OPTIMIST HOUR

File Working Group

Fatemeh (Saba) Ganji for OPTIMIST

2346138

Why an open-source ecosystem?

- Problem?
  - Number of engaging people using CW
    - Addressing the need for open-source tools, APIs, etc.
- Realizing an eco-system from already-established connections
  - Open-source ecosystem != open-source product
  - GitHub != Open-source ecosystem

You are a stakeholder!

- Citizens of the ecosystem that derive value
  - Key partners
  - Key activities
    - Key resources
- Value proposition
- Stakeholder segments
- Stakeholder relationships
  - Channels
  - Sign-up!
    - What you need for that an email address!
    - WHY?

OPTIMIST open source ecosystem

Value Propositions

- Helping *Learners* become *Doers*
- Reduce the barrier to entry (increase learners)
- Central Index with Artifacts, Datasets
- Solid HOWTO Documentation

Key Activities

- Define/Refine Common Interfaces
  - Target Hardware/Firmware
  - Measurement Hardware
  - Analysis Techniques
  - Datasets

Customer Segments

Learners
- Educators
- Students
- Early-stage Researchers

Doers
- Advanced Researchers
- Test Engineers
- Hackers

Key Partners

- Expert Users
- Standardization Bodies

File format vs. storage structures

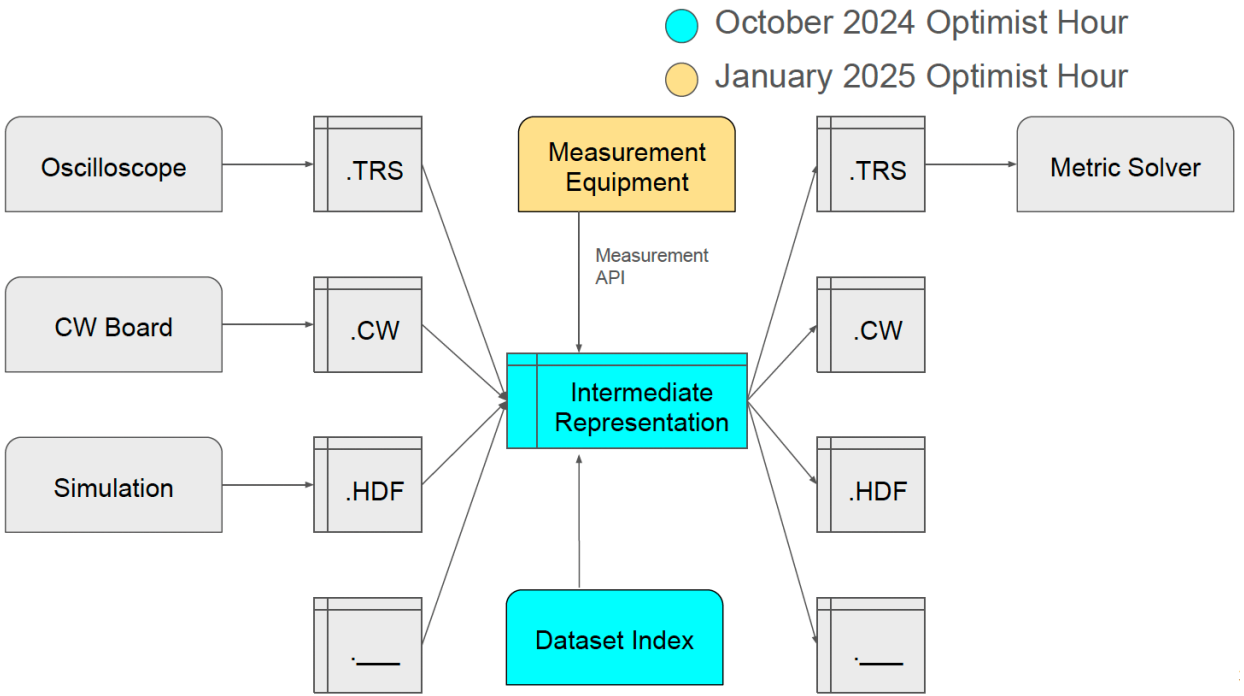
File format

- File format – type of datasets
- Data format
  - Numpy, trs, etc.
- Compressed vs uncompressed
  - Zip, tar, etc.
- Opensource compatibility
- Conversion between file formats
  - API: no reliable universal implementation
  - Documentation: scattered/unavailable
  - Learning curve: ~10 hours for beginner
  - A barrier to entry for students

Storage structures

- The hierarchy of data structure
  - json, hdf, zarr custom, etc.
- File sizes
- Multiple files vs single files for a trace set
- Accompanying information: documentation, scripts, metadata, etc.
- Test scripts
- Opensource compatibility
- Going hand in hand with file format

Can formats converge with each other?



Goals

- Foster collaboration and modularization
  - Groups can still collect traces using their convenient setups and formats
  - Datasets are more accessible
- Avoiding duplicate work
  - Incorporating different datasets into an analyzer's flow
- Promoting consistent notation
- Interoperability within a template approach

Missions

- A public draft of requirements for an intermediate representation
  - Allowing for additional public feedback
  - OPTIMIST facilitates different teams coming together to work on the draft
  - The document includes some common parts (e.g., “preliminaries”), defining consistent notation, building blocks, assumptions, and references.
- Benefits to working group?
  - Complete overview of (previously) adopted file format, including its pros and cons
  - Visibility
  - Reduce the barrier to entry (increase the number of learners)

Suggested discussion points

- What will the intermediate representation look like?
  - A standalone file format that can be converted to and from.
  - An API with a series of functions that can read and store in varying formats.
  - Some other solution.
- How will metadata be handled?
- How will users interact with/use the intermediate representation?
- Is an intermediate representation even the right path?

Let’s talk and take action!