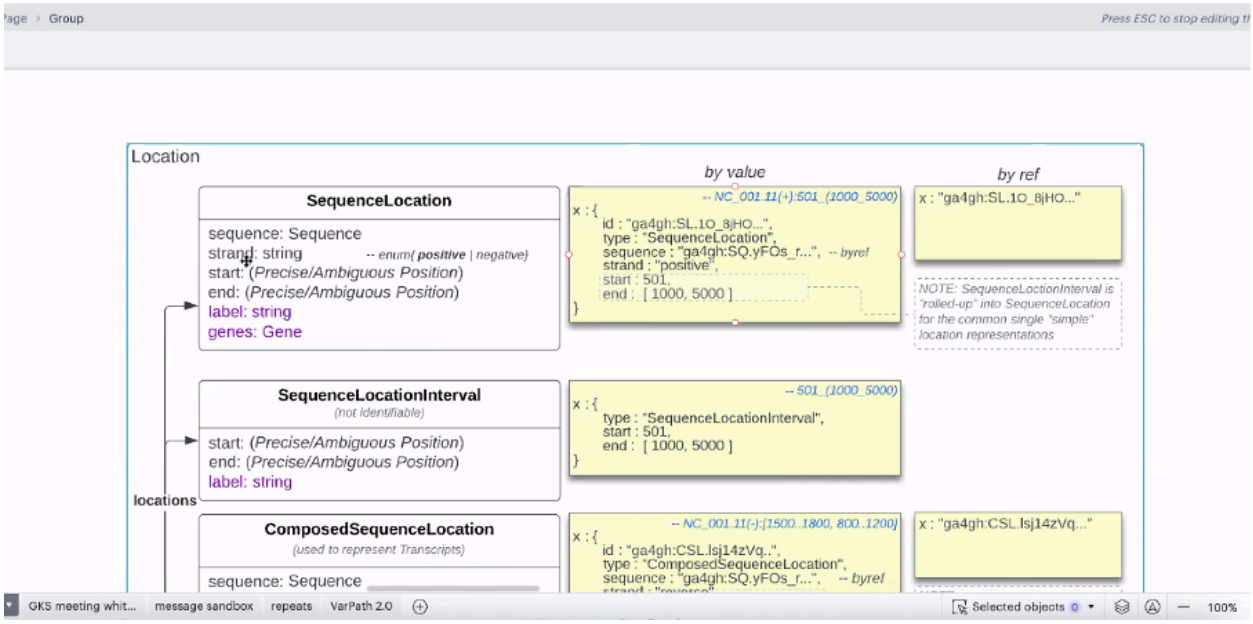
- We are defining the context. In allele we are represent
- ID we use the author ID.
- Microsatetelitle – 3 part deletion
- Identity are same
- Timestamp the identifiers .

AW – 1) we have lot of variation flat
We need to reduce duplication. Ex- haplotype – lot of alleles
Any way to roll that up and make the message simpler.
Retaining structure – in vrs 2.0 descriptive labels

LB - passing the message – we can change thing- can data completely.
LB – in normalization standpoint – should break seq at haplotype level.
AW – duplication of content – not necessary. Wanted to make it not flattened too much.
Ref seq derived from seq location. Not about a single level.

LB - trying to flatten as much as possible.
BF - 100% agree that conceptual model does not have to equal message structure (but the reverse is also true)





AS- What happen to clinvar allele ID

LB – Clini var receive variant – endpoints are same and ambiguous.

Notion must be different variants

LB – Clini var receive variant – endpoints are same and ambiguous.

Notions must be different variants.

LB – planning to get back.

AW – we don't have a schema yet. Going to implement before plenary.

WG – Expanding the schema and extend to define stuffs, dependency

LB – another class – seq position.

LB – Different than seq location or end points. Built a separate

LB – orientation . Called it as strand than orientation. Number is based on forward strand numbering and not reverse strand numbering

DC- breakpoint +and minus. Diff tools diff notion. Depending on the tool it changes

LB – talking about forward or reverse strand. Technically landing on one strand or other.

DC – going from position to 10 – 50. Someone will write in one way and another in another way.

LB – numbering, always on the forward strand.

LB – Planning to take all this account.

MT – Why vr2.0 location and start and end, why both?

LB – Don't need to have any necessarily all these here.

LB – Suggestion here – not every element here is an element used to identify alleles. What we use is type, location and state.

LB – we took start and end , rolled up and flattened it.

CP – decoration ?

LB – optional elements computationally not necessary.

CP – how do we dereference these alleles.

KF - I Find it a little confusing to have "location" be an object with properties and also have some Location properties be mixed into the Allele object adjacent to "location

VN - Shouldn't start, end, sequence and strand be attributes that can location can be decorated with?
LB – we are starting to implement these this week.
BF – interest – how it matches the information model that we are building? Are we change object structure?

2023-05-01: VRS 2.0 working/planning meeting

Hosting: Andreas Prlic, Larry Babb, Alex Wagner
Attendees "Name (Affiliation)": Tristan Nelson, Daniel Cameroon

	Actions Arising	Assigned To	Deadline

	Agenda Item	Speaker/Leader	Time
1	Working meeting to prioritize concepts for VRS version 2.0	Alex/Larry	50 minutes
2.	A.O.B	Group	Time remaining

Meeting recording :
https://us02web.zoom.us/rec/share/mMQM94TmdJ92zo1V3C0--jCOTBG8oO_8bD-sEVogHc9t7N1DSYpmVZqe6gWUlycH.BbY1cTy6TEI2nrv6
Passcode: h=0tP5M+

Minutes:
VRS 2.0 priorities:
Daniel Cameron (DC):

- Connecting genomic intervals (Sequence Locations)
  - Need a fundamental junction object (edges in a breakpoint graph)
  - Location and orientation
- Event model; how do things co-occur (relationship to Genotype model)
- "Anyone that uses VCF for SVs for anything vaguely complicated needs to deconstruct the VCF to a breakpoint graph"
- Terminology: what is an "event"?
  - Is an event in scope for VRS?
- Event phasing: SNPs on Sequence Locations that are part of a SV
- **Goal: fully describe a T2T representation of a cancer genome**

Andreas Prlic (AP):
What are the fundamental types of variants?

- Readthroughs, CNVs, Structural Variants, repeat expansion?
- Identifying policy / guidance around the use of VRS concepts

What about precision in VRS? For example, using confidence intervals to describe how likely a CNV is within a definite range?

Tristan Nelson (TN):
Do we represent CNVs with breakpoints?

- DC: no, these are abundance statements
- DC: also, having an inversion as a primitive is a mistake
- TN: that makes sense; we're basically talking about a vocabulary for creating composite sequences from other sequences

In the context of VRS 2.0, what breaking changes may occur to accommodate the SV structures, or is this purely an extension?

- DC: at a high level, I think this is non-breaking, but devil is in the details

Larry Babb (LB):
I would like to review of our VRS 1.3 / 2.0 change diagrams, here
First: remove the interval class
Second: no "Number" class wrapper
Third: typing sequences
Fourth: digest

Alex Wagner (AW):
Important to me: getting descriptive elements into VRS that do not impact computed digests

2023-04-24: CANCELED

2023-04-17: Connect preview

Hosting: Andreas Prlic, Larry Babb, Alex Wagner
Attendees "Name (Affiliation)": Anastasia,Katie Stahl,Kori Kuzma,Salem B,Amanda Spurdle,Hangjia Zhao,Tim Hefferon,Kyle F,Srikar C,Reggan Thomas

	Actions Arising	Assigned To	Deadline

	Agenda Item	Speaker/Leader	Time
1.	Maturity Model Preview and Discussion	Alex	25 mins
2.	VRS 2.0 Preview and Discussion	Larry	25 mins

Meeting recording :
https://us02web.zoom.us/rec/share/95UYvUcnoOjfo69l-WHej1Do4RGWhHFNdEGRS_RAYgOuPFKR9Os1BnOH8h6L7PPK.bofdUUDEdof5-6L2
Passcode: $r+zvT66

Minutes:
AW -
https://docs.google.com/document/d/1pqxeB9abRpLe0j6PiRBjmPYPKO9URyBY2QYVcyqO-m8/edit#

The idea behind the maturity model and our versioning policy is to apply a framework of common principles for how we discuss data classes developed in the GKS workstream broadly and, and then how we communicate those with the public

hen any given product is developed, there's kind of this adoption curve, right. And the idea is that early on, it's just the very interested participants like those on this call that that will innovate and pick up the technology and help develop it and then eventually like that grows and you get some early adopters. we're kind of in this phase now we're seeing a lot of larger initiatives picking up verse and using it in production.

we're trying to continue to build the models and develop them, people need to know what parts of verse are sticking around and are stable and what parts of verse are still changing or still kind of open to change so that they can get in and help

Walkthrough of Data class maturity Level
Branch management

Walkthrough of maturity and Versioning process flow diagram

SB - quick question can one class be beta and the other one be alpha but like can does every class have to be beta in order for it to be pushed into this like beta branch that you have like the 2.1 Or how does that work.

AW - you can have mixed classes and the idea here is, it's really that you want to make sure that every class within a branch at a certain level, meets the criteria for that so in the beta branch is for everything

Everything's going to be either a beta release candidate or stable level

LB - VRS 2.0 – Example drafting:

model that we came up with and that we have right now is this notion of a class of descriptors one for basically each of these different types, that sort of wraps its, its sibling type. So in other words we have sequence descriptors for sequences location descriptors for locations and so on and so forth. At the top of the line, you know, almost all of these things have calmed.

One of the things we'd like to do is essentially take the, the optional fields that are on these descriptors and allow verse to have these optional fields directly in the corresponding first object.

KF - concerned about the identifiers around the inverse value 100 identifier versus the descriptor identifier, making sure that they're clearly defined ?

LB - let's say you had the essentially same variant, and you're trying to represent the form that clean Gen allele registry creates versus the form that ClinVar creates. it might both have the same originating contextual build 38 genomic change, And therefore the truth is, we would consider them from a first standpoint to be the same variant. However, the way the allele registry decorates them and the number of expressions they compute and the way they do their alignments and all that kind of thing, really creates their own sort of descriptor around that value object then ClinVar does.

Key takeaways : nice to have some material examples in simplified form.
how are you capturing CNVs ?

2023-04-03: CNV,VRS 1.3 & SV-SIG

Hosting: Andreas Prlic, Larry Babb, Alex Wagner
Attendees "Name (Affiliation)": Wes Goar,Daniel C,Kori K,Alain R,Katie S,Salem B,Anastasia,Michael B,Kevin R,Catherine P,Bob F,James S,Vivek N,Mrinal T,Srikar C,Kyle F,Alain R

	Actions Arising	Assigned To	Deadline

	Agenda Item	Speaker/Leader	Time
1.	CNVs and VRS 1.3 Release Candidate	Alex	15 mins
2.	SV SIG introduction	Daniel	15 mins
3.	A.O.B	Group	Time remaining

Meeting recording:https://us02web.zoom.us/rec/share/lK43h4uAcpR6U88yog5XohXnDeCD3vHq4YLcwHqZU7UqeUGfWFFm9V3fOO7q95dW.zLTM3iRFyYvJdVhp
Passcode: zF6$!8vH

Minutes:

Notes:

- (Michael) regarding the CURIE discussion here is the previous GA4GH documentation https://schemablocks.org/standards-recommendations/#identifiers-and-curies
- (Michael) updated documentation of CNV definition in various systems (pls. add/improve!) https://cnvar.org/resources/CNV-annotation-standards/#cnv-term-use-comparison-in-computational-fileschema-formats

AW - High level walkthrough of Release 1.3
LB - https://registry.identifiers.org/registry/efo
MB - The codes you use are in principle CURIEs.
http://identifiers.org/EFO:0030070

DC – Walkthrough of SV
DC - So the first kind of fundamental question that we need to ask is, how wide are we actually expanding the scope? Are we taking kind of a narrow definition of structural variant in that? Are we wanting to just be able to represent DNA adjacencies? So where this section is this piece of DNA and this piece of DNA are actually physically connected on the same chromosome? So that's the narrow definition.

So you combine that with that, and copy number support. And you get kind of a VCF level of support there.
Do we actually want a wider scope than that, we want to be able to actually represent the more complex rearrangements, or even go towards an event based model that vaguely corresponds to the actual biology of the underlying rearrangements happening.

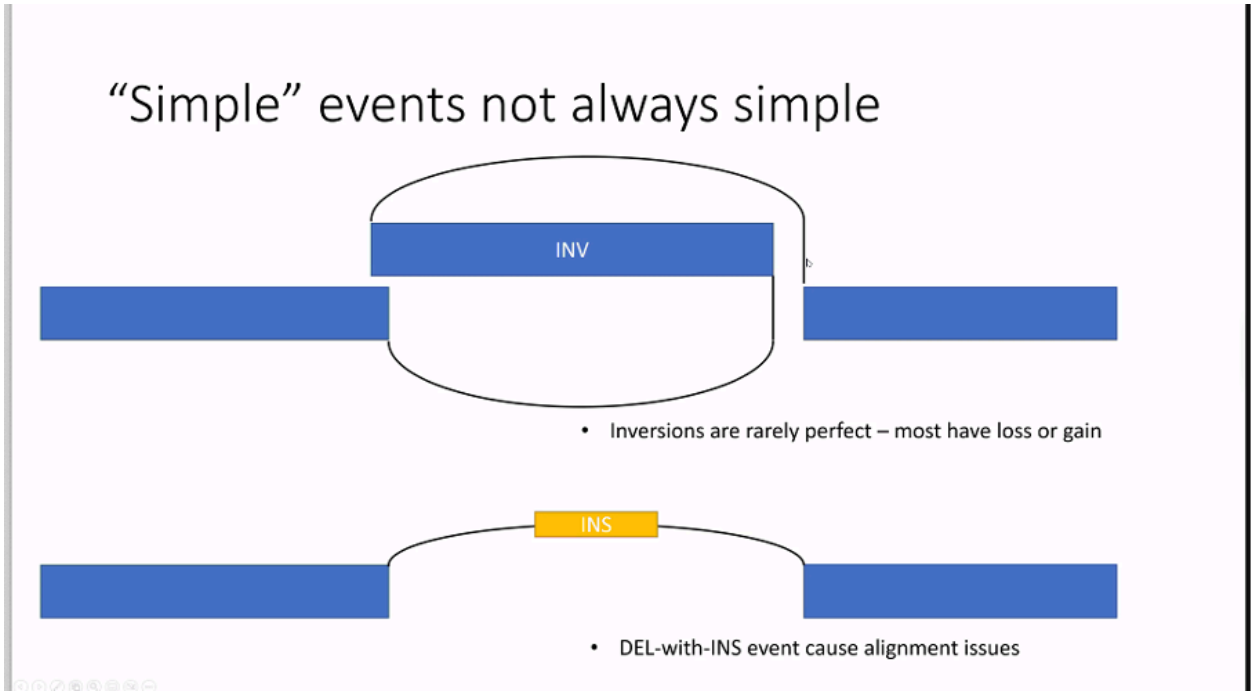
Because fundamentally, when we sequence or whatever technology we use to describe a genome, you'll have a set of structural variation, genomic rearrangements, you'll be describing the state of it.

So you can describe that state fully with copy number segmentations.

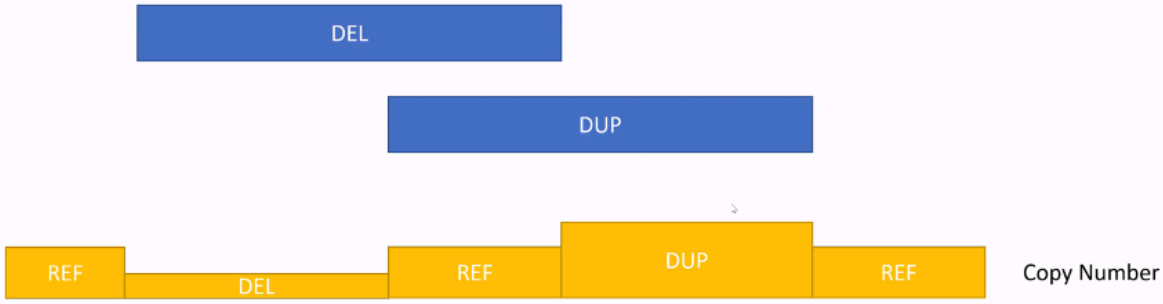
So DNA abundance, and where what's connected to what so your DNA adjacencies but that only gives you a descriptive outline of the state of that it doesn't actually tell you like the biology of what's actually happened, what events are actually there.

LB - Will we be reporting the events or attaching knowledge to the events? I presume the “events” are important and useful, I’m just wondering if VRS is about the final call or does it need to represent the events that make up the final variant representation?
MB - MO "representation according to available knowledge" is the goal; tools for each individual high resolution event in a genotype & compose, but fine for these "resulting outcome" representations when nothing else is available.
LB - Maybe a separate model for these event-based structures which can be composed and associated to a resultant representation that fits more into the existing VRS style variation.

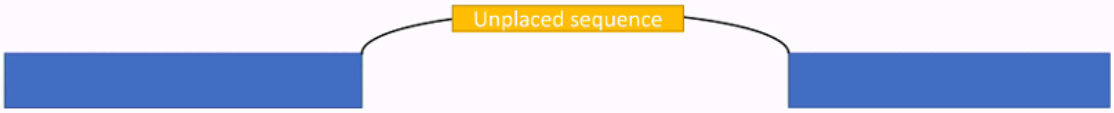
SV -VRS scoping discussion:



Overlapping events

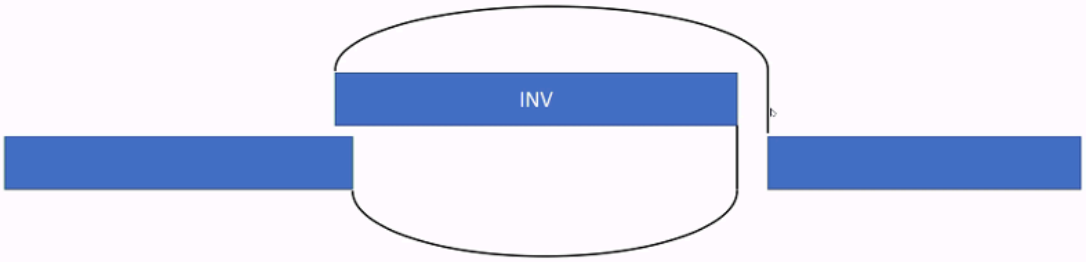


Gapped adjacencies



- Long read gaps up to ~1000bp
- Optical mapping gaps up to ~500kbp

“Simple” events not always simple

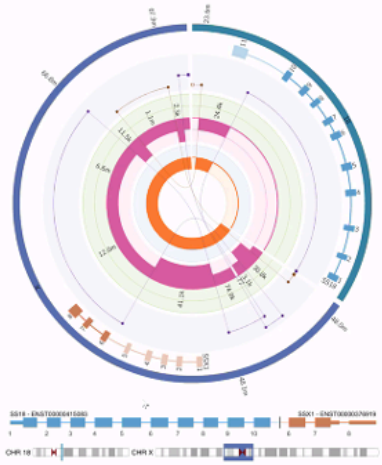


- Inversions are rarely perfect – most have loss or gain



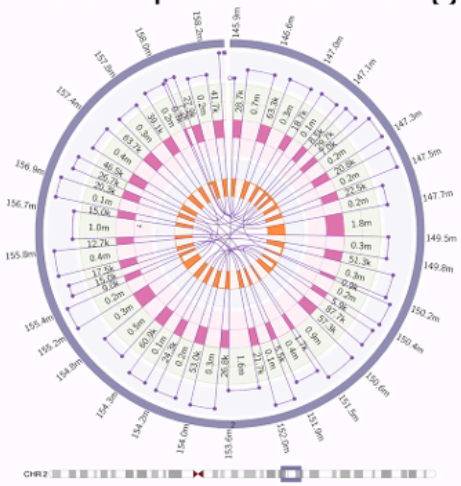
- DEL-with-INS event cause alignment issues

Complex rearrangements



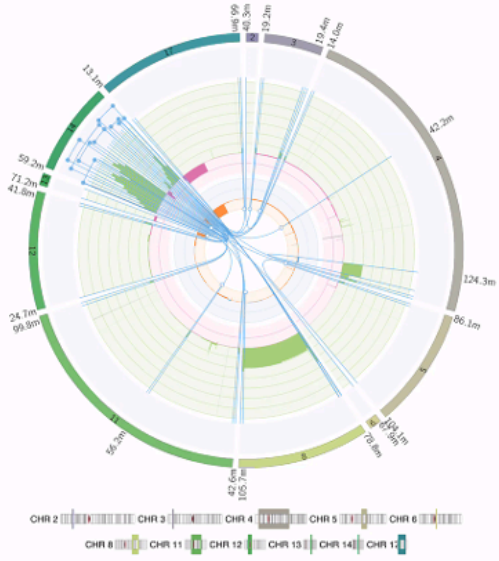
- Single complex rearrangement event
- Multiple derivate chromosomes

Complex rearrangements



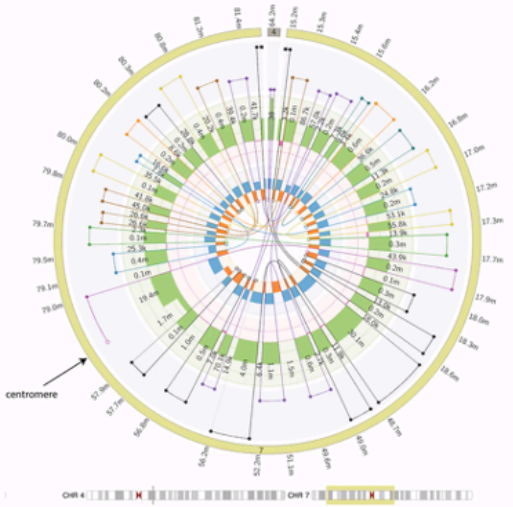
- Fully resolved chromothripsis
- Many overlapping DEL, DUP, and INV3/INV5 breakpoints

Complex rearrangements



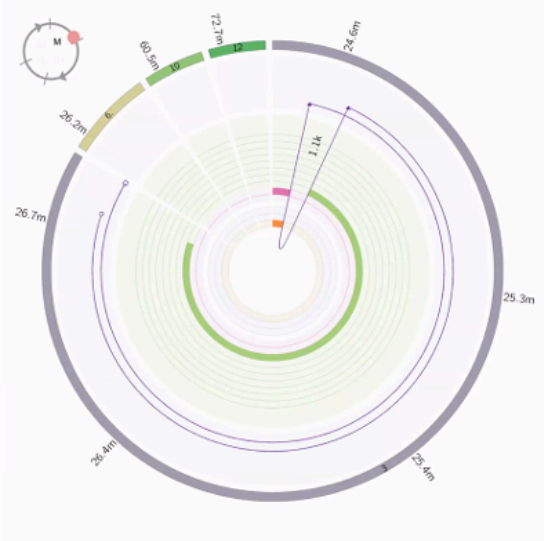
- LINE reactivation
- Many overlapping events

Ambiguous rearrangements



- Partially resolved chromothripsis
- Single breakends/ambiguous rearrangements

Temporal reconstructions



- Predictive reconstruction of ancestral rearrangement states

BF - Apologies if I missed this in the charter for the SV group. Could you please clarify if the intent of the group is to:

1. represent precisely the hypothesized rearrangement event(s) that led to a SV

2. represent the conceptual, feature-based outcome of an SV (e.g., region X was inverted)
3. represent the outcome state of a SV (e.g., the sequence)

#1 requires incredible precision and nuance, as Daniel is describing now. #2 seems to be primarily feature-based, may not be granular enough. #3 would be much more simple (sequence-based), potentially with annotated features that provide info about the hypothesized “source” of each region of sequence prior to the rearrangement event(s).

My personal bias is toward #3, but I recognize others in the group may have broader use cases.

LB - Maybe a form of “categorical” variation? (But I could be missing the point)
MB - @Larry E.g. chromothripsis is an annotation
BF – what is the linear reference genome space? Point of reference ?Are we working on feature based?

- Expressing the outcome of the operation? Linear seq- are we going to annotate them?

DC- we want to represent the genome of the event. Temporary is not something that we observe. Having the event allows us to do tertiary analysis . Lot of the details or most are relevant.
Work done on event based models.

AW - VRS does all three already!
MB - Exactly. As good as available & then compute to common reference resolution/use case (e.g. t(8;14)(q24;q32) may come from very different analyses but if you have a B-NHL classification use case ...).

BF – Are we trying to define the primary subj of the model that we building here? Or is this the subject of the output state?
DC – The reason the event based model is powerful – relatively simple.
MB - MO VRS variants could benefit from a technical provenance (besides using "cytoband" etc. a a proxy for such types...). Many variants won't be "best possible/true resolution" at time of description.
LB – Would love to see balanced translocation. Events lead to it. Want to know how we model these events. Need to understand – who uses these events? Scientist – final variant from ..

2023-03-27: CANCELED (Due to unavailability of the leads)

Hosting: Andreas Prlic, Larry Babb, Alex Wagner
Attendees "Name (Affiliation)":

	Actions Arising	Assigned To	Deadline

	Agenda Item	Speaker/Leader	Time

Meeting recording:
Minutes:

2023-03-06: Copy number,MNV & Connect

Hosting: Andreas Prlic, Larry Babb, Alex Wagner
Attendees "Name (Affiliation)": Catherine P,Mrinal Thomas,Michael B,Vivek N,Kori K,Hangjia Zhao,Anastasia,Kevin R,Kyle F,Salem B,Wes Goar,Likhitha S,James Stevenson,Katie S,Brad M,Tristan N,Sofia H,Reggan Thomas

	Actions Arising	Assigned To	Deadline
	Further discussion on ideas for CNV class names; please submit all proposals by end-of-week (issue 404).	All members	3/10
	MNV policy guidance	Andreas	

	Agenda Item	Speaker/Leader	Time

1.	Final proposal for CopyNumber and CopyNumberAssessment variation models.	Alex	10-45
2.	Multi nucleotide variants, do we want to make recommendations for how to represent them?	Andreas	10
3.	Open discussion/Q&A (eg topics - Connect, VRS related projects, etc...)	Larry and all members	Time remaining

Meeting recording:https://us02web.zoom.us/rec/share/glPji9VVstiQpgBg7GM99SXEdC1NDxWrhQxowOiF1clitTjBpL4M10jyvmzxvNCZ.lyjKCDKwTZzpR0n_
Passcode: !kQ!7u0q

Minutes:

AW - https://github.com/ga4gh/vrs/issues/404
AW - https://github.com/ga4gh/vrs/pull/416
AW - we're going to leave copy number unchanged, we're going to create the new copy number assessment. And we're going to use the efo codes that Michael provided
AW - copy number assessment class- a very similarly uses a subject, same fields. But then you've got this copy assessment and this copy assessment
AW - Some missing code that we don't have yet for high level loss, low level loss loss, Regional Base ploidy gain, low level gain or high level gain
LB - Assessment" may be confusing especially down the road when variant scientists and specialists need to assess a CopyNumberAssessment which would be a kind of CopyNumberAssessment Assessment.
MB - Still a bit confused by "high level loss" though I don't have a better way to name this
LB - We should really find a term that represents the "kind" of copy number it is, which means we may want to also qualify the current "CopyNumber" so that it is clearly distinguishable when folks want to treat all "kinds" of copy number from CopyNumber and CopyNumberAssessment
AP - How would we express tandem vs interspersed duplication variants?
LB - The only way to express "tandem" is using an Allele with the RepeatedSequenceExpression.
LB - If you want to express both the randomness and the quantity of copies you'd need both VRS object types. Or put it in a descriptor around one or the other
LB - we often call assessments what the variant scientists curators and physicians do when they're assessing a variant in terms of, in the context of clinical case or research or phenotype
LB - we should be careful not to not to muddy the waters too much here because you can see where a lot of a lot of times at the higher level we call assessments of alleles allele assessments or allele assertions
LB- we're just trying to do is come up with the notion of is this a gain or loss or a neutral thing? and maybe leave the interpretive assessment stuff to a higher level kind of VA kind of world
LB- it seems nice. to not have to change copy number to absolute copy number or quantified copy number, because we've already done it,

LB- I defined a class I tried to have the last part of the class B, the, the noun, the thing that it's being created and if the last part of it, it's an assessment, it makes me feel like what we're creating it as assessment, we're not creating a copy number, and these are two kinds of copy numbers.
TN - This feels like something where we could collect suggestions and arguments on the GitHub ticket.
LB - CopyNumber and CopyChange is pretty good by me
VN - Would we ever want to represent the copy number count and the copy number assessment together? In which case would one of the pieces of information be part of the variation descriptor?
LB- @Vivek IMO if one would want to “classify” the relative change of a copy number then that would be a higher level “assessment” or “descriptor” type of attribute. I think you are shining a light on the point that one of these things is derivable from the other but not visa versa. CopyNumberChange is somewhat ambiguous
Also, CopyNumber can be a bit ambiguous if one uses an indefinite range for he unbounded range of copies gained or lost (although loss is not really unbounded since it must stop at 0).
VN - would we ever want to represent the call number count and the copy number assessment together as one object, and if he wanted to do that with one of these pieces of information beyond the radiation descriptor level and set up the object itself?
AW - This also gets to the verse 2.0 thing I think if you're doing what we've been calling a copy number assessment, you're typically you'll you might parameterize that in some way.

AW- We're going to make a few comments from today's discussion on this PR everyone can go to 404 add their thoughts, and we'll pull them in for our next verse meeting
AP – MNVA – walkthrough
AW - f you're assessing those individual SNVs, you might get a very different result than what's actually biologically meaningful for that patient. Were there it's really an mNV that produces a different code on the one either s&p does or what either or what the reference does.
LB- Andre, Do you think you could open up a ticket on mNV naming and will will flag it as a, as a recommendation ticket or something like that. I think that's a good next step
LB – Walkthrough of the connect schedule

2023-02-27: CNV & VRS 2.0

Hosting: Andreas Prlic, Larry Babb, Alex Wagner

Attendees "Name (Affiliation)": Wes Goar,Kori Kuzma,Salem B,Likhitha S,Anastasia,Tristan Nelson,Kevin R,Katie S,Brad M,Kyle F,Hangjia Z,Bob F,Andy Yates,Mrinal Thomas, Vivek N,Catherine P,Tim Hefferson,Reggan Thomas

	Actions Arising	Assigned To	Deadline

	Agenda Item	Speaker/Leader	Time
1.	Review proposal for CNV (issue 404) with the intent to finalize by Connect	Alex	20
2.	Connect plans (VRS 2.0 flexible adoption, Maturity Model)	Larry / Alex	30

Meeting recording:
https://us02web.zoom.us/rec/share/4x_wWsPpjOUgfuoUSfhR_M5GWO_8wNIN48-ena7wGRVWvUO3clQ27WfwdWC0nFpa.CZ1lKBo2XPpLP8nf
Passcode: z$h9i!+d

Minutes:
CNV – Walkthrough of **https://github.com/ga4gh/vrs/issues/404**

AW – need to name the classes appropriately.
Will go from relative copy number to ?
MT - an algorithm or something of that sort is like a deterministic way to always go from an absolute copy number to a relative copy number?
AW - moving away from the absolute relative copy number terms in this proposal, but to go from a copy number variation to copy number assessment.
LB - you can go from a copy number to a copy number assessment so that everyone ends up getting the same copy number assessment from the same copy number.

AW - we would need a policy on what determines low level versus high level for us to be able to make that level of granularity in terms of our mapping.
AW - Very useful to have some community guidance on what constitutes low level and high level game for example
AW - what's being written into this proposal. we're not creating those policies.

LB - Sort of in favor of just having three enumerated terms - gain neutral loss

AW - CMP callers that use these types of statements. Like the GTK caller, there are basically five categories going from negative two to positive two and they basically correlate to or

correspond to,complete impartial loss, Copy, neutral, low level and high level game. and they have their own policy internal to that tool in terms of what makes it a lower high level game, but they do make that distinction in terms of the output from that caller?

. And this is a very commonly used tool. And, you know, like CMP calling in both the germline and somatic settings.
AW - Idea here is we want to be able to reflect the type of data that we see in CNV callers, and this is very commonly observed this kind of like five level representation

BF - I would prefer to build the structure and point to a terminology for the content. The policy discussion (debate) can then occur independently in the terminology space.

LB - one size fits all models is not going to work. And that doesn't mean that conceptually we have different models. It basically means from a pragmatic or use case

LB - we want to have different ways to model and structure the data so that everyone can do their best to adopt and actually use verse and be able to map back to the same conceptual model.

Walkthrough of an example of an allele:

LB - grabbed a bunch of records from ClinVar SCV records and we turned them into VA statements for pathogenesis statements/What you're seeing here is just the variant portion of one of those statement

LB - variation descriptors here are where we talk about building these wrapper classes. And from day one, when we talked about VRS and we talked about using VRS, we always said to ourselves, well, as great as this record, this, this pure computational record is, it doesn't really meet the needs of a lot of systems when they're exchanging data about variation. Just like in a patient, an EMR system, you don't send the bare minimum, two or three points of data about a patient to identify the patient. you need those maybe required and those may be the things that let you tie your master patient index or whatever. and they may be really important so you can know you don't have a duplicate patient or, you know, you've got the wrong patient

LB: We end up creating these classes for the variation descriptors in ClinVar. This is what we're done going back to the GK pod for ClinVar.

AW:
takeaways here are :
We want to extend the verse 2.0 model to allow for descriptive attributes inside the VRS objects. So you can imagine having a verse allele that has a reference sequence attribute, for example, that people often want to send associated with it, or the location embedded in that allele having the string associated with the sequence digest. So it could be something like an M accession or

NC accession as information that gets passed along with that entity. The idea here is to allow verse objects to remain self contained and be used improperly inside schemas but also give people the ability to decorate them with information that's often requested and valuable for you know, describing a variance to people without having to rely on a parallel structure of descriptors which do not nest very easily, Or well, which is one of the challenges that Larry was pointing out. AP - are we talking about kind of a workflow around data here and so when workflow is, step one is very registration in step two is annotation.?

AW - Looking for the advantages and shortfalls of VRS as you've seen them in your own applications.

2023-02-20: Canceled due to Holiday (President's day)

Hosting: Andreas Prlic, Larry Babb, Alex Wagner
Attendees "Name (Affiliation)":

	Actions Arising	Assigned To	Deadline

	Agenda Item	Speaker/Leader	Time

Minutes:

2023-02-06: Pipeline/Tool Rundown,Roadmap &CNV

Hosting: Andreas Prlic, Larry Babb, Alex Wagner
Attendees "Name (Affiliation)": Kori Kuzma,Jeremy Arbesfeld,Anastasia Smith,Tim Hefferon,Katie Stahl,Kyle Ferriter,Wes Goar,Kevin Riehle,James Stevenson,Mamana,Tristan Nelson, Salem Bajjali, Michael Baudis, Bob F

	Actions Arising	Assigned To	Deadline

	Review and add comments to the Maturity Model / Versioning policies	GKS Contributors	2/20/23
	Review issue 404 and today's discussion (recording link) and add thoughts / propose solutions on the GH issue	VRS Contributors	2/20/23

	Agenda Item	Speaker/Leader	Time
1.	Pipeline/Tool Rundown : ClinVar, gnomAD-vrs, MAVEdb-VRS, VCF annotator, AnyVar plans, Variation Normalizer	Larry	20
2.	2023 Roadmap : Maturity Model, SV SIG	Alex	20
3.	CNV class naming (issue 404),	Bob F and all	20

Meeting recording :
https://us02web.zoom.us/rec/share/wRHWfoPJKk5pDHfE5tlqzDETap1n_vA-_IjOwoV96deBGeXK8GroJSrEj0YcUzag.MyjymFwJjYSXU5ei

Passcode: %w77Ax@3

Meeting Minutes:
AW - https://docs.google.com/document/d/1pqxeB9abRpLe0j6PiRBjmPYPKO9URyBY2QYVcyqO-m8/edit#

Walkthrough of data class maturity model and versioning
Walkthrough of Data class maturity level - Alpha, Beta, RC and Stable

LB – Gnomad – Alex is working on this. Will be building an API. Will be done in next few months.Gnomad to produce a Json file
MAVEDB – Changes to amino acid residues, challenges- these are synthetic.
VN – is a open source service developed by Kori under the supervision of Alex Wagner. Improve the usability of VN.
MB - https://cnvar.org/resources/CNV-annotation-standards/

MB - Just had again some feedback about the "if we want to launch CNV Beacons we need VCF as import format" and wanting myself to do an example/workflow...
WG - Yeah we need to work on our documentation.
We do have test data in our repo that can be looked at. I tested this on the genome in the bottle gold standard data.

LB – Anyvar *https://github.com/biocommons/anyvar"* – provide python and REST interface to validate, normalize, generate identifier, and register biological sequence.

Copy number - Absolute and relative copy number
MB - Does "Beacon" mean "Beacon v2 variant default model"? Which ingests VRS (though allows "legacy" which no one ever used).
LB - https://github.com/ga4gh/vrs/issues/404
BF - "tandem" conveys structure, which is yet another concept that differs from CN
MB - @Bob That's why SVCLAIM in VCF4.4
BF - dup is not a copy number gain if the other chrom had a loss
MB - @Bob We have been there
BF - My point is that these are diff concepts, so we should consider capturing them separately. I posted examples of these here:
https://github.com/ga4gh/vrs/issues/404#issuecomment-1347398631
MB - @Tim Even that; we discuss a lot about precise representation where no precise data existed
MB - I think the notes for the EFO classes capture the meaning
https://github.com/ga4gh/vrs/issues/404#issuecomment-1419363538

LB - My argument is that is not our issue or charge. Our charge is to represent it as we get it (consistently). Then down streamers can figure out what policy they should apply

---

2023-01-23: Recap and 23Q1 Roadmap Plan(Canceled - MFW)

Hosting: Andreas Prlic, Larry Babb, Alex Wagner
Attendees "Name (Affiliation)":

	Actions Arising	Assigned To	Deadline

	Agenda Item	Speaker/Leader	Time
1.	Recap : GK Pilot developments, CNV class naming (issue 404), gnomAD-vrs, VCF annotator, AnyVar plans, Biocommons governance	Alex/Larry	

2.	23Q1 Roadmap : Maturity Model, SV SIG, SA-VA-VRS Crossover	Alex/Larry	

Minutes:
The meeting was canceled due to MFW (Meeting Free Week) and the agenda items were copied to the 02/06 VR meeting.

2023-01-16:Sync for Genes Phase 5

Hosting: Andreas Prlic, Larry Babb, Alex Wagner
Attendees "Name (Affiliation)": Bob Dolin,Mrinal Thomas,Justin Aronson,Wes Goar,Oscar Pastor,Salem B,Kyle Ferriter,Brad Munster,James Stevenson, Vivek N,Reggan Thomas (GA4Gh)

	Actions Arising	Assigned To	Deadline

	Agenda Item	Speaker/Leader	Time
	Welcome and review agenda (super brief rundown of current VRS activities - MM, PSB, ClinVar, Civic, Epic, gnomad)	Alex/Larry	10
	Sync for Genes Phase 5: A standards-based approach to sharing dynamically annotated genomic variants	Bob Dolin	20
	discussion topics related to presentation 1. VRS/VA implementation takeaways 2. Proposal to upgrade FHIR genomic molecular consequence to align with VRS 3. Complexity in handling SNV vs MNVs	all	as needed

	VRS discussion topics if time remains 1. MM updates 2. Epic KB pilot project raised concerns a. Tandem PGx star alleles '+' and 'x' nomenclature b. gene fusion design expectations c. ClinVar, Civic KB data delivery early adopter involvement 3. AnyVar restart - looking for interested users/developers	all	if time remains

Meeting recording
https://us02web.zoom.us/rec/play/8GwbXaiW3-x6zPzvfEbyUpBj49uLnFAr9EiOUJ6U9TuUzdV1HHpgeWvMghFGGGvD51DuA_23O46ITyk7.tm0K0LH9-U-6hu7T?continueMode=true&_x_zm_rtaid=rSDenPpXT_mi8B7788f2bA.1673938449631.9f6eccc62b80e6903e9995edc480d26d&_x_zm_rhtaid=968

Password : $SfU3#up

Minutes:

ONC (Office of the National Coordinator for Health Care) - a national entity in the United States that's primarily responsible for creating recommendations for interoperability standards for use within electronic medical records. They define what terminologies to use for medications or conditions. For labs, they define what protocols to use to communicate data from a laboratory over to electronic health records. And they've also been trying to advance the use of genetic data within electronic medical records to try to help figure out ways to drive precision medicine

Project overview -
ONC Syn for genes goal- Standardize the sharing of genomic information , mainly FHIR
ONC sync for genes 5 goal - Standardized sharing of annotated genomic information , mainly aligned with FHIR
Background :
Alignment with FHIR profile.

Justin looked into Molecular consequence

Developed FHIR genomics API
Lot of ppl started exploring the repository. We can store, VCF,FHIR. Previously experimenting with CIVIC, CLinvar.. Can store MC . Did a bit of enhancement

Use-case :
1) Clinician want to view up to date genetic implementation including germline cancer screening and ermine pharmacogenomic screening
2) Clinician want to filter and prioritize potentially causation variants as part of a rare disease diagnostic work

Approach:

Replaced the knowledge structure.Imported GA4gh encoded knowledge into the repository

Genetic data we are using is primarily VCF data.

We updated the operation so that we can return back molecular consequence
Return back Therapeutic implications

Face sheet app:
Shows the expert panel review
Shows broader spectrum of annotation

High level:
Successful demonstrated an open source pipeline for sharing of dynamically annotated genomics information using ga4gh encoded knowledge, delivered using FHIR genomics operation

Minor issue - Struggle when submitting multiple phenotypes

Molecular sequence:

Variant Filtration:

1) Implication/ annotation/ consequence
A) MC
B) Population freq
2) Phenotype
A) Human phenotype ontology

3) Screening

Data processing overview:

1) Annotation pipeline —> FHIR transformation —> Additional annotation

Pipelines Miss data :

MNV conversion:

AW - 60% of MNV / adjacent SNV events?
JA - Yes

AW - is it 60 % of MNV or 60 % of Patients?
JA - 60 % of MNV's

LB :
Dynamic Variant annotation:

Automatic variant annotation varies in complexity based on variant type and may require further application -layer calculation

MNV conversion :

— Phase data can be used to enhance molecular consequence prediction

Lack of standardization:

1) ClinVAR has no association between MNV and SNV
2) Pipelines may report only SNV , only MNV or both

Application level conversion :

1) require phasing data
2) relatively slow

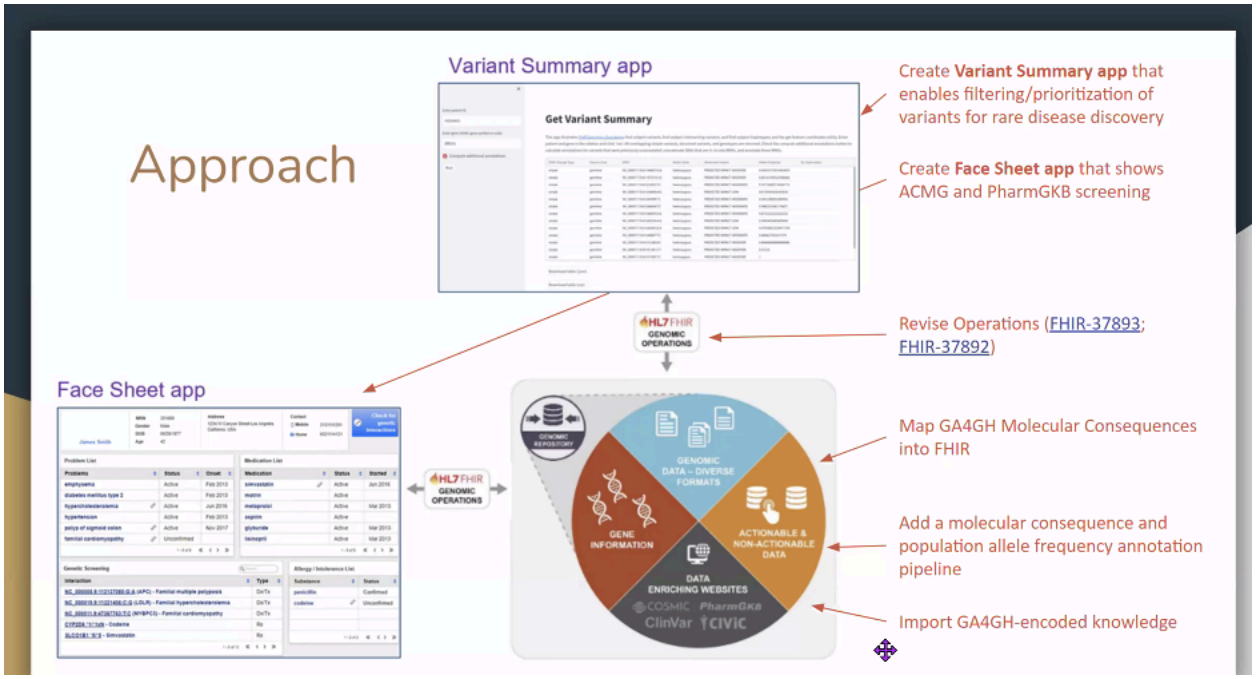
Recommendation :
1) Link component SNV to MNV
2) Report MNV's when possible

Demo on Molecular consequence:

Enter patient ID and Enter gene -

Click compute addition annotation - gets the additional SNV

BD - 60% statistic comes from https://www.nature.com/articles/s41467-019-12438-5



2023-##-##: TEMPLATE

Hosting: Andreas Prlic, Larry Babb, Alex Wagner
Attendees “Name (Affiliation)”:

	Actions Arising	Assigned To	Deadline

	Agenda Item	Speaker/Leader	Time

Minutes: