Welcome to the workshop on

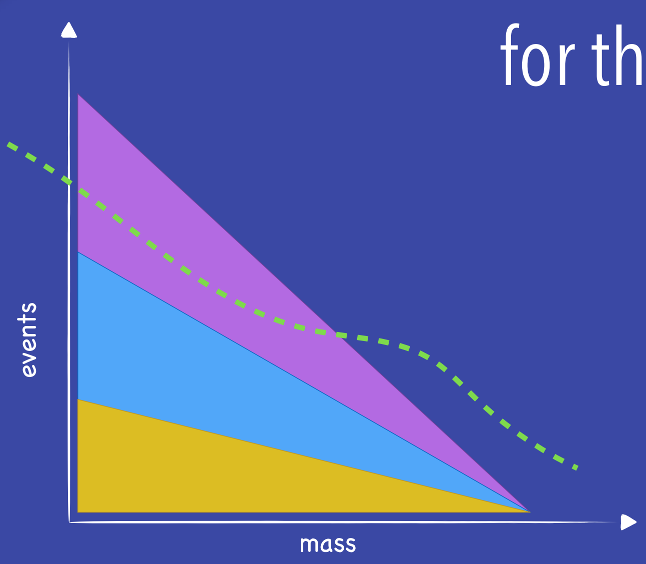
Analysis Description Languages

for the LHC

6-8 May 2019, Fermilab LPC

Sezen Sekmen (KNU)

for the organizers



Maximize the scientific Impact of the LHC

accessible **data**

accessible **simulation**

Maximize the Impact of LHC

accessible **results**

accessible **analysis information**

This workshop

General purpose or domain specific?

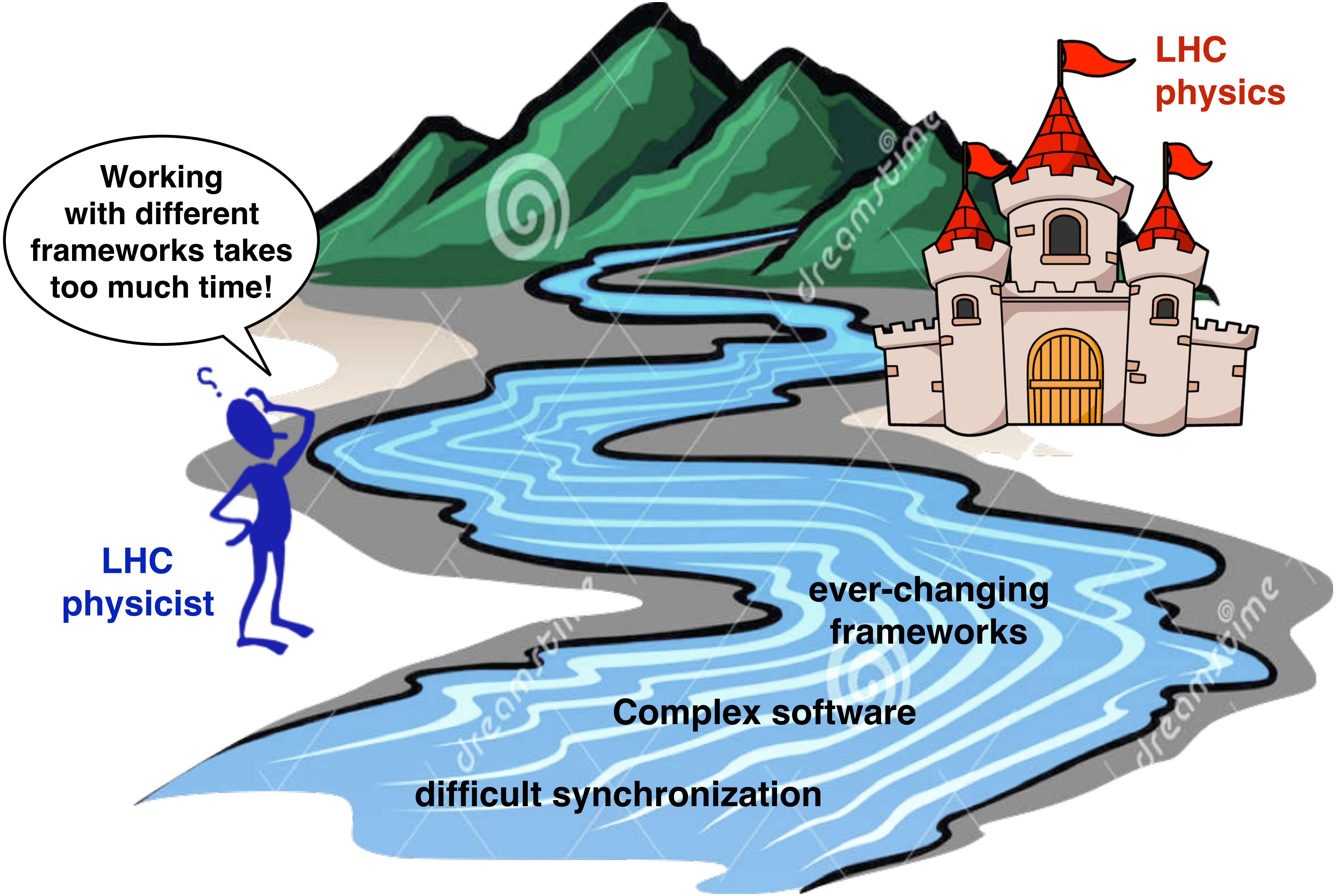
- Every discipline has its own way of thinking, its own concepts, and its own way of solving domain specific problems.
- In particular physics, we have generally used general purpose languages (GPL), first FORTRAN, now C++ and Python, to solve problems such as implementing analyses.
    - The advantage of GPLs is that they can solve a very large class of problems.
    - The disadvantage is that they do not model how a domain expert thinks about her or his domain of expertise.
- A domain specific language (DSL) is designed to model how domain experts think about the problems they wish to solve.

Analysis Description Languages for the LHC

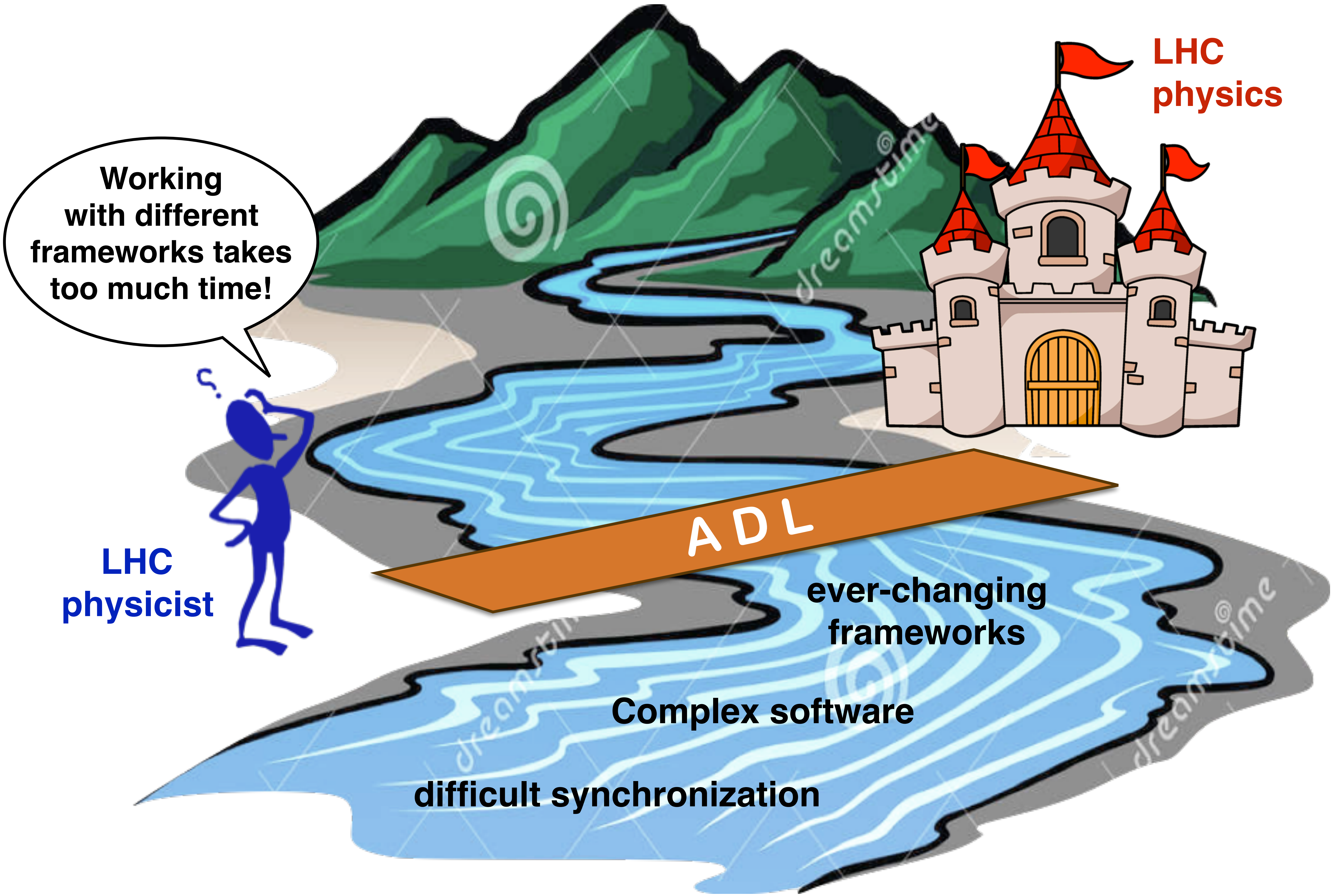
An Analysis Description Language (ADL) for the LHC could be defined as:

- A domain specific language capable of describing the contents of an LHC analysis in an unambiguous way. The potential scope for an ADL could include:
  - object and event selections
  - analytically and algorithmically defined observables, including those based on machine learning, efficiency maps, etc.
- Designed for use by anyone with an interest in, and knowledge of, LHC physics to abstract, visualize, validate, combine, reproduce, interpret, and communicate the contents of LHC analyses.
- Earlier HEP formats/languages proved successful and useful:
  - Les Houches Event Accord
  - SUSY Les Houches Accord

We have all been there...



We have all been there…



Principles for an LHC ADL

The principles of an analysis description language were defined in the Les Houches 2015 new physics WG report (arXiv:1605.02684)

Towards an analysis description accord for the LHC

D. Barducci, A. Buckley, G. Chalons, E. Conte, N. Desai, N. de Filippis, B. Fuks, P. Gras, S. Kraml, S. Kulkarni, U. Laa, M. Papucci, C. Pollard, H. B. Prosper, K. Sakurai, D. Schmeier, S. Sekmen, D. Sengupta, J. Sonneveld, J. Tattersall, G. Unel, W. Waltenberger, A. Weiler.

Abstract: We discuss the concept of an “analysis description accord" for LHC analyses, a format capable of describing the contents of an analysis in a standard and unambiguous way. We present the motivation for such an accord, the requirements upon it, and an initial discussion of the merits of several implementation approaches. With this, we hope to initiate a community-wide discussion that will yield, in due course, an actual accord.

Features of an ADL for the LHC

Basic requirements:

- Public
- Complete
- Easily learned
- Demonstrably correct

Desirable features:

- Self-contained
- General programming language-independent
- Analysis framework-independent

ADLs are good for everyone

Motivation / use case	Exp	TH/ Pheno	Public
Analysis preservation	✔	✔	✔
Analysis design, implementation	✔	✔	✔
Analysis communication, clarification, synchronization, visualization	✔	✔	✔
Analysis review by referees	✔	✔	
Interpretation studies, analysis reimplementation	✔	✔	✔
Easier comparison/combination of analyses	✔	✔	✔

ADL Scope

By construction, an ADL is not designed to be general purpose; therefore, getting its scope right is key.

- The **core** of any ADL for the LHC should include
  - Simple and composite object definitions
  - Observable definitions
  - Event selection definitions, event weighting
  - (Optionally) Standard reports and visualizations
  - … ?
- Further operations with selected events (background estimation methods, scale factor derivations, etc.) can vary greatly, and thus may not easily be considered within the ADL scope.

Purpose of this workshop

- **Learn** about DSLs, the ADL concept and the existing ADL prototypes and parsing/interpreting methods and tools through talks and hands-on exercises.
- **Discuss** ADL scope, physics content, technical details on language structure, syntax, parsing/interpreting methods, and realistic usage in full-fledged experimental analyses.
  - **Identify difficult cases** in language structure and parsing, and start discussing solutions.
- **Devise a roadmap** to proceed from current prototypes towards complete solutions.
  - Everyone is welcome to provide input and join the effort!

In this workshop: Talks

We will have several talks on existing efforts for building ADLs and tools to work with ADLs:

- **LHADA —> ADL, adl2tnm** (Harrison Prosper)
- **CutLang —> ADL** (Gökhan Ünel)
- **lhada2tnm** (Philippe Gras)
- **LINQ** (Gordon Watts)
- **YAML as ADL** (Ben Krikler)
- **NAIL**: A prototype analysis language on top of RDataFrame (Andrea Rizzi)
- **TTreeFormula** (Philippe Canal)
- **AEACUS** and **RHADAMANTUS** (Joel Walker)
- Fermilab colloquium: **Particle physics and programming languages** (Jim Pivarski - Wednesday, 4:30 PM.

In this workshop: Hands-on sessions

Two hands-on sessions on Monday, May 6th:

- **ADL / CutLang** (H. Prosper, S. Sekmen, G. Ünel)
  - Will include writing an analysis of your own choice using the ADL/CutLang format.
- **Language-making tools** (Jim Pivarski)
  - Formal introduction to language building and parsing.

Please join the **Mattermost channel ADL4LHCatLPC** for questions and technical discussions.

https://mattermost.web.cern.ch/signup_user_complete/?id=nwmm8c714jdrtxyu3drpyugt5r

https://mattermost.web.cern.ch/adl4lhcatlpc/channels/town-square

In this workshop: Discussions

- **Where/for what do we need a domain specific language?**
  - What different purposes can ADLs serve?
  - One or multiple ADLs?
  - The scope of ADLs? selecting events? filling histograms? fitting?
  - How would ADLs improve our way of thinking about our analyses?
- **What physics specific content should be included?**
  - Which elements of object and event operations can be described in an ADL? Where do we stop?
  - What logical / mathematical operations do we use?
  - How far do we go in describing composite objects, object groups, event variables?

In this workshop: Discussions

- **What are the language users' requirements?**
  - How to go from an ADL to performing a full fledged experimental analysis?
  - How much do the current interpreters meet the analysts' needs? What is missing?
  - Can we provide generic tools to parse ADLs in experimental analysis frameworks?
  - How can we use ADLs for combination of analyses?

In this workshop: Discussions

- **What kind of language/syntax do we need?**
  - Most generic and economic ways to cover the widest collection of needs from the biggest variety of analyses?
  - Objects, composite objects, event variables, definitions, selections, event weighting, …
  - Intrinsic loops, reducers, optimizers, object combinations/permutations/partitioning, ternary operations, …
  - Expressing the event input
  - Expressing external functions

In this workshop: Discussions

- **What tooling (parser / interpreter / compiler) do we need?**
  - Interpreters? Transpilers? Which method for which purpose?
  - How to create the most flexible tools for most generic usage?
  - Issues of speed, user-friendliness?
  - How to automate incorporation of event input formats? How to deal with different event formats? Common extensible object(s)?
  - How to automate incorporation of external functions?
  - Histogramming, event weighting?

In this workshop: Discussions

- **ADLs for analysis preservation**
  - What are the central LHC analysis preservation tools?
  - How can ADLs help analysis preservation? How can they be incorporated into the system?
  - Databases for event input formats, standard object definitions, event variables, external functions… ?
- **Feedback from other experiments**
  - What do physics analyses consist of in other experiments?
  - Would ADLs for other experiments be built on similar concepts as in ATLAS/CMS?
  - What can ATLAS/CMS learn from the other experiments?
- **How to move forward**
  - Summarizing the next steps and planning how to proceed.

In this workshop: Logistics

- **Dinner : Tuesday, at 7pm** at **Indian Harvest** in Naperville.
  For **car sharing** information, please see indico.

  *** Please confirm your participation **until the end of the afternoon coffee break today**!

- **Group photo : Tuesday morning coffee break**.

- Please **register and get your badge**!

Thanks…

Organizing committee:
Steve Mrenna (Fermilab)
Jim Pivarski (Princeton U.)
Harrison Prosper (Florida State U.)
Sezen Sekmen (Kyungpook Nat. U.)
Gökhan Ünel (U.C. Irvine)

LPC coordinators:
Cecilia Gerber (UIC)
Sergo Jindariani (Fermilab)

Local organization:
Gabriele Benelli (Brown U.)
Alexx Perloff (U. Colorado Boulder)
Marc Weinberg (Carnegie Mellon U.)

LPC events committee:
Gabriele Benelli (Brown U.)
Ben Kreis (Fermilab)
Kevin Pedro (Fermilab)

LPC administrators:
Carrie Farver
Terry Read
Sonya Wright

Workshop on

Analysis Description Languages

for the LHC

let's get started…