Data Visiting

Work Stream / Subgroup:

Regulatory & Ethics Work Stream

Date (Day/Month)**:** April 1, 2025
Time (EDT)**:** 4:00 PM-5:30 PM EST
Session chair(s): Donrich Thaldar, Yann Joly
Session chair(s) will be present: Donrich Thaldar, Yann Joly
Link to slides: presentation

Aim of meeting:

- To discuss different models of data visiting.
- To identify challenges associated with various forms of data visiting.
- To comment on plans for interviews aimed at understanding the specific needs and expectations of the genomics research community regarding data visiting.

Session description:

Data visiting is a form of data sharing in which analyses are conducted on shared data within the computing environment of the data provider, either through human or computational agents. Unlike traditional data-sharing approaches that require transferring large datasets across entities—and often across international borders—data visiting allows researchers, including analysts and machine learning algorithms, to perform analyses directly within the provider's secure online environment. While this model offers security advantages, its implementation presents legal, technical, and governance challenges that vary across jurisdictions.

At present, our insight into different models of data visiting is limited. This meeting will bring the GA4GH community together for an initial discussion to map out existing approaches, identify key challenges, and explore possible frameworks for data visiting agreements. Additionally, we will introduce plans for interviews with stakeholders to better understand the needs and expectations of the genomics research community. The findings from these interviews will serve as a foundation for legal research and the

development of practical, legal, and ethical guidance that is globally relevant and responsive to the realities of diverse stakeholders.

Agenda:

	Agenda item	Presented by	Time
1.0	Welcome	Yann Joly	5 mins
1.1	Background to the Group, progress with the publication	Donrich Thaldar	5 mins
2.0	Teaser: the data visiting models of GA4GH members	Donrich Thaldar	15 mins
2.1	Deep dive: Apheris	Robin Roehm	5 mins
2.2	**Brainstorm- challenges w/ DV**	**Yann Joly**	**25 mins**
3.0	Planned study: gaining a better understanding of the needs and expectations of the genomics research community	Paul Esselaar	5 mins
3.1	**Brainstorm- interview strategy**	**Paul Esselaar**	**10 mins**
4.0	Next Steps	Yann Joly & Donrich Thaldar	10 mins
5.0	**Questions & Meeting Close**	**Yann Joly**	**5 mins**

Resources and links:

- Data visiting lexicon
- Data Visiting Models
- Semi-structured interview questions

Attendees:

1. Monica Salani
16. Ariel Xue
33. Marc Fiume

2. Oliver Hofmann
3. Diya Uberoi
4. Dianne NIcol
5. Heather
6. Jan Stucke
7. Nathan Boyd
8. Christine Suver
9. Edmund Su
10. Gemma Brown
11. Palmira Granados
12. Benjamin Berk
13. Mette Peters
14. Andrew patterson
15. Harit patel
17. Jessica Seegobin
18. Donrich Thaldar
19. Chloe Migwon
20. Renee Potter
21. Melissa Basford
22. Robin Roehm
23. David Bujold
24. Ray Krasinski
25. Jaime Delagado
26. Cricket Sloav
27. Nicky Mulder
28. David Glazer
29. Anderson Brito
30. Vasiliki Rahimzadeh
31. Karen Cranston
32. Paul Esselaar
34. Brian O'Connor
35. Jaclyn Estrin
36. Kayla Socarras
37. Ma'n Zawati
38. Eberechi Wogu
39. Nathan Sheffield
40. Lee Lightenstein
41. Ashwin Budden
42. Cameron MacPherson
43. Yann Joly
44. Monica Salani
45. Oliver Hoffman
46. JP Pierce
47. Alison Hall

Meeting minutes:

Welcome:

YJ welcomed everyone and introduced the background of the data visiting group, including the aim and the status of the group. Since its start, the group has put forward a lexicon and a paper that discusses the various concepts associated with data visiting. The lexicon is now available on the GA4GH website and the paper is currently under review by Human Genomics.

Institutional Practices around Data Visiting

Given that Data visiting is practiced differently in jurisdictions around the world, **YJ** noted the importance of understanding the differences and similarities that exist around its implementation. **DT,** thus provided a thematic overview of data visiting, taking examples of 5 driver projects: Australian Genomics, UK Biobank, All of Us, Apheris, and ELIXIR Cloud & AAI. Each model varied as to three main elements: 1. Where the data is located, 2. How researchers are able to access data and 3. What can be taken home—and how that is monitored.

He emphasized that even if models differ as to what is viewable and or how data is stored, the key feature of data visiting is that data stays within the provider's secure environment and that researchers or algorithms can come to visit the data. **DT** highlighted that data could be stored by various approaches, including using central cloud platforms, institutional infrastructure, as well as federated clouds. Then he flagged different mechanisms of accessing data: de-identified data in secure platforms (UK Biobank & Australian Genomics); three-tier access model (All of Us); and researchers submitting code to obtain results (Apheris). Further details around these models can be found in the presentation.

Deep Dive: Apheris

RR, the CEO of Apheris, then provided an in-depth review around Apheris' experiences with data visiting and their approach around oversight. Computational governance is a method to control, supervise, and track all aspects of computations. He explained that one key challenge is around the adaptation of the model in different jurisdictions. While Computational Governance (CG) provides a potential solution around what can be computed and then what is returned, the implementation of CG faces challenges and it is not always clear how to comply with various legal requirements.

Discussion 1

YJ thus highlighted the importance of understanding how Data visiting is understood in different legal orders and opened up the floor for discussion. Given that the different platforms a provider uses to store data could influence responsibilities that a provider has, **VR** asked whether the practice of Data visiting is platform agnostic. **DT** clarified that DV is platform agnostic, as the provider could use any agent, but that agent or platform would also be considered a part of the environment. He did agree however, that this concept required further thought, given that it can implicate responsibilities providers have.

Considering the group's ambition to develop standards, **DB** enquired into what elements of data visiting, including, computational governance and asset policies could be standardizable. There are shades of grey in how this occurs in practice and it would be important to systematically understand what policies require and what is downloadable and what may not be. **CM and RR agreed,** bringing into the equation thoughts on standardizing the understanding of infrastructure.

While **DT** was in agreement, he explained that there is currently a debate as to whether one should use the term secure research environment or trusted research environment. This was something the group has tried not to delve into for the moment, but may be willing to get into, depending on directions people suggest.

Regardless of the concepts that will be explored through a potential data visiting tool, it will be important to understand the needs and challenges of the community and their experiences with data visiting. **PE** thus opened up the remainder of the discussion around the types of questions the group should be answering. He proposed holding interviews with the community.

Discussion 2: Interviews v. Deliberative workshop

While interviews may be one way to understand the perspectives of people and institutions on the ground, **YJ** suggested that we can get more diverse perspectives and ensure greater inclusion of bioinformaticians if a deliberative workshop is conducted. He proposed this as a possible next step for the group.

Highlighting the importance of representation and taking into account participants' perspectives, **AH** suggested including the voices of researchers on how data is used and how they understand the term. **YJ** agreed with **AH**, and also suggested that it might be more effective to have a solid lexicon before engaging with more participants. **DN** followed AH's suggestion and highlighted the importance of having participants involved in the study.

Key takeaways

- Data visiting offers many advantages to traditional data sharing, yet, challenges remain as to how the concept is understood legally.
- While it is implemented differently in regions around the world, similarities around key concepts exist, making it important to identify common elements that can be included in a future data visiting agreement.

Next steps

- The group will continue with background work on understanding Data visiting and its legal implications.
- The group will explore establishing a Data Visiting Agreement
- The group will touch base as to the next meeting and its plans towards a workshop and or interviews.

Should you have any questions, or if we can help with anything, please do not hesitate to reach out to rews-coordinator@ga4gh.org.