Summary of 2019 Community Workshop

On April 12-13 of 2019 the Second FDSI Workshop was held in Boulder, Co. The workshop website can be found at [fdsi-full-community-workshop](). It started at 1:00 to allow morning travel from most US destinations and held two breakout sessions on the 12th and two breakout sessions on the 13th. The format of each breakout session was to have a 15 minute introduction by leaders in the field to define the topic and prime the discussion that followed in 60 minute discussions at tables of 10-20 participants each. After this hour of table discussion, each table would present a summary of the discussions that occurred in their breakout. The full output of the workshop was captured and is still available on GitHub at [Second-FDSI-Workshop]()

Session A

The topic of the first breakout in Session A was "Essential Components of FDSI" which was led Bob Moser and Scott Dawson. Note there was an attempt to pair a senior person with a (relatively) younger person in each breakout sessions leadership to expand the leadership beyond the PI team. Each table's discussion was captured in a google doc that all participants shared live at the table which can each be found here https://drive.google.com/drive/folders/1YXYM-YqWt_3EWel6n5e0mTHLzjJ3KnUB.
During this breakout, each group discussed potential components associated with a sustainable fluid dynamics software infrastructure. There were five dedicated breakout sessions: A-1) experimental data definition and data management, A-2) validation and other data analytics, A-3) model hierarchies and coordination of models, A-4) simulation problem definition, execution, and control and data management, and A-5) post-processing of data from experiments and simulations

In A-1), discussion focused on two points: i) the need to full describe experimental setups including boundary conditions and potentially inlet conditions and ii) the need for collaboration with CFD modelers to identify required data for proper definition and validation of CFD simulations, especially for coordinated experimental-computational studies. FDSI was identified as a vehicle for providing a community forum to help link experimentalists with computational scientists and engineers, as well as a vehicle for polling the CFD community on what data sets and experiments would be useful in validating their work.

In A-2), discussion focused broadly on the problem of validating numerical models using experimental data as well as higher-fidelity simulation data. A lack of standards as well as education were identified as particularly vexing issues in this regard. It was noted that the generation of community-accepted loss/likelihood functions for specific cases would enable clearer comparison of models, and FDSI could support such an effort through the development of common infrastructure for loss/likelihood function definition and exchange as well as the offering of educational activities such as best practices workshops and summer schools.

In A-3), discussion focused on the need to share data across a hierarchy of models (e.g., DNS, LES, RANS in the context of turbulence modeling and simulation), but there is a difficulty associated with sharing such data as i) the models are often implemented in different codebases and thus produce data in different formats and ii) data associated with different models in a hierarchy often describe different phenomena (e.g., RANS produces statistical quantities while LES and DNS produce time-instaneous quantities). FDSI was identified as a vessel for developing universal modules for analyzing data sets arising from a wide suite of CFD software packages as well as consolidating common data in one place, especially for standard test and validation cases.

In A-4), discussion focused on the issue that the CFD community does not have uniform, documented, and agreed upon standards of reporting simulation inputs and outputs. FDSI can address this issue through the establishment of a repository that includes enough, or all, of the information to simulate a specific instantiation of a problem, including a problem statement that includes a physical description of the problem and a description of the expected quantities of interest regardless of the method used to conduct the simulation.

In A-5), discussion focused on the fact that a lot of data is generated by experiments and simulations, but there is a lack of standards in terms of both data storage and data processing. Moreover, there is a barrier to entry for implementing state-of-the-art post-processing techniques. To address these issues, FDSI can help to create standards for input/output data, aid in the development of scale post-processing/visualization algorithms so that they may be easily applied to arbitrary experiments and simulations, help in the construction of a toolbox of post-processing/visualization algorithms, and provide documentation.

Session B

The topic of the second breakout in Session B was “FDSI Software Infrastructure Sharing” which was led by Charles Meneveau and Lorena Barba. Each table’s discussions can be found here https://drive.google.com/drive/folders/1uGkmXUVEEoHfSiLCVHrW0_B72rUJliN5.

During this breakout, each group took a slightly different angle to the overarching topic of software sharing.
The first group focused on the challenges of sharing common problem definition software. It was noted that this was easiest to interpret in the context of setting up the inputs and desired outputs from a CFD simulation as almost anyone who has done CFD simulations can identify challenges in setting up the same problem in one or more CFD solvers. While it is recognized that there are analogs to experiments, the experience of the participants in this group was mostly in CFD so the discussions naturally focused on CFD. In the area of pre-processing the most significant problems identified were: 1) different grids used (structured vs. unstructured), 2) specifying boundary conditions (especially inflow and outflow), 3) specifying/documenting the uncertainty or sensitivity in parameters (both in boundary condition and non-boundary/mesh solver parameters). It was noted that there were two, somewhat distinct use cases, for common problem definition software. The first use-case was to enable comparison of several solvers results to the same problem as is often done in workshops (e.g Drag Prediction workshop). In this case committee grids were observed to be very beneficial. The second use case involves generally making it easier for users to migrate their experience with one solver to different solvers for problems of their interest. This is a topic that needs more discussion as the former use case dominated the conversations and the subsequent report.

The second group was focused on the sharing of data analytics software. Here discussions echoed the importance of sharing data analysis software since it is inefficient for every research group to develop their own software BUT, in the current community environment, it can take a very long time to learn how to use someone else's code. The difficulty stems from the fact that other group's codes are usually poorly documented (there is seldom grant support to make documentation for software developed) and poorly supported (again, few labs have staff to support the software they develop). Clearly, this is an opportunity for FDSI as the software scientists that they would employ could gather software from the community, improve its sustainability (through best software engineering practices), improve its efficiency through improved algorithms, document its proper use with examples and user guides, and generally support further development with the community. The group also recognized that there was a role for FDSI to be a "matchmaker" between developers (who may know more about the math and algorithms of the data analytics) and the users (who may know more about the science they would like the software to elucidate).

This provides a natural segway to the third group that discussed what the software sharer/consumer want/needs from FDSI. In addition to the topics discussed in the previous paragraph, this group delved a bit deeper into the issue that software development is not incentivized in the promotion and tenure process. They also added more detail to the discussion of the issue of the challenges for many groups to employ the best practices in software engineering (e.g., version control, continuous integration, portability, etc. see document for more detail). As noted above, there is a clear need and role for FDSI to play in the education, leadership, and support of the community in sustainable software development.

This ties in very well with the fourth and final groups discussion of existing examples from different fluid dynamics areas. This group discussed existing community fluids codes (e.g., WRF), research-group-level codes that are currently shared, and software libraries that are not fluids-specific but have been used with success in several codes (e.g., GMSH for meshing, DAKOTA for UQ, etc.). A fairly extensive list was developed in this one hour and FDSI should work to grow this list further to be more inclusive. After the list is grown, it will likely be important for FDSI to determine a subset of packages/components that the best combination of high community interest, high and broad potential community impact, and adequate engagement by that software development community to focus on near term versus longer term. This group's breakout report closed with the vision that FDSI will provide a site / platform / facility that provides higher level vetting and support to enable quality software to thrive, and less reliable software to find methods to improve.

Lightning Round

To build connections, each participant was invited to present a slide on the topic of their choice in a Lightning Round of presentations that kicked off the day on Saturday. Almost all participated and this was viewed as a huge success. While it might have been nice to have held it at the start of the workshop, such a plan would have missed participants with travel delays or who otherwise could only join for the Saturday part of the meeting.

Session C

The topic of the third breakout in Session C was "Community-Sourced Plan" led by Beverley McKeon and Andres Tejada-Martinez. Each table's discussions can be found here https://drive.google.com/drive/folders/1a50BFr9t9Cpq74rRU-8SRhKXzbbsvvN4.
<INSERT summary of breakout C here>

The discussions focused around engagement with the community, in terms of further defining the need for the Institute and associated goals, means for long-term sustaining the community effort, and proposals for educational and outreach programs within FDSI.

The need for a communal organization such as that embodied by FDSI to connect communities working in areas related to fluid dynamics was clearly elaborated and agreed upon. There is a clear need for a mechanism to connect communities involved in algorithmic and infrastructure development with those exploiting such tools for physics discovery. The fluid dynamic problems of today are of a complexity that it is not realistic for researchers to span the full range of tool development and discovery; there is a significant opportunity to advance science by creating a forum for connecting researchers and practitioners with a common goal. Establishing standards for reporting and archiving would be a very valuable early goal for FDSI, as would identifying and documenting shared computational codes and procedures. Associated with these steps should be a strong training component, targeting the education of a new generation of graduate students and advanced researchers equipped with knowledge of validation and verification techniques and armed with community validated computational and data analysis tools.

With regards to long-term sustenance of FDSI, we view the centralization of fluid dynamics techniques within the Institute as a valuable commodity which can be exploited via a conference series intended to educate and disseminate the outcomes of the Institute. With an appropriate fee structure, such a series could provide a steady annual or biennial stream sufficient to maintain operations of the Institute at a subsistence level.

Further, both the establishment of standards of practice in the area and the balanced enabling of innovation offer opportunities for engagement with industry, especially with regards to student internships within the Institute targeted at solving industry-guided issues.

FDSI occupies a unique space with regards to integrating between disciplines which can be exploited for the development of courses targeted at graduate students and also continuing education.

Many effective outreach strategies for FDSI were identified, with a view to expanding the pool of options rather than duplicating those already in place through professional societies.

Session D

The topic of the third breakout in Session C was "Governance, Leadership Team Composition, and Prioritization of FDSI Budget/Effort" led by Jed Brown and Magarette Jadamec. Each table's discussions can be found here https://drive.google.com/drive/folders/1pL6RiQq2CCLSlHFkIiVKMLdo06ultyNG. Having only four sessions forced consolidation of two very important topics into a single session. While governance are budget prioritizations are related, they also have very distinct elements and would have been worthy of full discussions if time had allowed.

The first table took on the task of discussing a governance model for FDSI. Early discussions centered upon identifying the key constituents of the community. Here two categorizations were identified. The first was based on affiliation/sector (academic institutions, industry, national labs). The second was based on interests (software share, software use, data share, data use, technical area). The discussions recognized the need for FDSI to quickly and steadily grow more broad community participation. Engagement in the governance process is essential to that growth and thus the group discussed modes that have proven successful for other community-based efforts. Forming an incorporated non-profit (e.g., NumFOCUS) was discussed. Engagement of industry was discussed with a potential to associate some "voting" or direction based on level of support but this raised the challenging question of keeping a balance between the academic community needs (with their "votes"/direction) balanced with industry. There was also considerable discussion of the organizational structure. Consistent with the original proposal, an executive committee with representation from academia, national labs, industry, and at large was discussed with options to appoint vs. being elected, possibly having each constituent community vote separately their representation.

The second breakout group discussed the tradeoff between resource allocation for improving existing software to make it more sustainable vs developing new software to meet community identified needs. The premise, that under a finite budget to hire software engineers their time would likely be split in some manner between these two categories. It was discussed that MOLSSI had faced the same issues. It was posed that it would have to be part of the governance structure to make these resource allocation decisions with guidance from the community and oversight from the executive committee. Similar discussions regarding whether those hired by the center should be collocated (as MOLSSI does) or distributed to multiple software excellence centers. Without getting into all of the details which are available at the link above, the group had a lively discussion about how an NSF funded Implementation of FDSI might allocate its budget and how it could transition itself to a sustainable software institute.

The third breakout group discussed resource allocation to education and outreach. A diverse set of activities in this area included workshops, hackathons, development of educational modules, fellowships for code documentation, tutorials, social media, and K-12 engagement. The relative cost and impact of these activities was vigorously discussed.

The fourth breakout discussed resource allocation for data stewardship. In addition to the software side of this topic which includes archiving, analysis tools, searching, the group also discussed governance/policy issues such as meta-data standards and best practices which should be shared throughout the community. There was further discussion of hardware within FDSI that might physically host fluid dynamics data from computations and experiments. While this was generally considered to be valuable, the usual concerns about sustainability were also discussed.

Closing Discussion

The final session of the workshop was an open discussion on the topic of "Community Priorities and Next Steps Discussion" Notes from this can be found at this link on the third page. Much of the discussion reiterated the important points from the sections above but an excellent new topic was raised-- How can FDSI broaden participation and diversity. While it is true that the three disciplines that train fluid dynamicists and data scientists (e.g., engineering, mathematics, and computer science) all suffer from small pipelines that need to be improved via STEM efforts, the bigger problem facing the fluid dynamics community is that the pipeline leaks--underrepresented groups and women that should today be in a position of leadership have left the field. This topic was discussed vigorously and a consensus was reached that the leaky pipeline requires confronting biases, both explicit and implicit, in our community at all levels There was a strong sentiment that FDSI had an opportunity to play a strong role in raising the discussion of this issue within the fluid dynamics community and thereby be an agent of real change.