

# Breakout\_A-Remote (can map to issue name later)

Issue: [https://github.com/CFDSI/Kickoff\\_Workshop/issues/#](https://github.com/CFDSI/Kickoff_Workshop/issues/#)

Related Issues:

Issue Statement: 18

Discussion topic: [What critical reference parameters need to be stored in meta data files to ensure the usefulness and longevity of archived data sets?](#)

Moderator: John Farnsworth

Note taker: Riccardo Balin

Reporter:

Group Members:

- John Farnsworth (CU Boulder)
- Laura Villafane (Stanford)
- 

Please address these topics in your discussion (moderators please make sure that there is enough time to cover all three before the session ends).

1. Describe the problem:

- What format should the data be stored in?
- Where should the data be stored and how can we ensure it is available over time?
- How should the data be managed, and by who?
- Should there be a minimum standard/participation to be able to access data?
- Should there be information on how to reproduce the data?
- Databases can become very large, what should be stored? Raw data or processed data? What is useful?
- There is uncertainty in defining the problem of interest and determining what data to seek and share from the beginning of a project. This definition determines how the experiment or simulation is setup and what data is stored.
- Sharing data and contributing to the community takes valuable time from your researchers. No incentive in investing that time because no immediate or personal reward comes from it.

## 2. What are potential solutions?

- Guidelines must be set on how to share the data
- Make a platform where all the data is shared over time and available to everyone
- Have a library that points to where the data is stored
- The platform or library needs to be maintained and updated
- Along with data there should be description on how the data was produced, and how it can be reproduced by someone else
- Reward and incentivise investing more time in improving the community, sharing data and experiences. Instilling this mindset in students.

## 3. What can CFDSI do to help?

- Provide a platform to store data or the library to direct people to the data
- Manage the platform or library
- Provide guidelines on how to share data to reduce time invested
- Make workshops on how to share data, use data, use the platform or library

## 4. Misc ideas so they don't get lost (e.g., Did you find new issues? If yes, create the issues on GitHub!):

- Boundary conditions (inflow, outflow, tunnel walls, inviscid core, geometry) in experimental studies need to be characterized better for reproducibility of same flow problem in CFD, and vice versa.
  - CAD models need to be shared between groups, and notes on how CAD differs from model that was built and put in the tunnel
- Currently most collaborations between experiment and CFD involve one-on-one conversations between the groups involved, usually with a lot of learning for both sides. When new collaborations start, these conversations and learning steps are repeated, wasting time and slowing down progress.
  - There needs to be sharing of experiences and best practices for collaborations

## 5. Summary for report-back (Alternatively, just bold the key points above):

- New Issue: How do we catalogue/report to the community the 1-on-1 experiences that arrive out of close collaboration between researchers (EXP - CFD)? Many of these arrive out of personal discussions and discoveries and do not get readily reported in the present literature.
- What platforms and tools exist to host and catalogue data resources jointly for the community?
- Much of this effort evolves out of altruistic service to the community. Many established researchers (i.e. with tenure) can pursue these endeavors without concern that they

compromise focus on their individual research and advancement. How do we incentivise junior researchers and students to spend time on these necessary community tasks. This relates back to engaging students early on into a shared community.