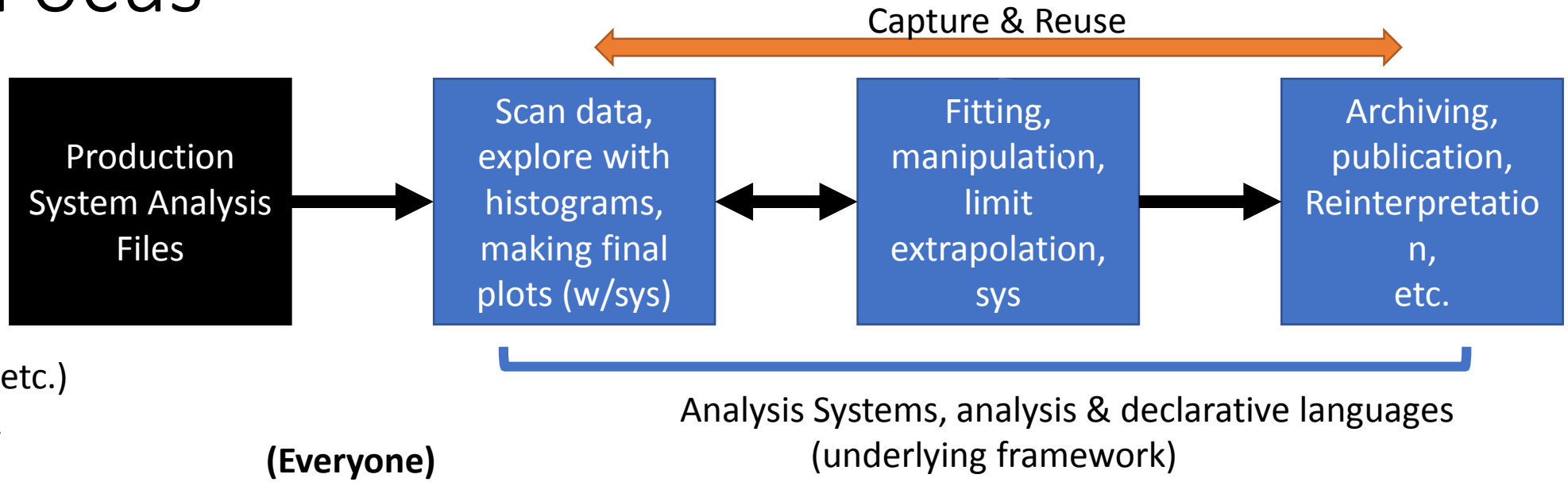
PI Chat
University of Washington

G. Watts (UW/Seattle)

Group Focus



<u>Analysis Systems</u>

- Interactive
- Full Scope (systematics, ML, etc.)
- Modern tooling & capabilities
- Links into production to enable analysis

(Everyone)

<u>Management</u>

- US ATLAS/ATLAS connection
- Steering Board
- ~External HEP Coordination
- Helping with EB meetings & running Day-to-Day
- Pester people about an advisory board meeting

(G. Watts)

1

Who is actively working on the project from the group and what are they doing?

Group leader
- func_adl to ATLAS xAOD
- func_adl to CMS Run 1 AOD (w/ fellow)
- hep_tables
- Open Data DID Finder
- ServiceX/coffea/func_adl integration
- ServiceX

Graduate Student
- func_adl to flat ROOT (uproot)
- Analysis Description Language Benchmarks
- General python ecosystem integration with (his) analysis

Postdoc: Left last June
- func_adl to RDataFrame

Tal is just starting… (see future discussions)

What projects are connected with the institute's grand challenges?

Analysis Grand Challenge

The ~200 TB of data will start with ServiceX and a func_adl query

Initial data files will be:
- ROOT Ttree's
- ATLAS xAOD's
- CMS OpenData

The first two are well implemented, and the second is being brought up

Tools like hep_tables, if successful, may be a more concise way of expression simple queries to ServiceX

We have started an effort linking ServiceX, Coffea, Awkward, DASK
- If successful, may be the way we eventually do scale out for the grand challenge
- N.B. UW is far from the only group working on this: Ben is pushing hard on this currently in the context of the Nebraska coffea-casa AF

2

What are contributions of those individuals in your group to the various area(s), i.e. AS, IA, DOMA, SSL, SSC, OSG-LHC, MGMT

	Analysis Systems	DOMA (SX)	Management
Gordon Watts	40%	10%	50%
Mason Proffitt	100%		
Tal van Daleen	100%		

What specific collaborations has the group made with other groups within IRIS-HEP?

(I’m ignoring management responsibilities)

Champaign-Urbana

Close working relationship with Ben with both myself and Mason

UChicago

Building a closer working relationship with ServiceX team there (like Andrew)

NYU

Starting one with Alex Held (see future directions discussion)

4

What specific collaborations has the group made with other groups or individuals external to IRIS-HEP? What are plans to integrate the work with the wider HEP community? (I'm ignoring management responsibilities)

U Texas

Working with P. Onyisi and K. Choi around ServiceX. The uproot backend which is part of their recent TRexFitter demo; TCut to qastle translator

ZTH

- Recent collaboration with CS group lead by Muller
- Paper submitted to CS journal, will be posted to archive after first comments back
- Compares ROOT, BigQuery (Google), Athena (Amazon), Rumble, etc. on the benchmarks we created in IRIS-HEP
- Will invite for topical meeting, etc.

Evaluating Query Languages and Systems for High-Energy Physics Data

[Experiments, Analyses, and Benchmarks]

Dan Graur
Department of Computer Science
ETH Zurich
dan.graur@inf.ethz.ch

Ingo Müller
Department of Computer Science
ETH Zurich
ingo.mueller@inf.ethz.ch

Ghislain Fourny
Department of Computer Science
ETH Zurich
ghislain.fourny@inf.ethz.ch

Gordon T. Watts
Department of Physics
University of Washington
gwatts@uw.edu

Mason Proffitt
Department of Physics
University of Washington
masonLp@uw.edu

Gustavo Alonso
Department of Computer Science
ETH Zurich
alonso@inf.ethz.ch

Where does this work go next?

What specific collaborations has the group made with other groups or individuals external to IRIS-HEP? What are plans to integrate the work with the wider HEP community? (I'm ignoring management responsibilities)

PONDD – Amy Roberts

- Helped write the grant (we were originally on it)
- Will participate in making the SX aspects of it a success

There are possibly others…

Tiled group (Tom Caswell, Dan Allen) - BNL

- Tiled serves data, combines catalog, format translator
- Some parts of what are missing from Sx
- Collaborating to adapt it to awkward data types
- Aimed at smaller experiments

5

Which contributions has the group made to the "Intellectual Hub" aspect of IRIS-HEP? (Including community building activities, training, outreach and broader impact.)

We are very present on Slack

We evangelize Slack and IRIS-HEP inside ATLAS and in the HSF

Mason has helped with training (and is interested in doing it more)

Tracking some of our beyond-HEP acitivites

2 Fellows

- David Liu (physics), last summer, partially successful building a test suite for ServiceX.
- Biadyanath Kundu, Current, CMS Run 1 AOD C++ backend transformer targeted at CMS Open Data.

Potential Third Fellow

- ServiceX as a RDataFrame data-source

Please list any papers/presentations/software-repositories connected to the above

Github Organization	Repos
Iris-hep	• adl-benchmarks-index • func_adl_xAOD • func_adl_uproot • func_adl_servicex • qastle • func_adl
Ssl-hep	• ServiceX • ServiceX_uproot_transformer • ServiceX_cpp_xAOD_transformer • ServiceX-DID-finder • ServiceX_coffea • Servicex_clients • servicex-backend-tests • Servicex_CodeGen_xAOD
gordonwatts	• hep_tables • dataframe_expressions

Plus lots of examples, scratch work, etc.

How is your project managed? How is progress measured? How are risks identified and mitigated?

Management is informal:

- Coordinate project development using github milestones and project Kanban boards
- Meetings are informal and semi-regular
- Regular reports at Analysis Systems meeting and ServiceX meeting

Risks

- Frequent presentations and conversations via Slack to understand community pain points
- Turned into issues or projects (e.g. hep_tables)
- Discussions in the Analysis Systems meeting

Are metrics aligned with the project? Do metrics accurately reflect the progress, success or impact of the project?

Metrics are monitored at the Analysis Systems level for this work

Integration into the Analysis Challenge means there are constant integration milestones

Future Plans

New money from IRIS-HEP

1. Continue with Deputy Director Position
2. Worried about IRIS-HEP projects and ROOT's integration
   - RDataFrame as a source
   - New hire, 50% to work on C++ and ROOT issues important to IRIS-HEP and Analysis
3. Evolve physicist interface to ServiceX and coffea
   - Supporting multi-tasking at SX and beyond level with little code requirements on physicist
   - func_adl is complete, but perhaps too general, awkward is much more concise for simple tasks
     - Use hep_tables to explore better ways to bridge this gap in understandability and capability
   - Support linkage to coffea processors, or deep integration into DASK-like awkward DataFrames if Jim and others head in that direction.
4. Cabinetry Interface to ServiceX and func_adl and DASK and coffea, differentiable programming pipeline
   - Tal is well positioned to work with Alex H (short commute)
   - New effort starting with Tal
5. Fill out the SX backends
   - xAOD: Run 3 dataformat (new software release), implement a few more primitives
   - Uproot: add primitives as they are needed. Will loose Mason in about a year; need to understand sustainability of this
   - CMS Run 1 AOD for OpenData: Complete and test against CMS examples. This will be the end of the CMS fellow: need to decide where to go next
     - Doesn't really make sense for a ATLAS group to evolve this beyond open data

Future Plans

New Hires at UW (possibly)

1. Post-doc or Staff (new money)
    - 50% on IRIS-HEP, connect with the ROOT team, Analysis related topics, helping connect IRIS-HEP analysis to this effort
    - 50% from US ATLAS working on low level I/O, etc., in ROOT
2. Physics Post-doc or Graduate Student (burning down an older grant)
    - Shift Tal to 50% and replace other 50% with this person
    - Support for 2 years, perhaps more
    - Tal and this person would work on physics and IRIS-HEP to connect the two efforts
    - Use our tools in Run 3
3. Tal (new hire, IRIS-HEP currently)
    - Similar to above person