

The LHCb Ntuple Wizard & Ntupling Service

22.10.2024

Christine Aidala¹, **Dillon Fitzgerald**¹, Kai Habermann², Ludwig Kramer², Adam Morris³, Sebastian Neubert², Piet Nogga², Eduardo Rodrigues³, Marco Donadoni³, Daan Rosendal^{3,4}, Tibor Simko³



¹University of Michigan, ²University of Bonn

³CERN, ⁴Windsheim University



Outline

- I. [Open Data](#)
- II. [A new approach: the LHCb Ntuple Wizard & Ntupling Service](#)
- III. [Request workflow](#)
- IV. [Additional resources](#)
- V. [Beta release](#)

Open Data

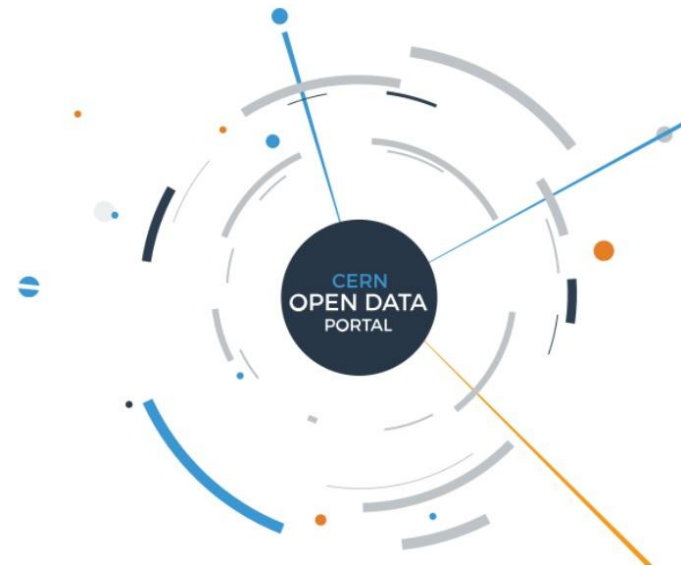
CMS Open Data User Story: [The Future of Particle Physics is Open](#) [2017-12-01 by Jesse Thaler (MIT)]
 (includes links to 2 published papers with open data!)



Open Data

The data collected at the LHC is very valuable! It should be made available to the public in accordance with the [CERN Open Data Policy](#) and [CERN Open Science Policy](#)

The CERN Open Data Portal (<https://opendata.cern.ch/>) provides a location for LHC experiments to host open data



CERN Open Data Policy

The [CERN Open Data Policy](#) outlines the commitment to make the data collected at the LHC publicly available at several levels of complexity, as established by the Data Preservation in High Energy Physics Collaboration ([DPHEP-2012-001](#))

- Level 1: Published results
 - This can include tables and figures but also preprocessed Ntuples or binned and unbinned fit likelihood functions.
- Level 2: Outreach and education
 - Usually in the form of highly preprocessed Ntuples.
- Level 3: Reconstructed data
 - These data have been preprocessed to derive physics objects, such as charged particle candidates, photons, or particle jets. Reconstructed data may or may not be corrected for detector effects, such as efficiency and resolution.

Target: Release research quality data mainly for theorists and phenomenologists
- Level 4: Raw data
 - the basic quantities recorded by the experimental instruments.

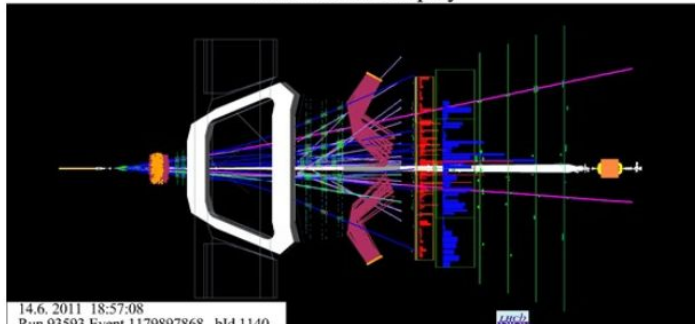

[News](#) › [News](#) › Topic: Knowledge sharing

LHCb releases first set of data to the public

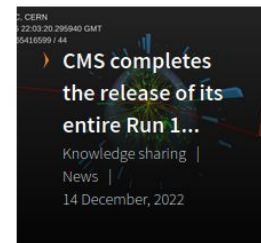
The LHCb collaboration has released data from Run 1 of the LHC to the public for the first time, allowing research to be conducted by anyone in the world

8 DECEMBER, 2022 | By LHCb collaboration

LHCb Event Display



Related Articles



<https://home.cern/news/news/knowledge-sharing/lhcb-releases-first-set-data-public>

December 2022



LHCb releases the entire Run I dataset

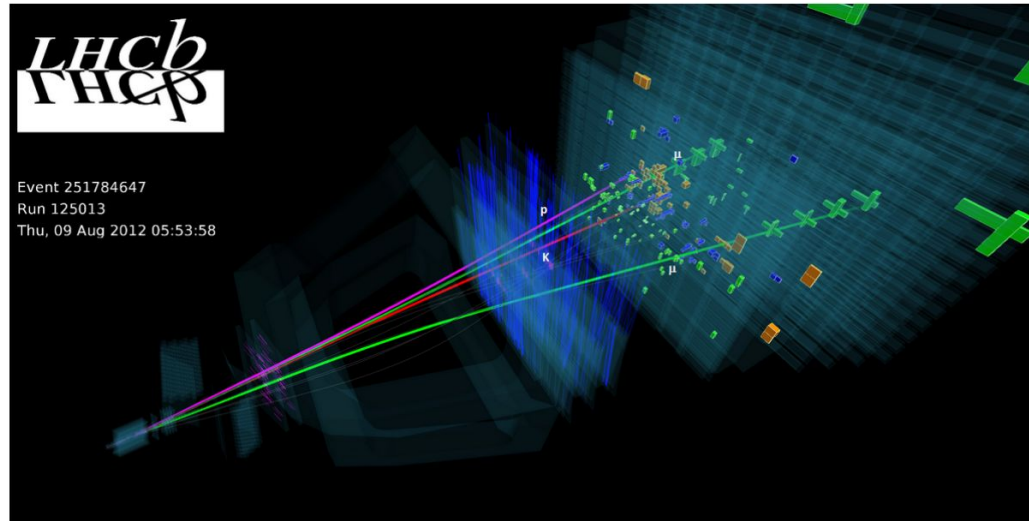
LHCb Collaboration

News

Date of publication: 2023-12-20

Today the LHCb collaboration completes the release of the data collected throughout the Run I of the Large Hadron Collider at CERN. The sample made available amounts to approximately 800 terabytes (TB) of data. These data, collected by the LHCb experiment in 2011 and 2012, contains information obtained from proton-proton collisions. The format made available provides pre-filtered data, suitable for a wide range of physics studies. The image below displays an event recorded during 2012.

December 2023



<https://opendata.cern.ch/docs/lhcb-releases-entire-run1-dataset>

LHCb Open Data

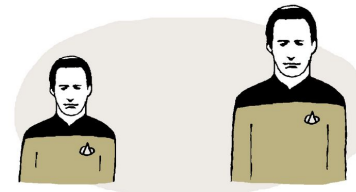
LHCb released (end of 2023) all of the Run 1 data (~800 TB) on the CERN Open Data Portal:

<https://opendata.cern.ch/search?page=1&size=20&experiment=LHCb>

Releases for Run 2 and beyond will be challenging due to the volume of data...

	ALICE	ATLAS	CMS	LHCb
Run-2	2 PB	0.5 PB	2 PB	10 PB (including Run-1)
Run-3	4 PB	1 PB	4 PB	45 PB
Total	6 PB	1.5 PB	6 PB	55 PB

DATA BIG DATA



Dataedo /cartoon

Probr@Dataedo

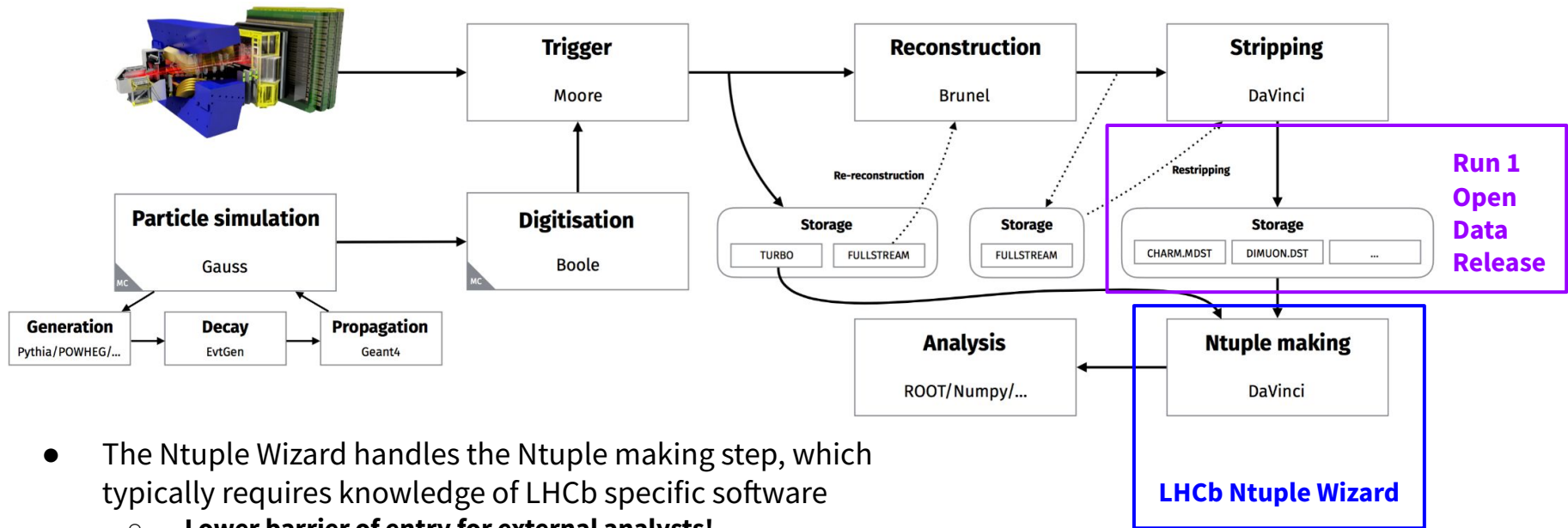
This is not scalable! This prompted the development of a new system...

The LHCb Ntuple Wizard

Previous presentations

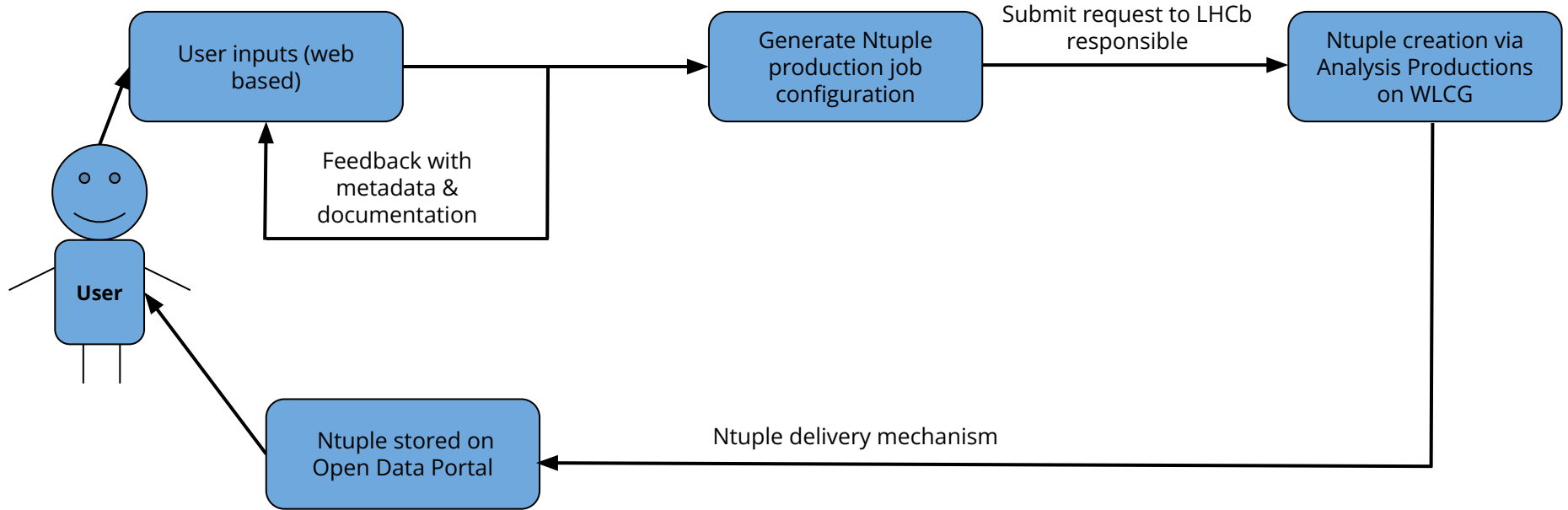
- [CHEP 2023](#): Presented idea and mature prototype (screenshots) publicly
- [4th DPHEP Collaboration Workshop](#): Showed fully functional service developed in collaboration with CERN IT

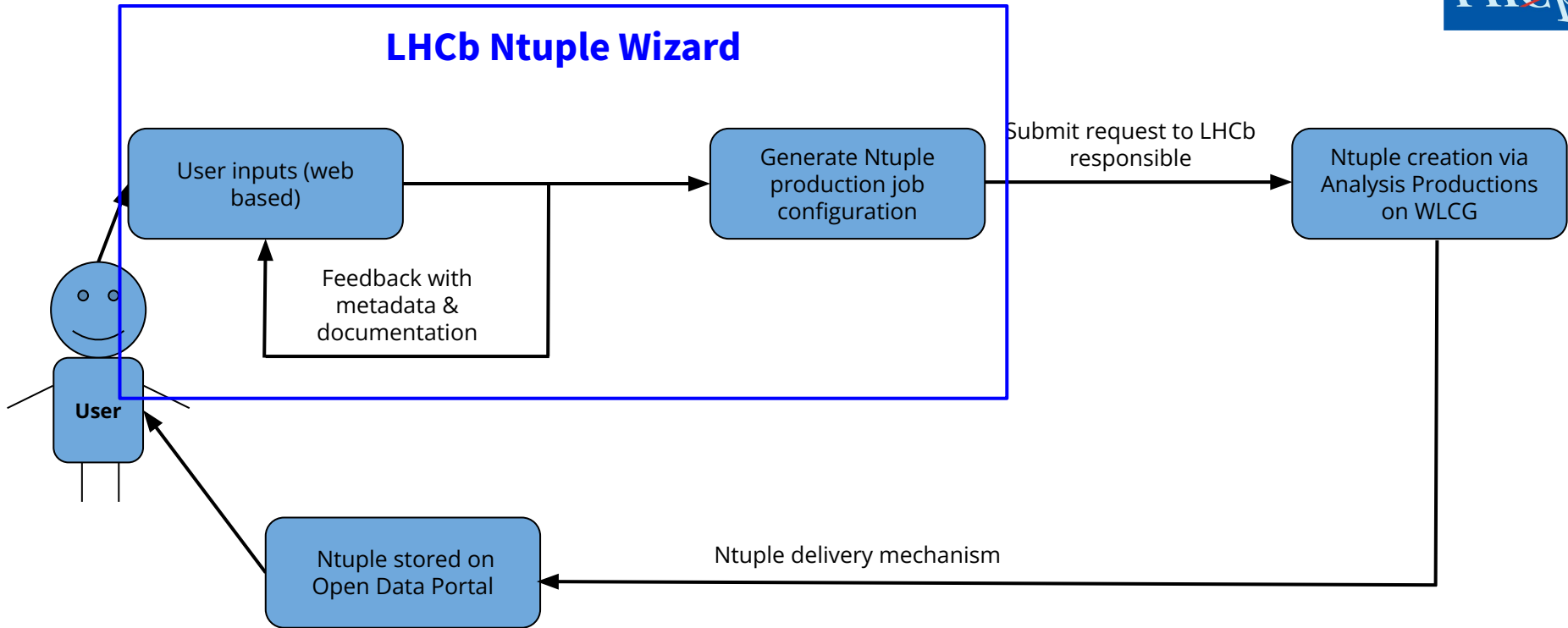
LHCb Run 1 and 2 Data Flow

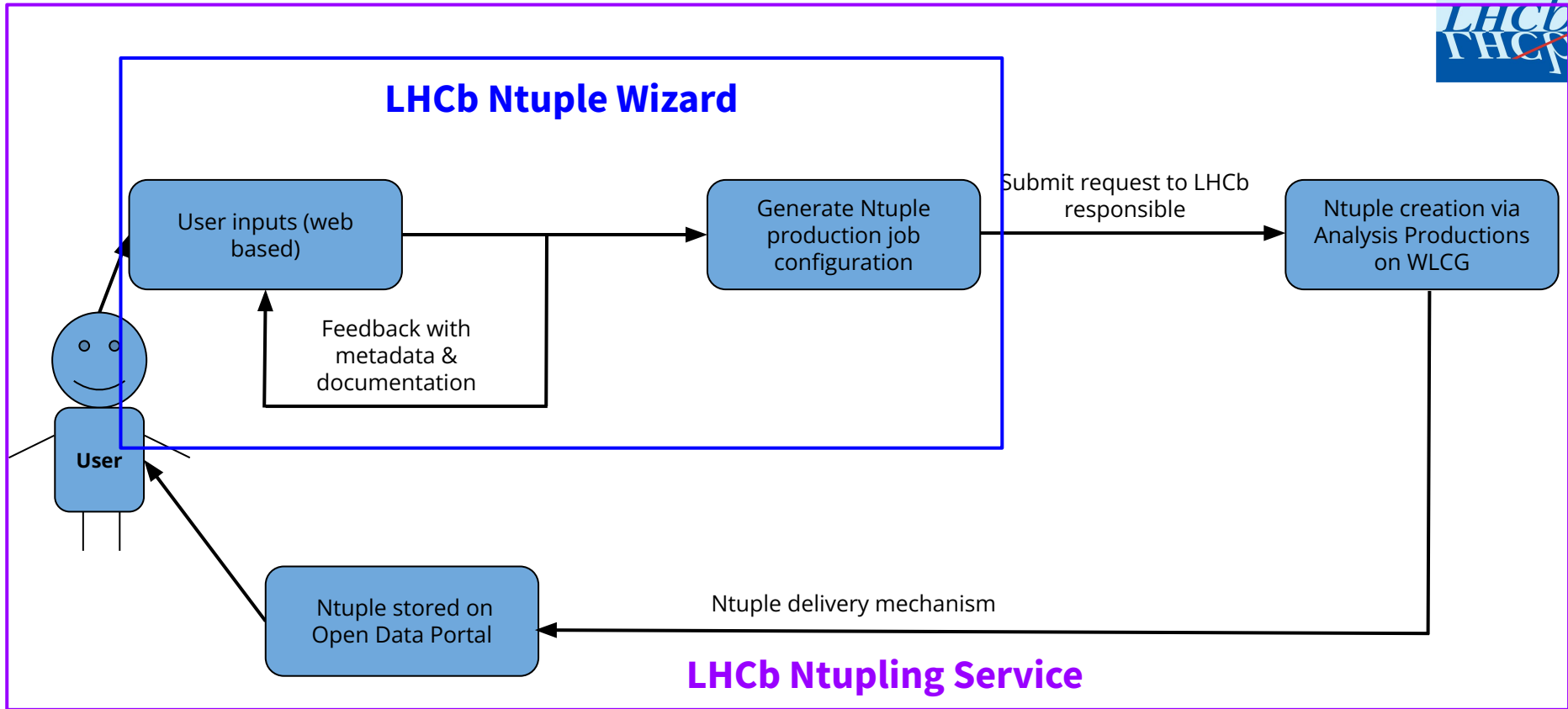


- The Ntuple Wizard handles the Ntuple making step, which typically requires knowledge of LHCb specific software
 - **Lower barrier of entry for external analysts!**
 - **Convenient means of parsing documentation and available data for internal analysts!**

The basic idea...



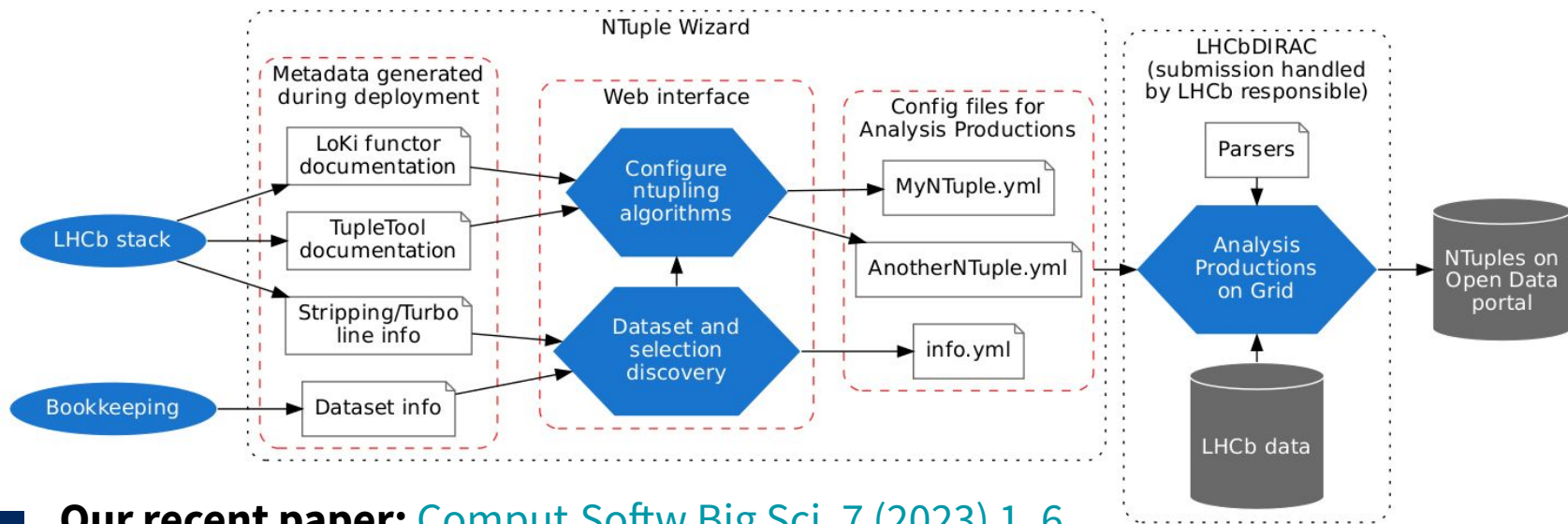




LHCb Ntupling Service

The LHCb Ntuple Wizard

- Intuitive web interface (wizard) guides the user through formulating a query, key features include:
 - Dataset discovery/selection
 - Ntuple configuration
- Input (metadata/documentation) and output (configuration files from user) have secure design features

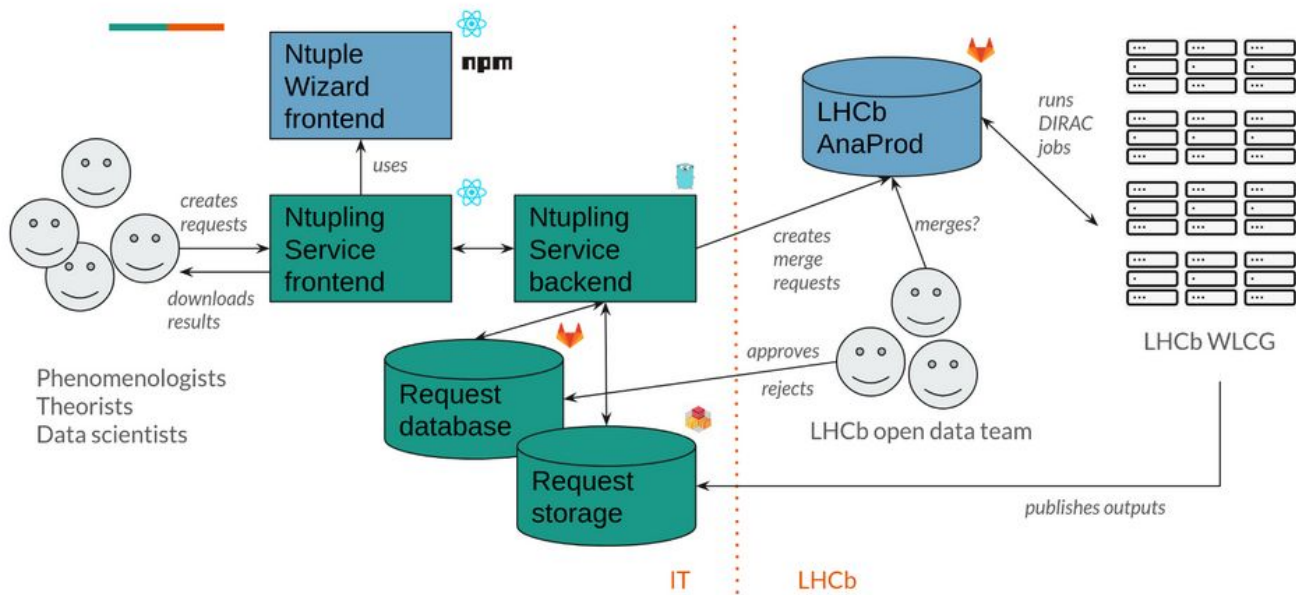


Our recent paper: [Comput.Softw.Big Sci. 7 \(2023\) 1, 6](#)

The LHCb Ntupling Service

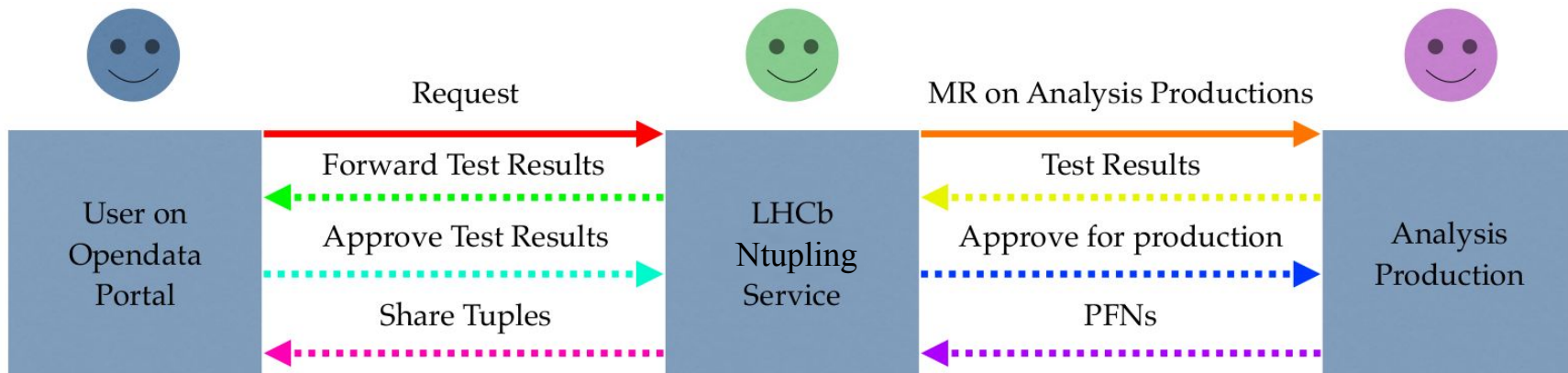
The LHCb Ntuple Wizard has been integrated with the CERN Open Data Portal and the LHCb Analysis Productions batch processing system thanks to the very fruitful and ongoing collaboration with the Open Data team from CERN IT. The combined application is called the LHCb Ntupling Service.

- See [talk at CHEP 2024](#) for more details!



Request Workflow

The LHCb Ntupling Service



Request

- A request can be made on the LHCb Ntupling Service, to be accessible from the [CERN Open Data Portal](#) for the public release
- Create custom Ntuples!
 - Choose from available decays, and data streams
 - Configure your own Ntuple
 - Add and customize [TupleTools](#) to specify quantities written to output

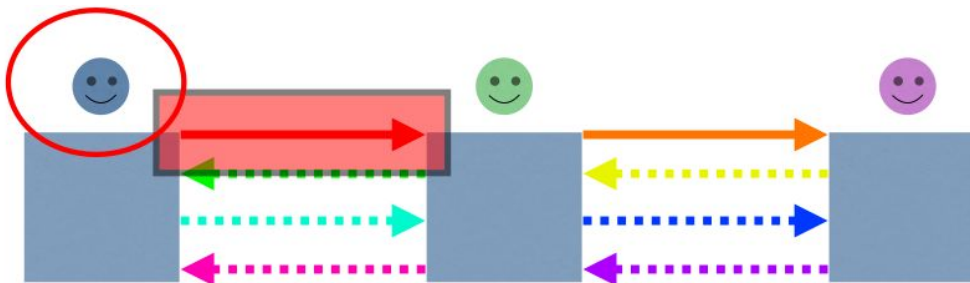
LHCb Open Data Ntupling Service

Logged in as dillon.fitzgerald@cern.ch [Edit profile](#)

Your requests

Create new request

Create new request



Web Interface: Dataset Discovery

***Key feature:** Find available dataset by first choosing physics object of interest!

Decay search

Head (exactly): B^+ Contains (all of): D^0 Show only selected:

Tags (none of): undefined-unstable x charge-violating x lepton-flavour-violating x Stripping line

<input type="checkbox"/>	$B^+ \rightarrow (\bar{D}^0 \rightarrow K^+\pi^-(\pi^0 \rightarrow \gamma\gamma))\pi^+$	2 Stripping lines
<input type="checkbox"/>	$B^+ \rightarrow (\bar{D}^0 \rightarrow K^+\pi^-\pi^-\pi^+)\pi^+$	3 Stripping lines
<input checked="" type="checkbox"/>	$B^+ \rightarrow (\bar{D}^0 \rightarrow K^+\pi^-)\pi^+$	6 Stripping lines
<input type="checkbox"/>	$B^+ \rightarrow (\bar{D}^0 \rightarrow K^-K^+(\pi^0 \rightarrow \gamma\gamma))\pi^+$	2 Stripping lines
<input type="checkbox"/>	$B^+ \rightarrow (\bar{D}^0 \rightarrow K^-K^+K^+\pi^-)\pi^+$	2 Stripping lines
<input type="checkbox"/>	$B^+ \rightarrow (\bar{D}^0 \rightarrow K^-K^+\pi^-\pi^+)\pi^+$	3 Stripping lines

Lists physics objects available in the LHCb database (primarily decays)

- List filtering options include:
 - Decay head (top level decaying particle)
 - Particles in the decay
 - Tags related to specific physics (include or exclude)
 - “Stripping line” name

- Can make multiple selections from the list

Web Interface: Dataset Discovery

Selection of a physics object exposes the corresponding available datasets for the user to choose from

Stripping line selection

- Specifies algorithms applied to identify candidates of the selected physics object

Dataset selection

- Specifies the dataset to run over -- multiple selections can be made

Production configuration

Btree

$B^+ \rightarrow (\bar{D}^0 \rightarrow K^+ \pi^-) \pi^+$

Title MyAnalysis

Email name@example.com

Done Clear

StrippingB2D0PID2HHBeauty2CharmL...

S21r1 S21rp2 S21 S21r0p2 S24r2

S28r2 S29r2 S34

BHADRONCOMPLETEEVENT.DST

Data 2016 MagDown S28r2

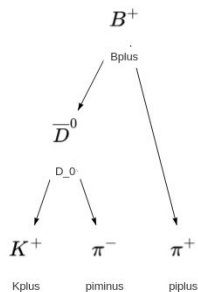
At this stage, the user can initiate configuration of the Ntuple corresponding to the selected physics object(s)



Web Interface: Ntuple Configuration

Configure $B^+ \rightarrow (\bar{D}^0 \rightarrow K^+\pi^-\pi^+)$

BQ



Select by category

Hadron Meson X+ X0 X- Up Beauty Charm Strange Down LongLived Stable StableCharged Scalar

Current selection: $B^+ \rightarrow (\bar{D}^0 \rightarrow K^+\pi^-\pi^+)$

5 TupleTools		+
TupleToolANPID	<input type="checkbox"/>	<input type="checkbox"/>
TupleToolEventInfo	<input type="checkbox"/>	<input type="checkbox"/>
TupleToolGeometry	<input type="checkbox"/>	<input type="checkbox"/>
TupleToolKinematic	<input type="checkbox"/>	<input type="checkbox"/>
TupleToolPid	<input type="checkbox"/>	<input type="checkbox"/>

Ntuple configuration via an interactive node tree

- Particles in decay rendered as nodes in tree
- Each node can be configured independently, or in various groupings
 - Labels provided to select nodes by similar categories
- Node configuration proceeds by adding, removing, or configuring **TupleTools**, which save various physics quantities to the Ntuple
 - Can be performed on entire tree, single node, or selection of nodes
- The entire node tree includes 5 standard TupleTools for LHCb analysis by default

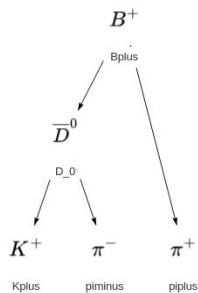
Web Interface: Ntuple Configuration

Configure $B^+ \rightarrow (\bar{D}^0 \rightarrow K^+\pi^-)\pi^+$

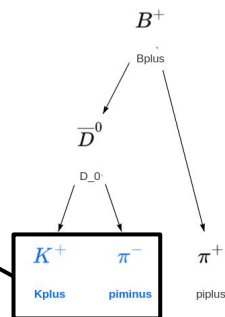
🔍

Configure $B^+ \rightarrow (\bar{D}^0 \rightarrow K^+\pi^-)\pi^+$

🔍



Selected nodes



Select by category

- Hadron
- Meson
- X+
- X0
- X-
- Up
- Beauty
- Charm
- Strange
- Down
- LongLived
- Stable
- StableCharged
- Scalar

Current selection: $B^+ \rightarrow (\bar{D}^0 \rightarrow K^+\pi^-)\pi^+$

5 TupleTools		+
TupleToolANPID	<input type="checkbox"/>	<input type="checkbox"/>
TupleToolEventInfo	<input type="checkbox"/>	<input type="checkbox"/>
TupleToolGeometry	<input type="checkbox"/>	<input type="checkbox"/>
TupleToolKinematic	<input type="checkbox"/>	<input type="checkbox"/>
TupleToolPid	<input type="checkbox"/>	<input type="checkbox"/>

Select by category

- Hadron
- Meson
- Up
- LongLived
- Stable
- StableCharged
- Scalar

Current selection: $B^+ \rightarrow (\bar{D}^0 \rightarrow K^+\pi^-)\pi^+$

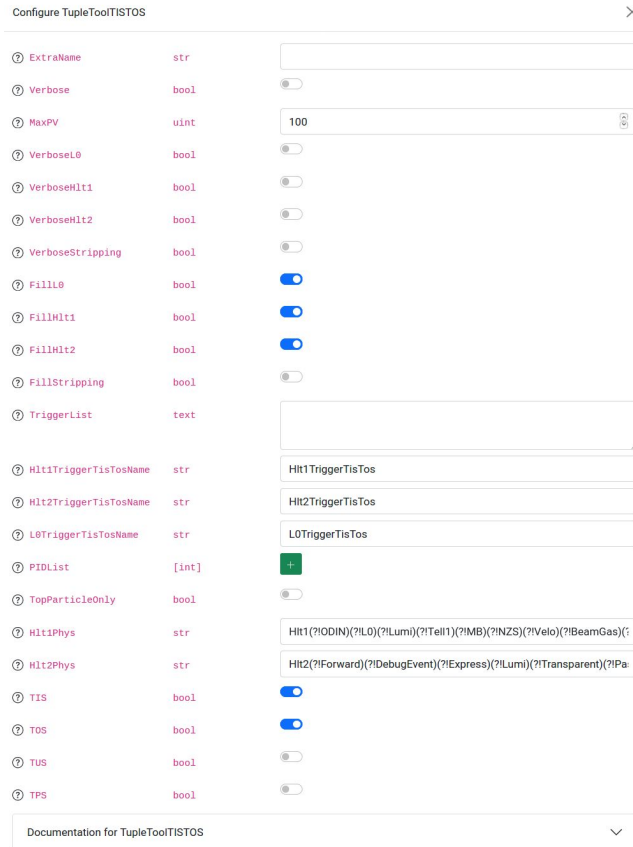
1 TupleTool		+
TupleToolTISTOS	<input type="checkbox"/>	<input type="checkbox"/>

Launch TupleTool configuration

Web Interface: Tuple Tool Configuration Example

Example of TupleTool configuration interface for TupleToolTISTOS (**T**riple **I**ndependent of **S**ignal/**T**riple **O**n **S**ignal)

- Configurable names, data types, and user input fields are included
- Mouseover tooltips and links to documentation are included for guidance
 - This includes LHCb Doxygen documentation
- Each TupleTool has specific configurables
 - For many tools, the standard configuration is perfectly fine
 - Only certain tools (e.g. related to the trigger) need specific configurations, to be specified in the documentation



Request

- Submitting a request will open an issue on GitLab
 - Users will be regularly informed when stages of the production process change

LHCb Open Data Ntupling Service

Logged in as dillon.fitzgerald@cern.ch [Edit profile](#)

[Your requests](#) [Create new request](#)

Create new request

✎ 🗑️

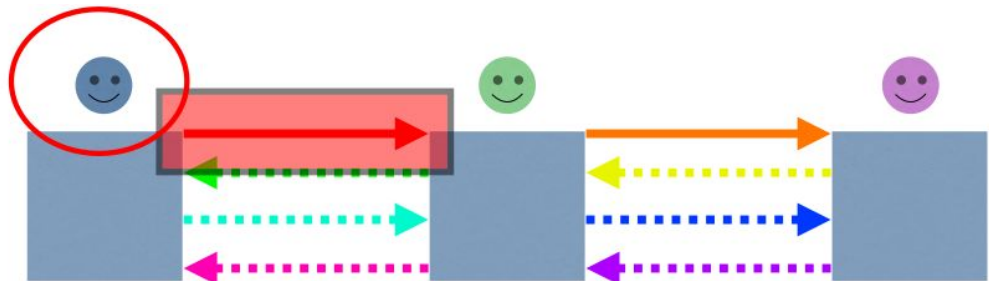
$B^+ \rightarrow (\bar{D}^0 \rightarrow K^+ \pi^-) \pi^+$

✕

✕ 🗑️

✓

Submit Clear



Approval for Test Production by LHCb Open Data Team

- LHCb reviews and approves requests manually
- Approval triggers a draft merge request (MR) in [Analysis Productions](#)
 - This launches a test production via the Analysis Productions pipeline
- Merge request left on draft until user confirms results of test production

LHCb Open Data Ntupling Service

Logged in as dillon.fitzgerald@cern.ch [Edit profile](#)

[Your requests](#)

[Create new request](#)

Request ID: 71

Request date	2024-02-07 14:07:22
Production name	B_Dpi
Reason	I am testing the system and getting some example screenshots!
Status	Awaiting LHCb review

Comments

No comments yet.

Type your comment here. Markdown is supported.

[Add comment](#)

Ntupling Service Request | B_Dpi | dillon.fitzgerald@cern.ch

[Open](#) Issue created 5 minutes ago by CERN Open Data DEV

Requester

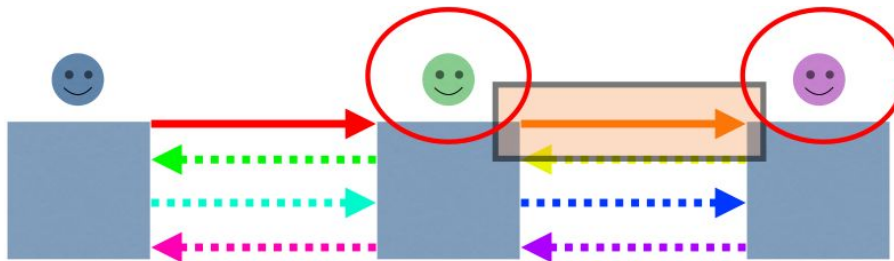
- Name: Dillon Fitzgerald
- Email: dillon.fitzgerald@cern.ch
- Field of research: Physics (experimental)
- Position: Graduate student
- Experiment: LHCb
- Remarks: Testing the system

Request Details

- Production name: B_Dpi
- Reason for request: I am testing the system and getting some example screenshots!

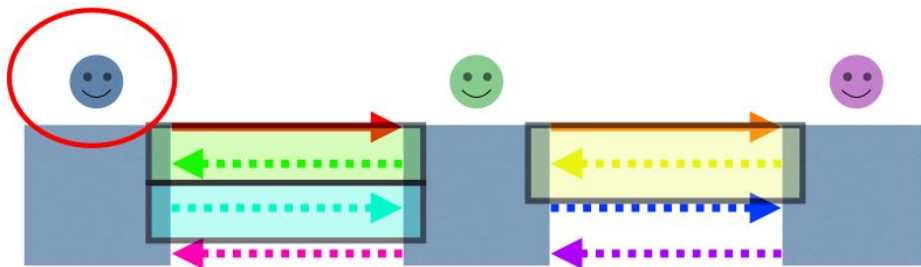
Instructions

- See production configuration files for this request here: [IG4 \(diffs\)](#)
- React to a comment with the 📧 (mailbox) emoji to publish a comment to the requester



Return Test Results to User for Verification

- From the test production output, a markdown file is provided to the user with information about the production
- The user can verify the test results at this stage
 - Are all expected branches present in the Ntuple?
 - Is the estimated output size within a feasible range?
- After user approval, the merge request on Analysis Productions is marked as ready



test-production.md

Welcome to the LHCb Open Data service!

Congratulations! Your request has been approved by the LHCb Open Data Team and submitted to the LHCb Analysis Productions service!

The internal testing system has been successfully completed, creating a small test sample.

Please take a moment to assess the test statistic below and consider the following:

1. Is the estimated output size within a feasible range for you?
2. Do all expected variables exist in the ROOT file?

If the test statistic fulfills the above criteria and you are satisfied with the results, please confirm your request here.

Next Steps:

- Once you confirm you are satisfied with the test result, the LHCb Open Data Team will send the request to full production
- You will be notified once the production is finished. The ROOT files are then transferred to the Open Data Portal.
- You can then access the data freely. Have fun on your Analysis!

Thank you for your patience. If you have further questions or concerns, please feel free to reach out by leaving a comment on your request, by sending an email to opendata-support@cern.ch or by opening a ticket on the [Opendata Data Forum](#).

B_3Ps1k will process 135.6TB of data and create about 17.8GB across 6 samples. Please note that the actual size of the output files may vary from the estimate.

See Branches in Production Tuples

```

▼ jobs
  ▼ Btree/DecayTree:1 (394 Branches)
    Bplus_ENDVERTEX_X
    Bplus_ENDVERTEX_Y
    Bplus_ENDVERTEX_Z
    Bplus_ENDVERTEX_XERR
    Bplus_ENDVERTEX_YERR
    Bplus_ENDVERTEX_ZERR
    Bplus_ENDVERTEX_CHI2
    Bplus_ENDVERTEX_NDOF
    Bplus_ENDVERTEX_COV_
    Bplus_OWNPV_X
    Bplus_OWNPV_Y
    Bplus_OWNPV_Z
    Bplus_OWNPV_XERR
    Bplus_OWNPV_YERR
    Bplus_OWNPV_ZERR
    Bplus_OWNPV_CHI2
    Bplus_OWNPV_NDOF
    Bplus_OWNPV_COV_
    Bplus_IP_OWNPV
    Bplus_IPCHIZ_OWNPV
    Bplus_FD_OWNPV
  
```

Download Close

Approval for Full Production

LHCb approves the final production and merges to Analysis Productions

- This triggers the Analysis Productions system to process the full request over all selected datasets

LHCb Analysis Productions Log out Piet Nogga (piet.nogga@cern.ch)

Home
 Productions
 Pipelines
 Settings
 Documentation

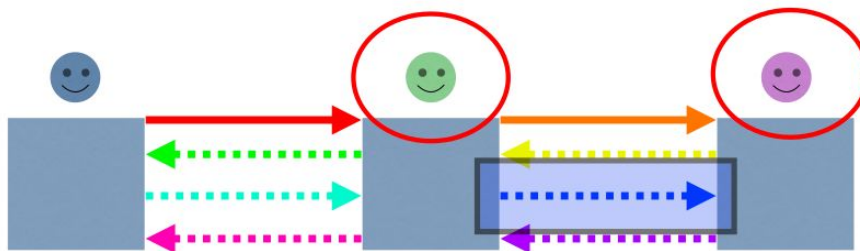
Tree display

This section displays the samples split by tags and is the recommended way of requesting datasets. Clicking on one of the boxes will filter the list of samples shown below. See TODO for more information.

Grouped tags: config datatype eventtype polarity

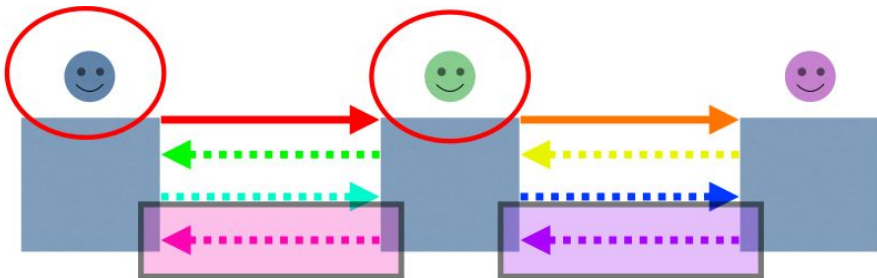
Drag to sort: config datatype eventtype polarity

wizdexamplexcreep 12			
2011 2	2015 2	2017 2	2018 2
90000000 2 magdown 1	90000000 2 magdown 1	90000000 2 magdown 1	90000000 2 magdown 1
magup 1	magup 1		
2012 2	2016 2	magup 1	magup 1
90000000 2 magdown 1	90000000 2 magdown 1		
magup 1	magup 1		



Deliver Ntuples to CERN Open Data Portal

- Upon completion, the output Ntuples are copied to *eospublic* and hosted on the CERN Open Data Portal, accessible by viewing your request in the Ntupling Service interface
- The produced Ntuples can now be downloaded by the user!



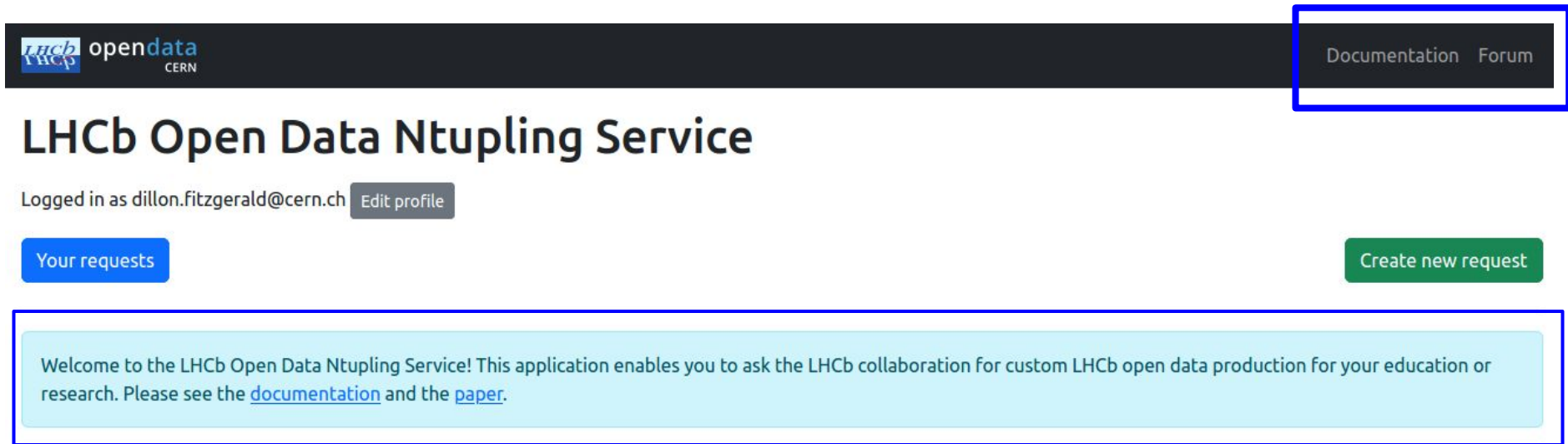
The real production for your request has been completed! You can download the results below.

Real production results

Data	Size		
00245692_00000001_1.dvntuple.root	30.171 MiB		
00245692_00000002_1.dvntuple.root	143.751 MiB		
00245692_00000003_1.dvntuple.root	243.268 MiB		
00245692_00000004_1.dvntuple.root	336.708 MiB		
00245692_00000005_1.dvntuple.root	423.539 MiB		
00245692_00000006_1.dvntuple.root	487.739 MiB		
00245692_00000007_1.dvntuple.root	552.622 MiB		
00245692_00000008_1.dvntuple.root	621.955 MiB		
00245692_00000009_1.dvntuple.root	694.136 MiB		
00245692_00000010_1.dvntuple.root	791.876 MiB		
00245692_00000011_1.dvntuple.root	177.133 MiB		
00245692_00000012_1.dvntuple.root	567.146 MiB		

Additional Resources

You can access additional resources such as the [LHCb open data guide](#) and the [CERN open data forum](#) directly from the LHCb Ntupling Service landing page



The screenshot shows the LHCb Open Data Ntupling Service landing page. At the top left is the LHCb logo and 'open data CERN'. At the top right, a dark navigation bar contains 'Documentation' and 'Forum', which are highlighted with a blue box. Below the navigation bar is the main heading 'LHCb Open Data Ntupling Service'. Underneath, it says 'Logged in as dillon.fitzgerald@cern.ch' with an 'Edit profile' button. There are two buttons: 'Your requests' (blue) and 'Create new request' (green). A light blue box contains a welcome message: 'Welcome to the LHCb Open Data Ntupling Service! This application enables you to ask the LHCb collaboration for custom LHCb open data production for your education or research. Please see the [documentation](#) and the [paper](#).'

LHCb Open Data Guide

Home

Ntupling Service

[About](#)

Dataset Discovery

[Selecting Decays](#)[Choosing a Stripping Line](#)[Choosing a Dataset to Process](#)

Ntuple Configuration

[Ntuple Creation](#)[Ntuple Configuration](#)[Tuple Tools](#)[Submitting Requests](#)

Monitoring and Retrieving Output

[Monitoring Output](#)[Retrieving Output](#)

The LHCb Ntupling Service

Welcome to the LHCb Ntupling Service Documentation!

The LHCb Ntupling Service is a fully functional system integrating the LHCb Ntuple Wizard application ([Comput.Softw.Big Sci. 7 \(2023\) 1, 6](#)) with the [CERN Open Data Portal](#) and the Worldwide LHC Computing Grid (WLCG), allowing users to interface with LHCb open data. Users can generate a query by following steps laid out in this tutorial, which will be processed by the LHCb Open Data team, and if accepted, ultimately will result in Ntuples delivered to the [CERN Open Data Portal](#). The service is hosted on the Open Data Portal with the landing page shown in the screenshot below. The procedure to formulate a query of the LHCb data can be initiated by clicking the green "Create new request" button.






LHCb Open Data Ntupling Service

Logged in as [dillon.fitzgerald@cern.ch](#) [Edit profile](#)[Your requests](#)[Create new request](#)

It is recommended to use the Ntuple Wizard in a computer browser rather than on a mobile phone, as there is important contextual information displayed in mouseover tooltips that may not be accessible on mobile devices.

LHCb Open Data Guide

Home

- Ntupling Service 
- About 
- Dataset Discovery 
- Selecting Decays
- Choosing a Stripping Line
- Choosing a Dataset to Process
- Ntuple Configuration 
- Ntuple Creation
- Ntuple Configuration
- Tuple Tools
- Submitting Requests
- Monitoring and Retrieving Output 
- Monitoring Output
- Retrieving Output

Many pages with detailed steps on various stages of the process!

The LHCb Ntupling Service

Welcome to the LHCb Ntupling Service Documentation!

The LHCb Ntupling Service is a fully functional system integrating the LHCb Ntuple Wizard application ([Comput.Softw.Big Sci. 7 \(2023\) 1, 6](#)) with the [CERN Open Data Portal](#) and the Worldwide LHC Computing Grid (WLCG), allowing users to interface with LHCb open data. Users can generate a query by following steps laid out in this tutorial, which will be processed by the LHCb Open Data team, and if accepted, ultimately will result in Ntuples delivered to the [CERN Open Data Portal](#). The service is hosted on the Open Data Portal with the landing page shown in the screenshot below. The procedure to formulate a query of the LHCb data can be initiated by clicking the green "Create new request" button.

LHCb Open Data Ntupling Service

Logged in as [dillon.fitzgerald@cern.ch](#) [Edit profile](#)

[Your requests](#)[Create new request](#)

It is recommended to use the Ntuple Wizard in a computer browser rather than on a mobile phone, as there is important contextual information displayed in mouseover tooltips that may not be accessible on mobile devices.

CERN Open Data Forum

<https://opendata-forum.cern.ch>

opendata		Sign Up	Log In	Q	☰
Using Scikit-Learn to classify signal using Secondary vertex characteristics ■ CMS		3	18	8 Aug	
Library conflict in Simplified Run 2 Analysis lessons from 2023 Workshop ■ Software tools		0	14	23 Jul	
Large data set analysis using only Python ■ Containers		4	30	22 Jul	
Alternatives to Jupyter Notebooks in Python Docker (Windows) ■ Containers		8	25	19 Jul	
Regarding the slimmed secondary vertex from reco::Vertex collection ■ CMS		2	21	9 Jul	
Extracting Dxy and NDf for secondary vertices from the class "reco::VertexCompositePtrCandidate"		4	42	1 Jul	
VNC Server connection error in CMSSW docker container ■ Containers		5	100	20 Jun	

categories ▾ Latest Hot Categories

Search...

- CMS × 22
Ask any question about CMS Open Data
- Site Feedback × 5
Discussion about this site, its organization, how it works, and how we can improve it.
- News × 2
General news about CERN Open Data
- ATLAS × 1
Ask any question about ATLAS Open Data
- LHCb
Ask any question about LHCb Open Data

CERN Open Data Forum - LHCb Category

<https://opendata-forum.cern.ch/c/lhcb/13>



Ask any question about LHCb Open Data

LHCb Latest Hot

Topic	Replies	Views	Activity
About the LHCb category			
Ask any question about LHCb Open Data	TS	0	377 Nov 2022

There are no more LHCb topics.

subcategories ▾ Latest Hot

Search... 🔍

no subcategories

LHCb Ntupling Service

Ask any questions about the LHCb Ntupling Service.

CERN Open Data Forum - Ntupling Service Subcategory

<https://opendata-forum.cern.ch/c/lhcb/lhcb-ntupling-service/15>

■ LHCb ▶

■ LHCb Ntupling Service ▶

Latest

Hot

Topic		Replies	Views	Activity
 About the LHCb Ntupling Service category Ask any questions about the LHCb Ntupling Service.	TS	0	3	5h

There are no more LHCb Ntupling Service topics.

Access to the Beta Release!

A link was sent to your email to access the LHCb Ntuple Wizard and overarching Ntupling Service

Try things out for yourself!

Feedback is greatly appreciated!

- For questions specific to your request
 - You can communicate with us via comments on the LHCb Ntupling Service
- For more general inquiries or suggestions about the LHCb Ntupling Service
 - You can make a post on the forum: <https://opendata-forum.cern.ch/c/lhcb/lhcb-ntupling-service/15>

Comments

No comments yet.

Type your comment here. Markdown is supported.

Add comment