Global Health Data Science Hub working group meeting notes summary

Date: 2 May 2023

1. Review and agree Hub key priority areas for the next 6 months:

- **Data governance, sharing and reuse** (to support information retrieval)
  - Identifying data governance principles per institution/country
    - Identifying and developing an understanding of the existing principles
  - Harmonising across various areas
    - Secondary plan
  - Analytical pipelines and workflows
    - Federated analysis/learning model to support data sharing
- **Develop supportive and collaborative infrastructure for the team.**
  - E.g., GitHub account for the hub for developing different tools to support code sharing.
    - Group GitHub account
    - Guideline and training for team members
    - Finding projects that have git accounts or resources that can be reproducible e.g., in publications.
    - Training team members on collaborative infrastructure

- Data governance and sharing is an important focus area. This is an area which is required for all data science work and can be quite complex
- Identifying and surfacing all the governance tools and processes that exists across regions could help engage and empower our health data lifecycle
- Harmonising data sharing standards / governance processes is important as it would allow data to be more easily accessed and re-used
- Identifying and sharing the current state of data sharing processes and capabilities is an important first step on this journey
- Include Data Life Management (DLM) as part of the data research lifecycle

2. Hub metrics

- Metrics will be used to monitor site traffic and pages views and to increase usage and engagement with hub.

3. Proposal for bi-monthly 'Spotlight on' pages

- A new idea for 'Spotlight on' series was discussed. This would showcase curated content that is focussed on a priority topic and would be published on the hub every two months.
- This would provide an opportunity to invite experts to deliver webinars on the specific topic
- Data science clinics and training around R could be a good first topic for the series
- Due to the multiple different streams of activities on the hub we will create and circulate a working document that lists all the hub activities.