Data Analysis and Interpretation: Charge and Scope

- **Charge:** Look at the future of analysis at scales up to that of the HL-LHC
  - Future analysis models
  - Sociology of data access in groups
  - Challenges and opportunities surrounding analysis code development and infrastructure
  - Approaches to preserve and disseminate knowledge

- **Scope:** Current and future analysis techniques and the toolkits that support them
  - Consider how factors like data volume, type and parallelism of data access, and number of parallel analyses impact the ability to carry out a comprehensive physics program.

- **Not a goal:** Reach community consensus
- **Goal:** Create a roadmap for R&D projects
  - Experiments can develop an informed plan for the HL-LHC era (pick and choose)
  - Create a framework for contacts and discussions to work together on similar aspects of the R&D roadmap

Visualization (everything that turns data into pictures) includes now also histograms, etc.

Discussion during the workshop

Structure the topics of the paper:

1. Challenge
2. Use cases
3. Roadmap

Where the roadmap should discuss plans for

- The first year
- Up to year 3 (TDR time scale)
- Up to year 5 (Run 3)

Roadmap

1-year time frame

- enabling new open source tools to be plugged in dynamically in the existing ecosystem and mechanisms to dynamically exchange parts of the ecosystem with new components
- Develop requirements and design a next generation analysis facility concept, incorporating fast caching technologies to explore a query-based analysis approach
- finalize full support of python in our ecosystem including long term maintenance
- evolve policies to minimize this effort by retiring less used components from the integration and validation efforts

3-year time frame

- research a comprehensive set of bridges and ferries, where a bridge enables the ecosystem to use an open source analysis tool and a ferry allows to use data from the ecosystem in the tool and vice versa
- Analysis facility: conceptualization of an end-to-end low-latency response high-capacity analysis facility with an initial prototype
- develop a prototype functional or declarative programming language model
- evolve our policies how to replace components with new tools, maybe external, and solicit the community help in bridging and integrating it

5-year time frame

- Analysis facility: evaluate chosen architectures and verify design or provide input for corrective actions (Run3 data should be used)
- in year 5, a blueprint for remaining developments and system design should become available, in time for deployment.

Missing: Finalizing Analysis Model Chapter, Adding chapter about Analysis Interpretation (statistical tools)

Paper draft Special Thanks to Lothar, JimP, Matt and everyone who helped

Next Steps

- Paper draft
  - Topics are complete (except statistical tools, but that is under control)
  - Need work on the text
- Want to work more on interconnection with other WGs
  - Talked to Preservation WG, Framework WG, Computing Model WG (through text contribution), Data management WG
  - Saw overlaps with ML WG, etc.
- End of July deadline
  - Think we can make it
  - Will send around draft to wider audience as soon as above steps are complete