Extending RECAST for Truth-Level Reinterpretations

ALEX SCHUY (UNIVERSITY OF WASHINGTON), LUKAS HEINRICH (CERN), KYLE CRANMER (NEW YORK UNIVERSITY), SHIH-CHIEH HSU (UNIVERSITY OF WASHINGTON)

State of RECAST Development

Past

- Collaboration-specific analysis preservation & reinterpretation using internal tools

Present

- Truth-level reinterpretation using external (public) tools

Future

- Expanded public catalogue
- Smart grid selection
- REANA backend

Past

COLLABORATION-SPECIFIC ANALYSIS PRESERVATION & REINTERPRETATION USING INTERNAL TOOLS

01

There is a **steadily growing** number of **theoretical models** to analyze, but making analyses is **time-consuming**.

02

Analyses often utilize signatures that make them **sensitive to a wide range of models**, more than are initially explored.

03

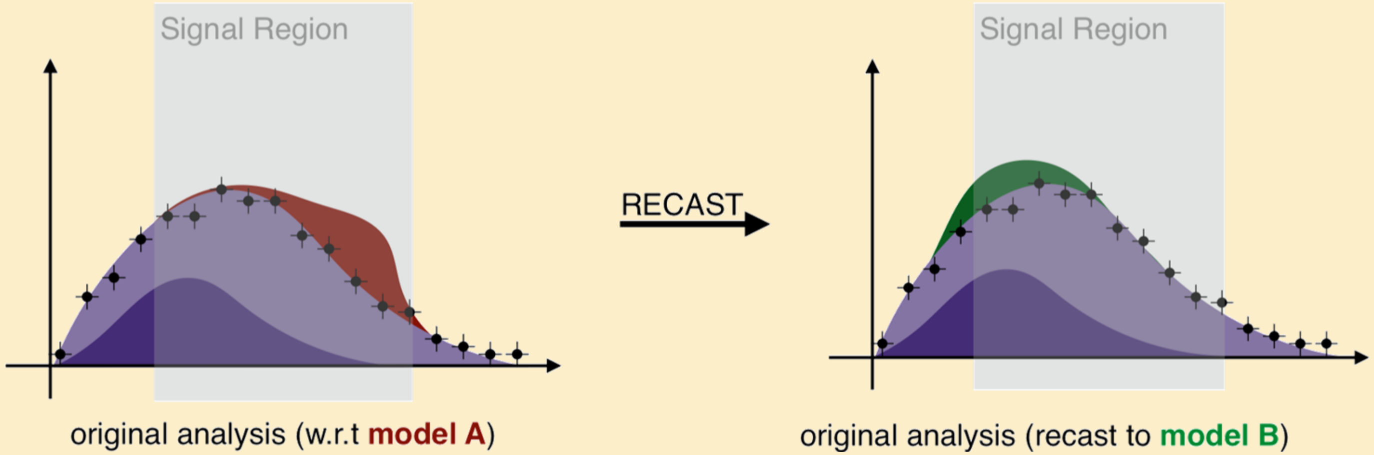
Make the most of LHC results for years to come by **preserving full analyses**.

Motivation

Idea

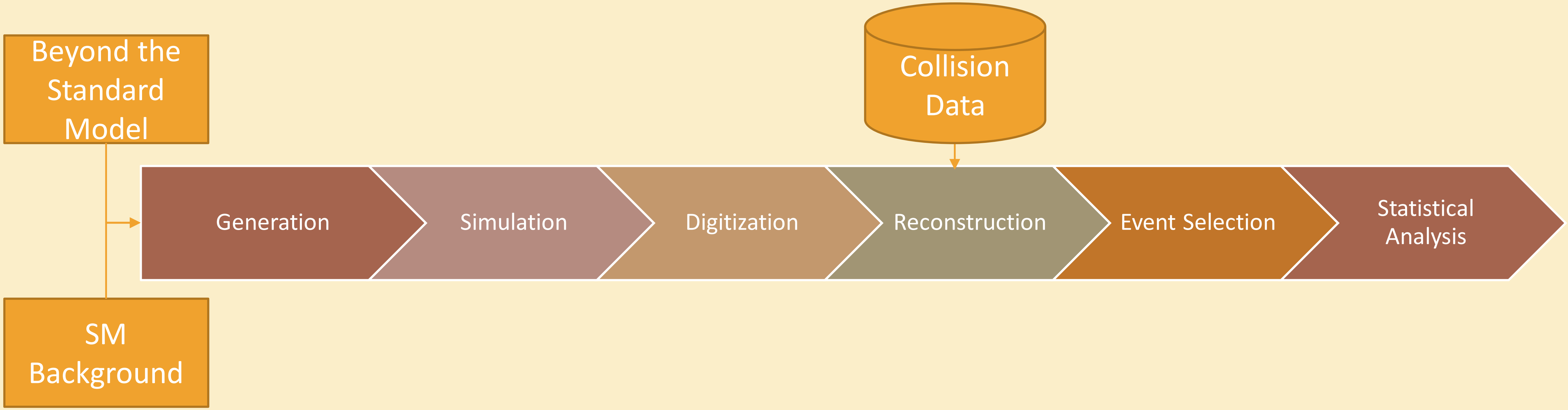
Useful limits can be found for a new model by **exploiting an existing analysis with a similar experimental signature.**

All we need to do is **generate new signal samples** and pass them + **existing background** into the analysis.

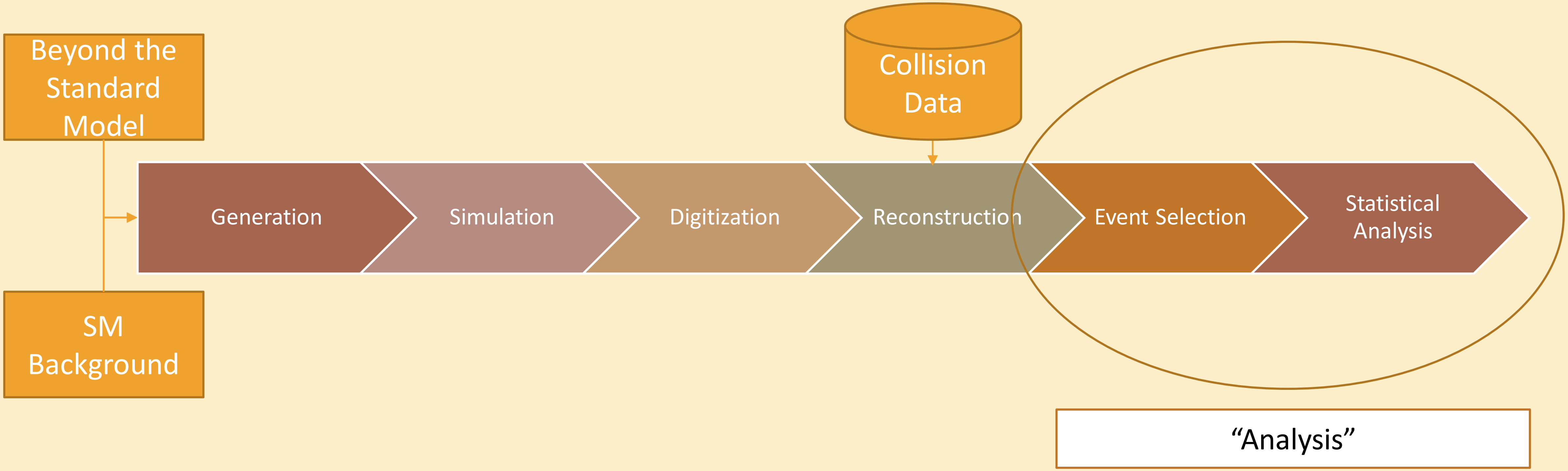


See https://indico.cern.ch/event/639314/contributions/2726367/attachments/1542484/2419628/ReinterpretationWorkshopFNAL.pdf

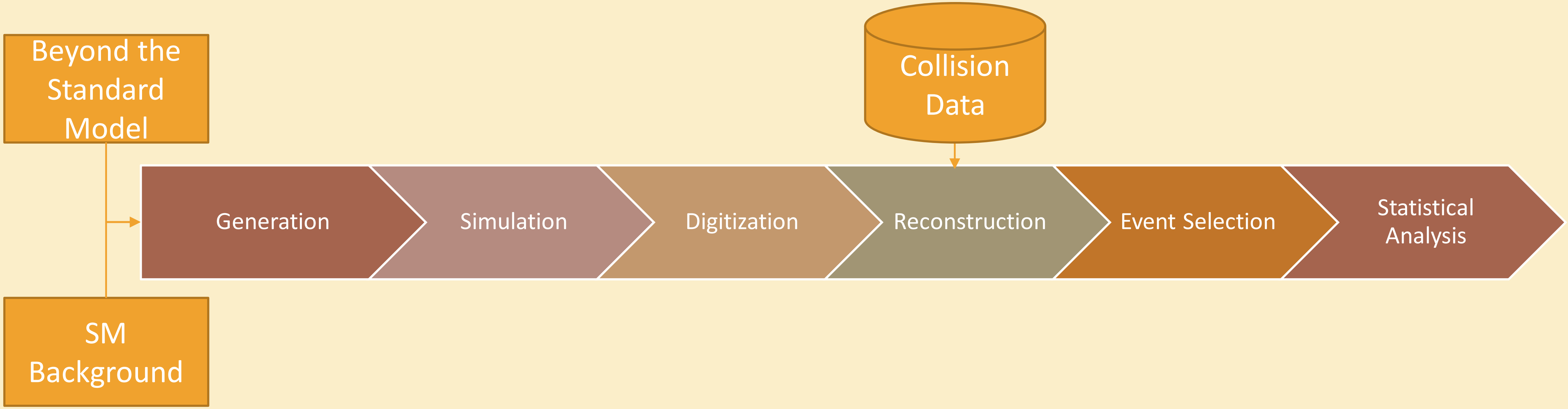
Typical Search Workflow



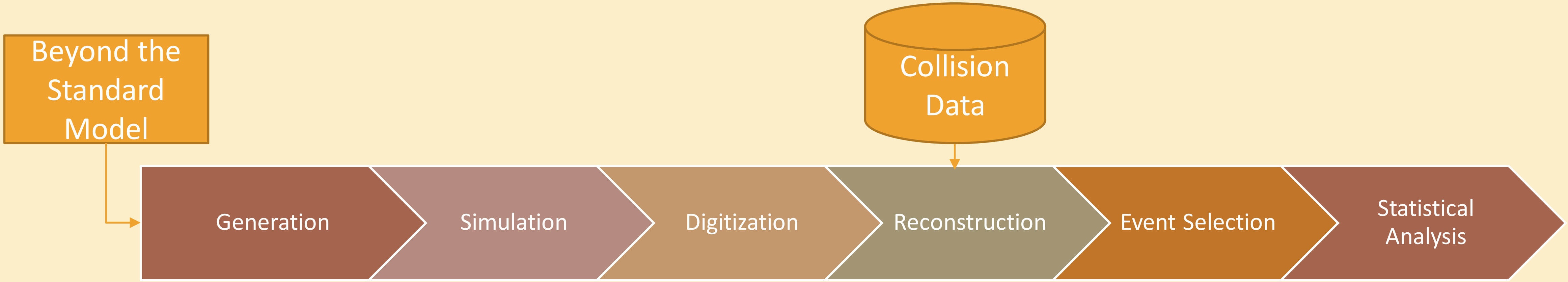
Typical Search Workflow



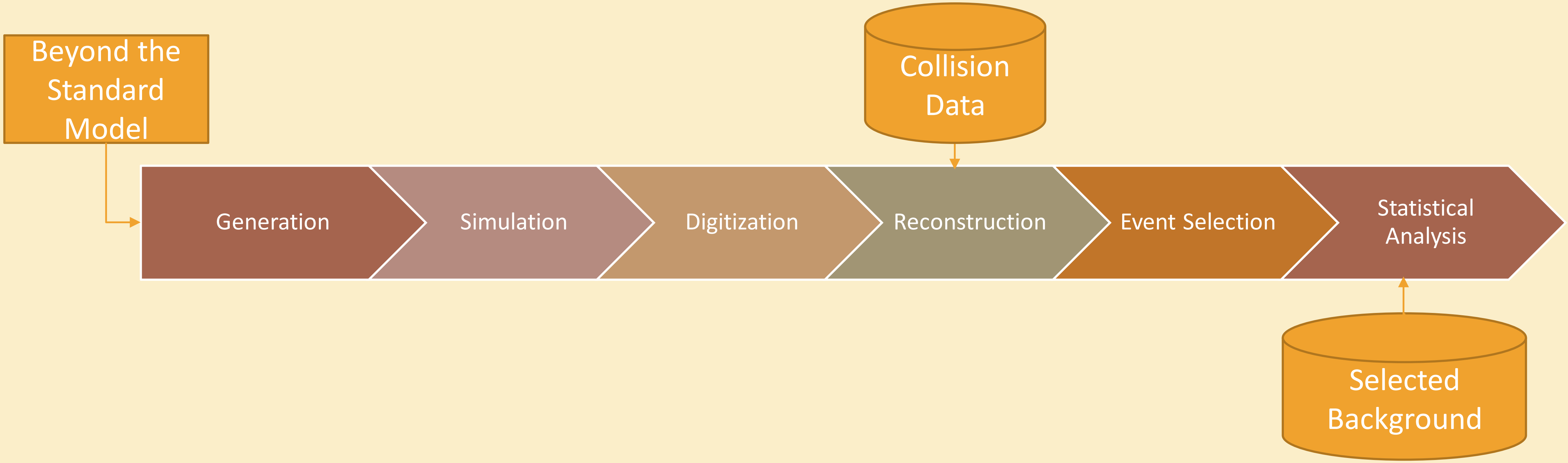
Recast Workflow



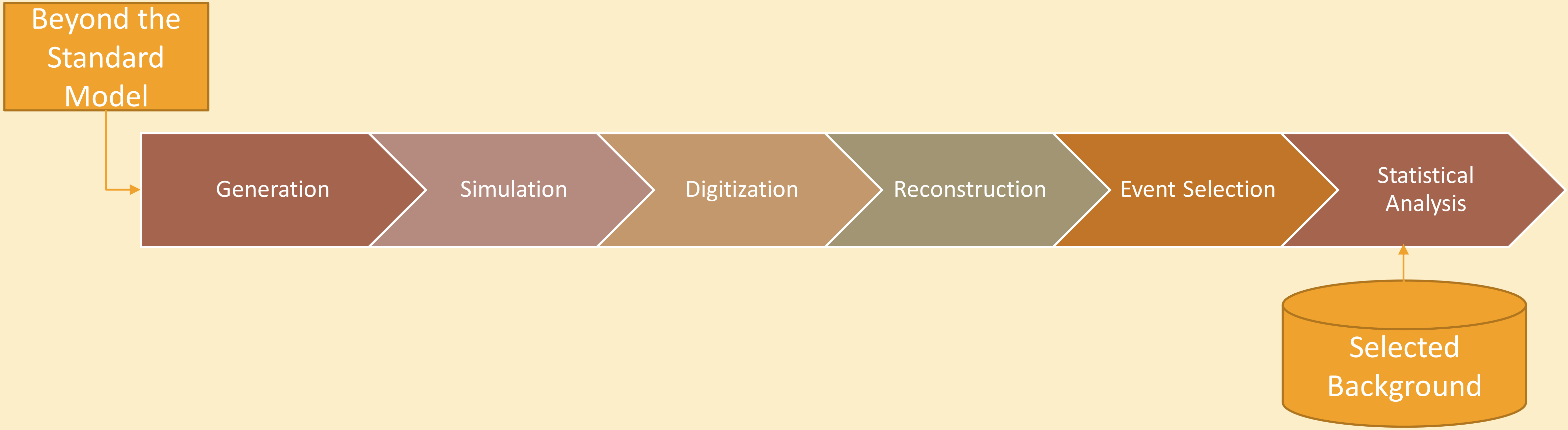
Recast Workflow



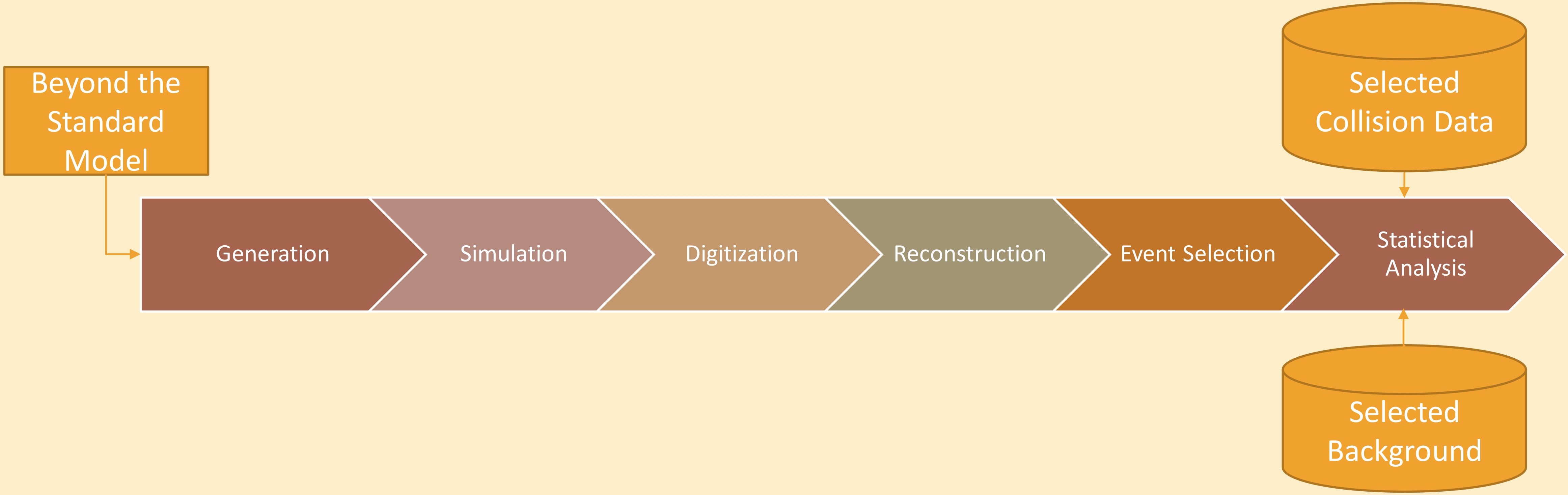
Recast Workflow



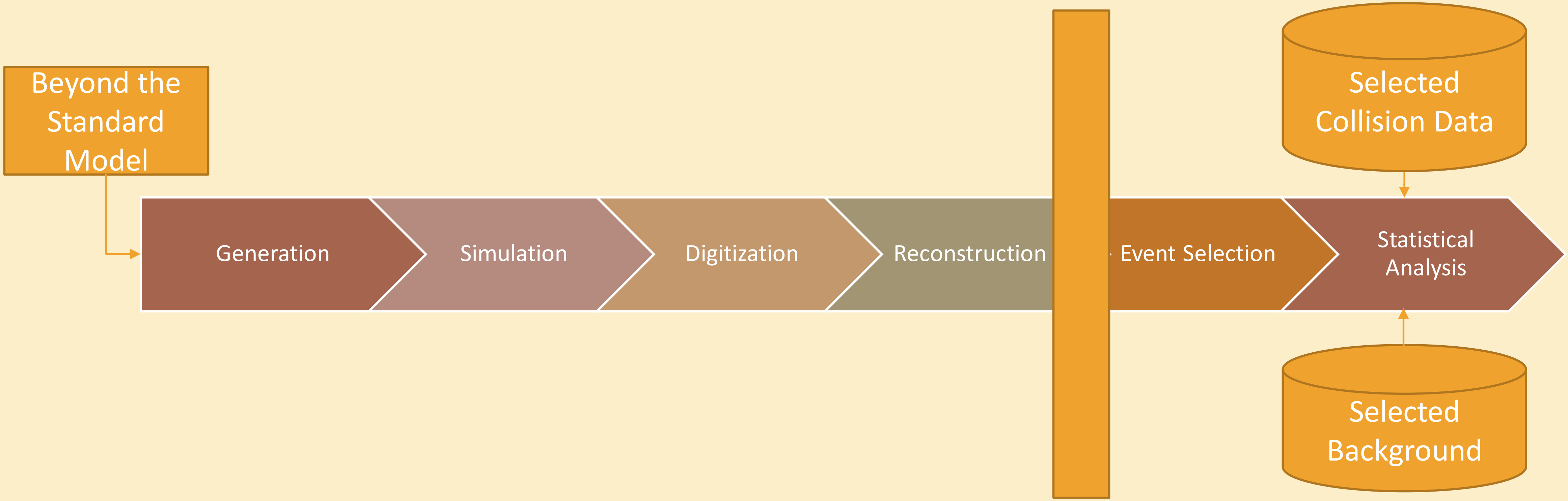
Recast Workflow



Recast Workflow



Recast Workflow



How It Works

To preserve an analysis, three components are necessary:

1. **Software** – what framework(s) does the analysis use and what are the dependencies?
2. **Commands** – what do I need to do to use the framework(s) for each stage of the analysis?
3. **Workflow** – how do I connect the analysis stages?

How It Works

To preserve an analysis, three components are necessary:

1. **Software** – what framework(s) does the analysis use and what are the dependencies?
2. **Commands** – what do I need to do to use the framework(s) for each stage of the analysis?
3. **Workflow** – how do I connect the analysis stages?

Step 1 uses industry-standard containerization: docker images.

How It Works

To preserve an analysis, three components are necessary:

1. **Software** – what framework(s) does the analysis use and what are the dependencies?
2. **Commands** – what do I need to do to use the framework(s) for each stage of the analysis?
3. **Workflow** – how do I connect the analysis stages?

Steps 2 and 3 are specified using a 'workflow language' (Yadage).

How It Works (extended)

More details available online at: https://recast-docs.web.cern.ch/

- Walkthrough on how to preserve and reuse an analysis.
- Examples.
- Integrated with GitLab CI.
- Analysis preservation was made a requirement for ATLAS.

Present

- TRUTH-LEVEL REINTERPRETATION

Full simulation is computationally expensive and difficult to use.

Determine which regions of phase space would be interesting for a full reinterpretation.

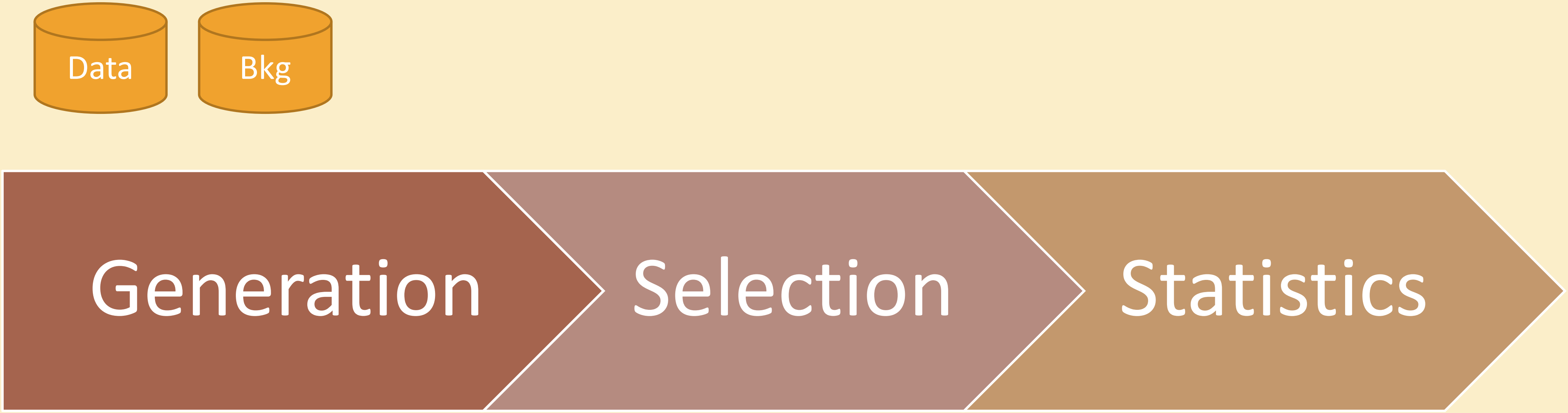
Estimate and explore other quantities, such as signal theory uncertainties.

Motivation

Catalogue

The following slides are lightly adapted from Lukas' FNAL talk:
https://indico.cern.ch/event/639314/contributions/2726367/attachments/1542484/2419628/ReinterpretationWorkshopFNAL.pdf

Catalogue



Many possibilities for each step...

Catalogue

Generation
- MG5+Py8
- Sherpa
- Powheg
- Herwig

Selection
- Rivet
- CheckMate
- MadAnalysis

Statistics
- Pyhf
- Roofit

Catalogue

Generation → **Selection** → **Statistics**

Build a catalogue!

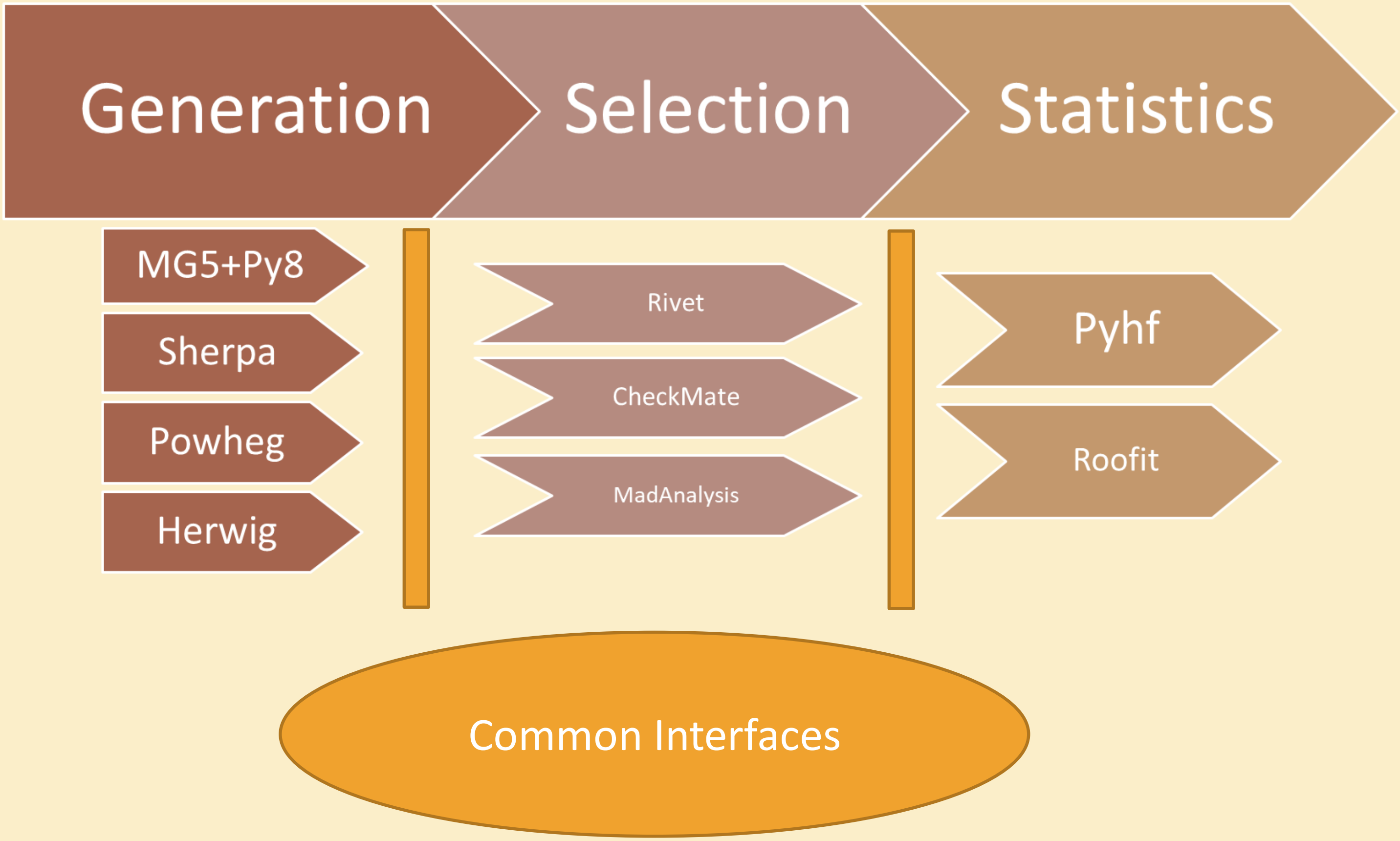
- Generation
  - MG5+I
  - Sherpa
  - Powheg
  - Herwig
- Selection
  - Rivet
  - CheckMate
  - MadAnalysis
- Statistics
  - Pyhf
  - Roofit

Idea

Use the **same workflow language** as with internal reinterpretations (Yadage).

Write **sub-workflows for each step and tool**.

Form a **complete workflow** by combining sub-workflows with **matching interfaces**.

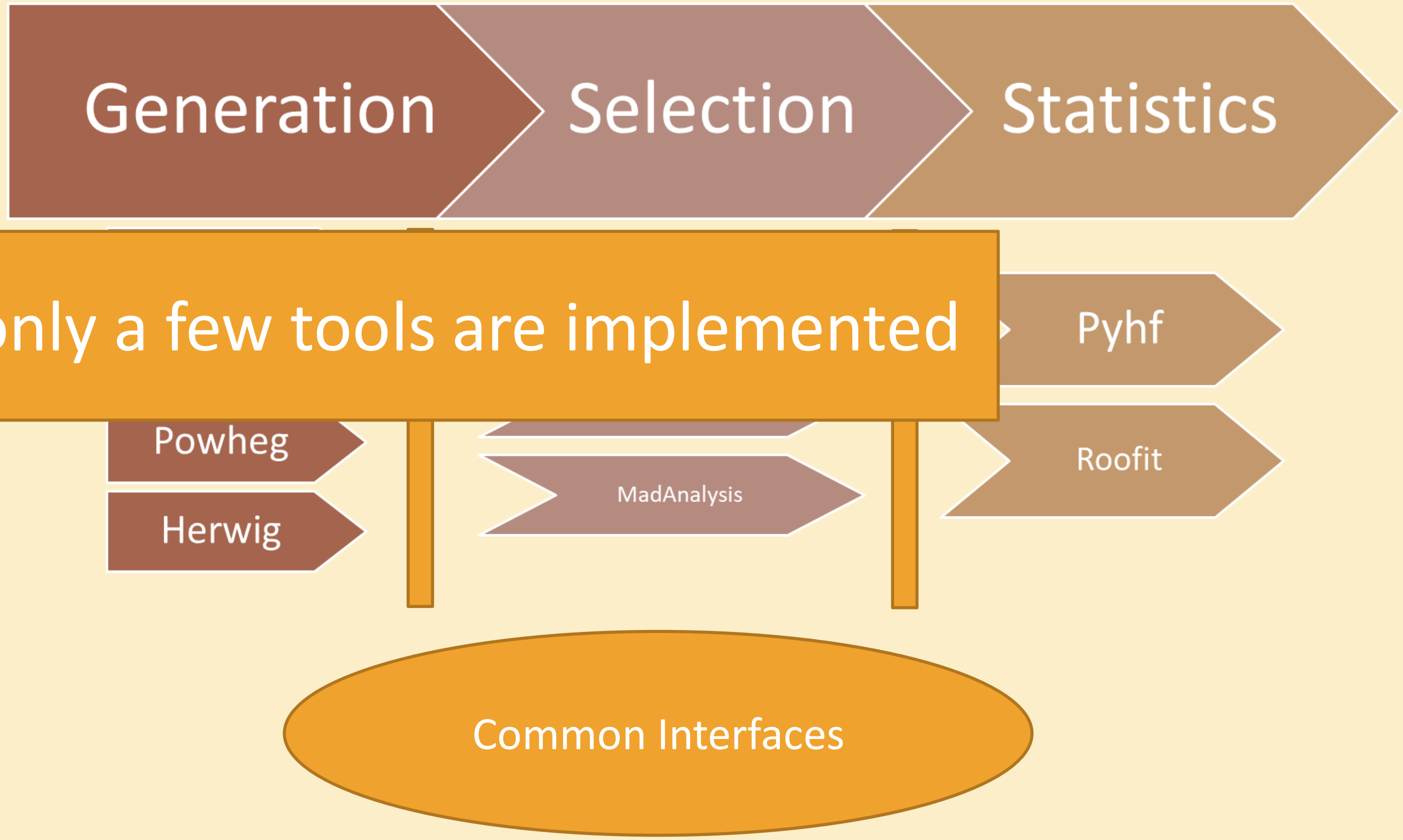


Idea

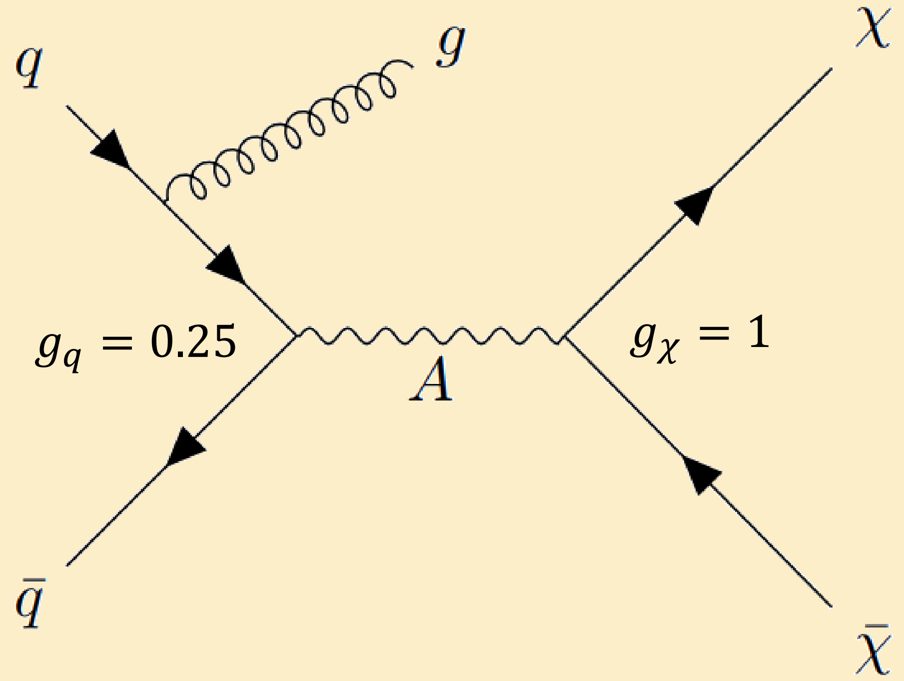
Use the **same workflow language** as with internal reinterpretations (Yadage).

Write **sub-workflows for each** **and tool**.

Form a **complete workflow** by combining sub-workflows with **matching interfaces**.



Example – Simplified DM Model



LO Feynman diagram from https://arxiv.org/1707.03263

Simplified DM model with WIMP pairs produced through **s-channel spin-1 mediator exchange**.

The model has **four free parameters**: the quark-mediator coupling $\boldsymbol{g_q}$, the DM-mediator coupling $\boldsymbol{g_\chi}$, the mass of the axial-vector mediator $\boldsymbol{m_A}$, and the mass of the DM particle $\boldsymbol{m_\chi}$.

We set $g_q = 0.25$ and $g_\chi = 1$ and vary m_A from 10 GeV to 2 TeV and m_χ from 1 GeV to 1 TeV.

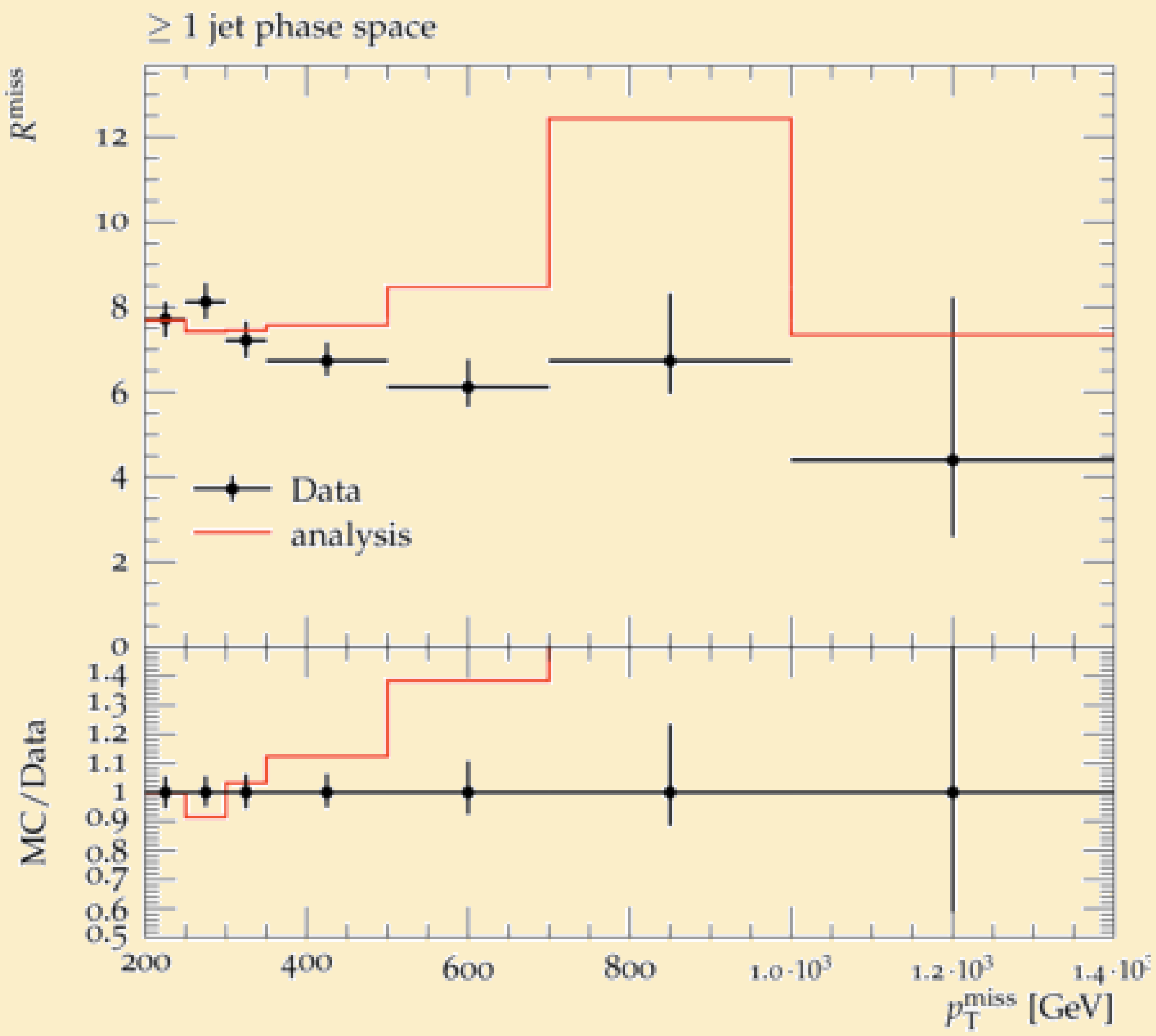
Example – Analysis

$$R^{\text{miss}} = \frac{\sigma_{\text{fid}}\left(p_{\text{T}}^{\text{miss}} + \text{jets}\right)}{\sigma_{\text{fid}}\left(\ell^{+}\ell^{-} + \text{jets}\right)}$$

We use a 2017 ATLAS analysis (ATLAS_2017_I1609448) that searched for DM models by using a **MET + jets** signature.

The analysis defines two regions: the ≥ 1 jet region and the VBF region.

The analysis examines R^{miss} as a function of: p_T^{miss} in the ≥ 1 jet region, p_T^{miss} in the VBF region, m_{jj} in the VBF region, and $\Delta\phi_{jj}$ in the VBF region.



Example – Results

Future

- REANA BACKEND
- SMART GRID SELECTION
- EXPANDED PUBLIC CATALOGUE WITH WEB INTERFACE

Reana Backend

www.reanahub.io

Cloud-backed computational workflow platform.

Supports yadage.

Supports submission of non-interactive batch workflows.

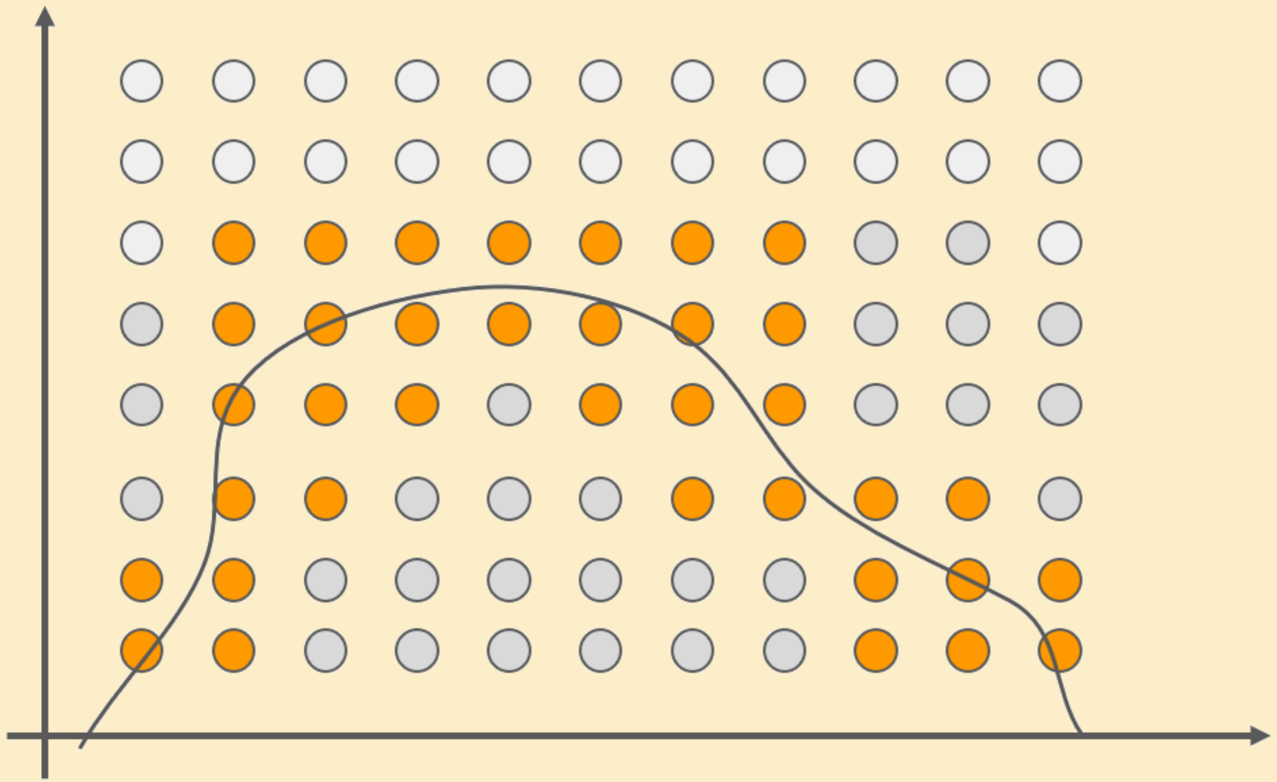
Smart Grid Selection

Problem:

- Typical grid procedure samples parameter space uniformly, but we **only care about the contour** – a lot of **wasted computation**.

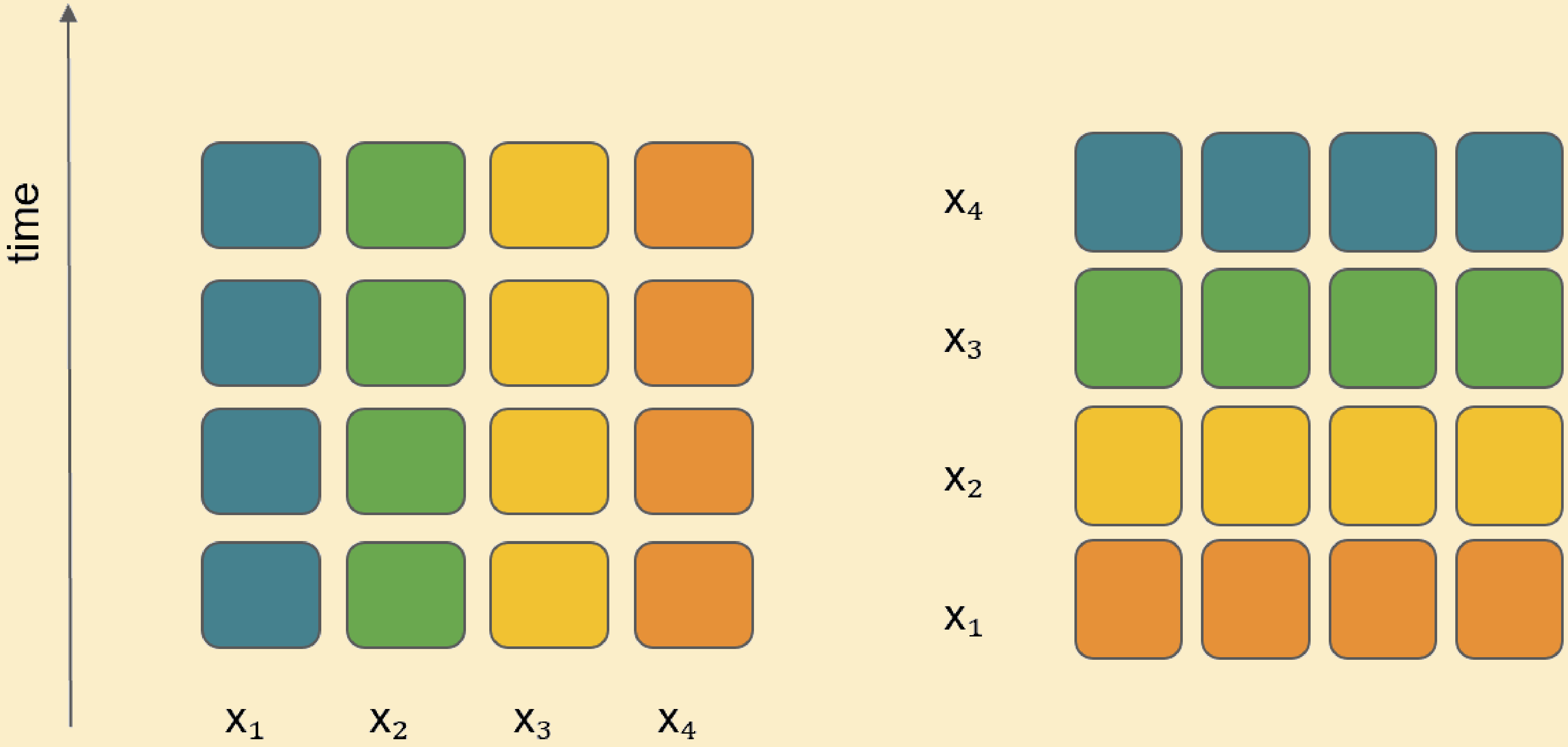
Ideas:

- Use **active learning** to sample near the contour.
- Use approximate evaluations to inform our prior.



https://indico.cern.ch/event/708041/contributions/3269754/

Wall-clock Time



Summary

RECAST is an **analysis preservation and reinterpretation tool** that has been used for **detector-specific analyses**.

RECAST is now being **extended to truth reinterpretations** using external tools. To illustrate, we implemented a RECAST workflow with madgraph+pythia, rivet and pyhf for an unfolded, parton-level 2017 ATLAS analysis of a simplified DM model.

Many exciting things to come! Contributions are welcome.

Backup

Reference Plot