









Valeriia (Lera) Lukashenko

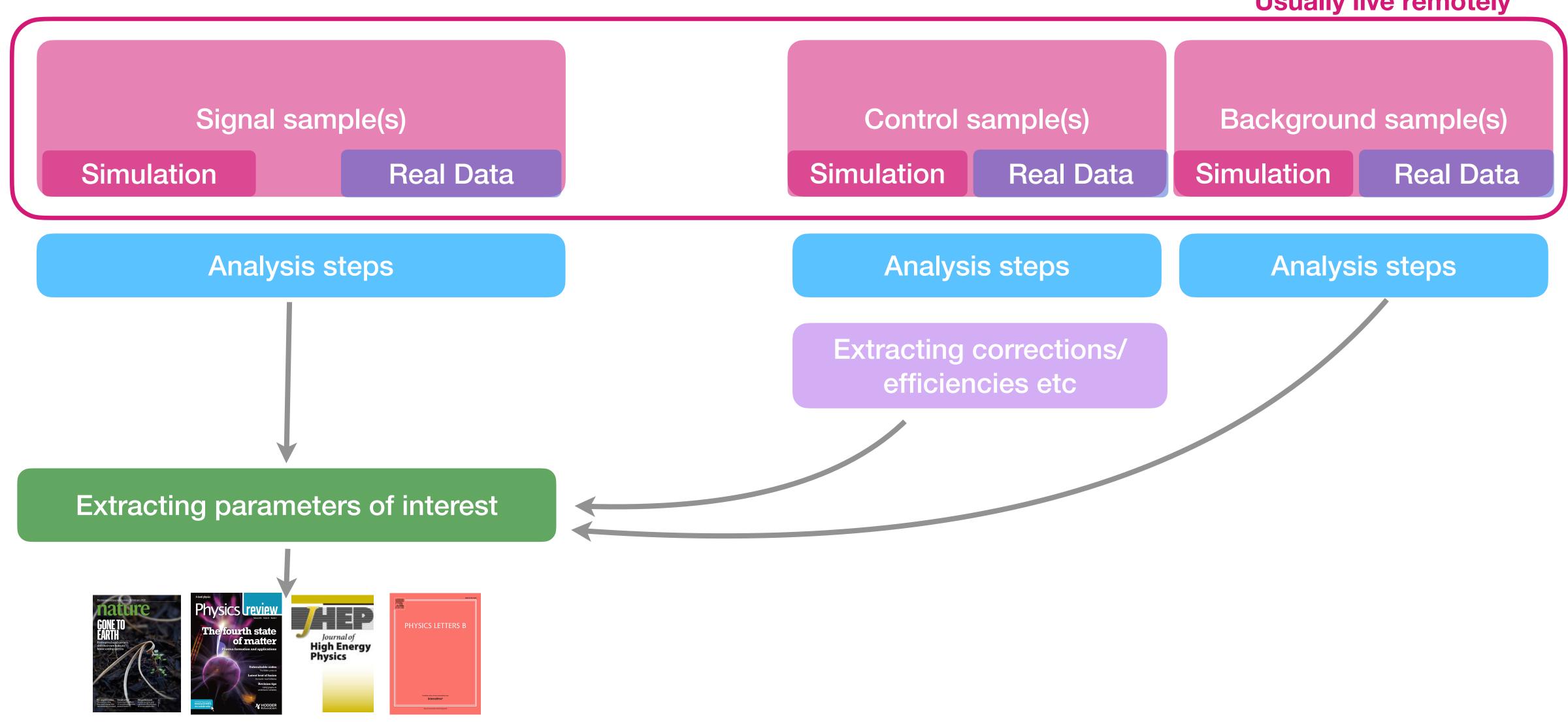




~1500 collaborators ~20 countries **100 institutes**







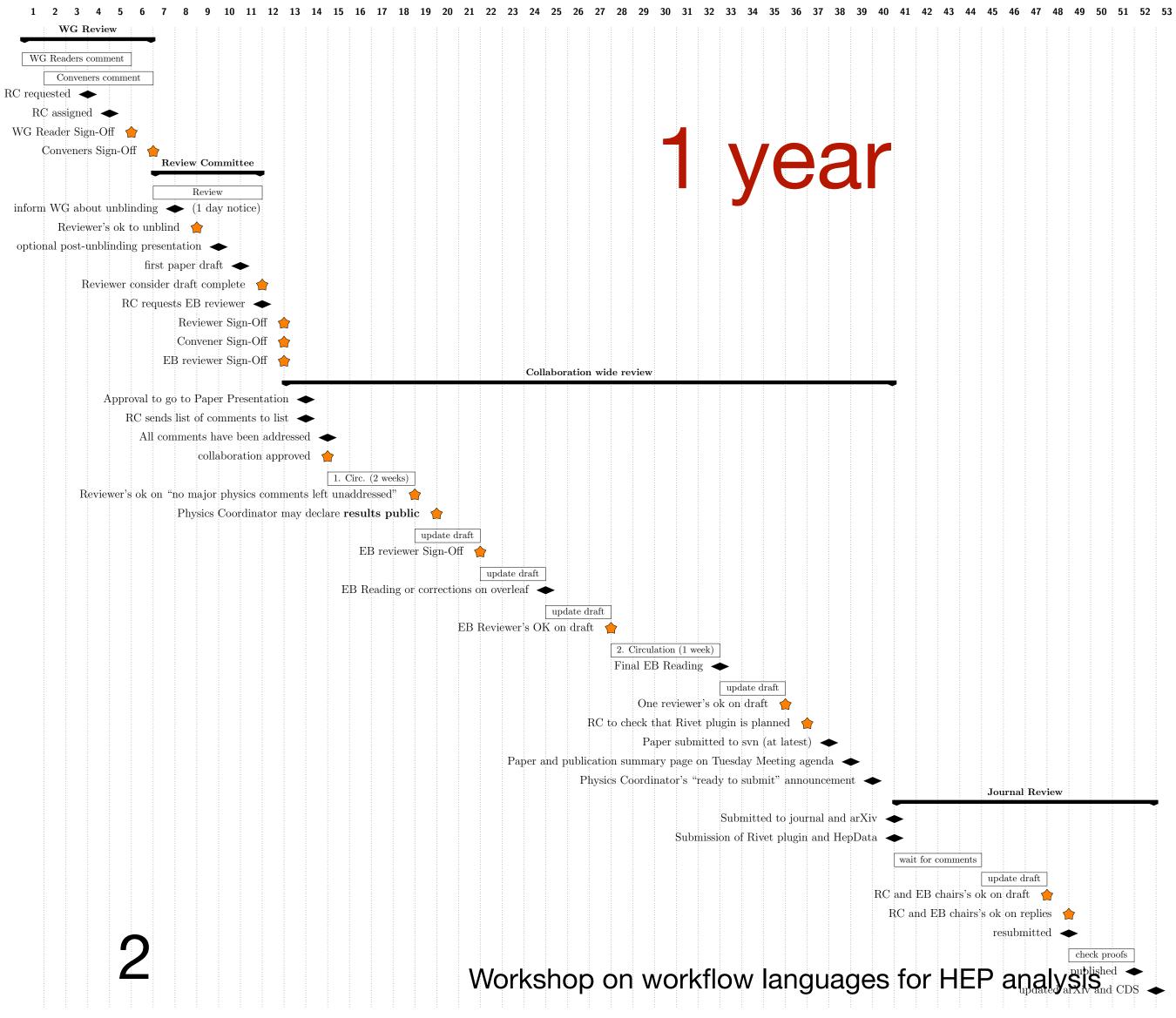
V. Lukashenko, Experience with Snakemake in LHCb

Usually live remotely

Timeline of a modern LHCb analysis

years of development

V. Lukashenko, Experience with Snakemake in LHCb

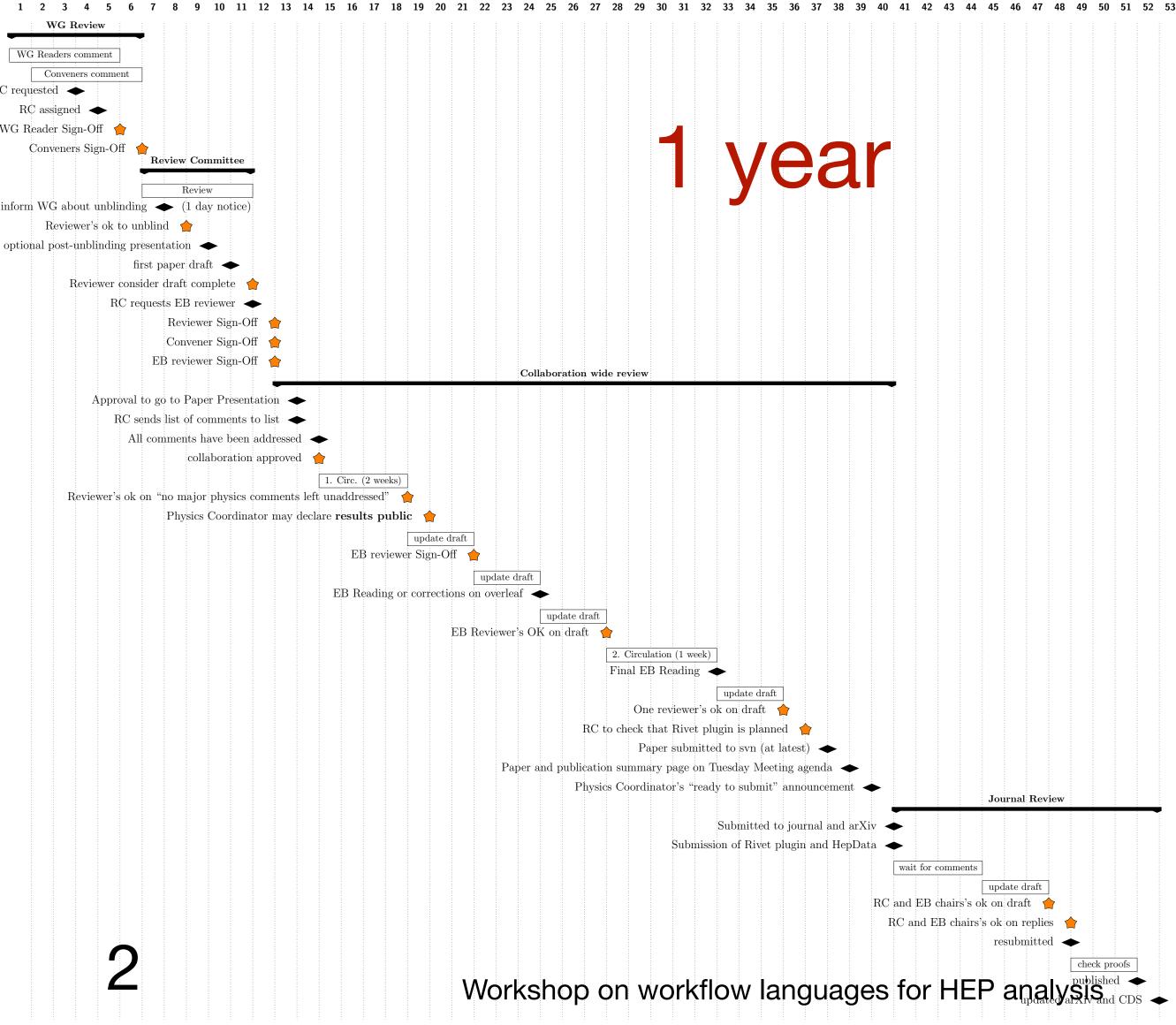




Timeline of a modern LHCb analysis WG Review WG Readers comment Conveners comment RC requested \blacklozenge 6 years ago RC assigned \checkmark WG Reader Sign-Off Conveners Sign-Off **Review Committee** Review inform WG about unblinding \blacklozenge (1 day notice) Reviewer's ok to unblind

years of development

V. Lukashenko, Experience with Snakemake in LHCb





Approximate timeline of a $B_s^0 \rightarrow J/\psi\phi$ analysis

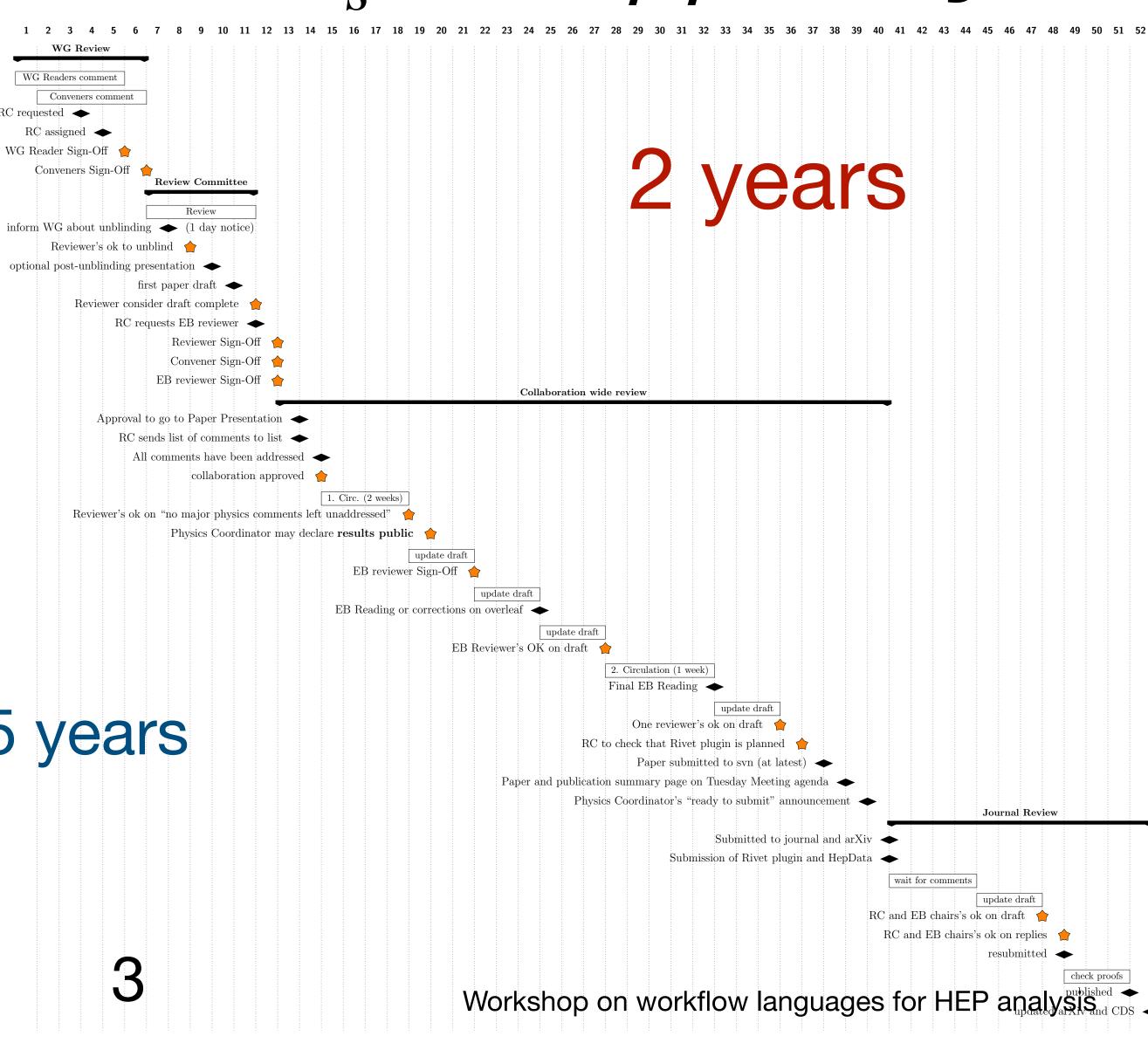
RC requested \blacklozenge

years of development

if not including partial Run 2 analysis, which another ~ 3 years

Average PhD lifetime ~ 3.5 years

V. Lukashenko, Experience with Snakemake in LHCb



Approximate timeline

WG Rea С RC requeste WG Reade

inform W optional j

years of development

if not including partial Run 2 analysis, which another ~ 3 years

Average PhD lifetir

Note: that people who left and people who joined where not necessary hired in the same group or had an overlap of their contracts!

V. Lukashenko, Experience with Snakemake in LHCb

\mathbf{P} of a B_s^0	$\rightarrow J/\psi\phi$ analysis
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19	9 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
WG Review	
Readers comment	
Conveners comment	
sted 🔶	
assigned 🔶	
der Sign-Off 🔶	
onveners Sign-Off 🔶 Review Committee	2 years
$\frac{\text{Review}}{\text{VG about unblinding}} \bigstar (1 \text{ day notice})$	
Reviewer's ok to unblind	
l post-unblinding presentation 🔶	
first paper draft 🔶	
Reviewer consider draft complete 🏾 🔶	
RC requests EB reviewer \blacklozenge	
Reviewer Sign-Off 🍵	
Convener Sign-Off 🔶	
EB reviewer Sign-Off 🔶	Collaboration wide review
Approval to go to Paper Presentation \checkmark	

While I was there:

4 PhDs left (one unexpectedly) 3 new PhDs joined (including me) 1 postdoc left 1 postdoc joined

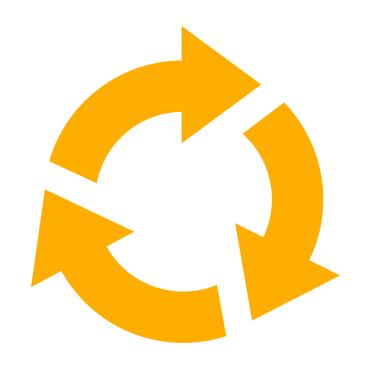
Workshop on workflow languages for HEP analysished +

RC and EB chairs's ok on replies \blacklozenge

resubmitted \bullet



The difference of big analysis - it is not just me!



Huge people turn over

V. Lukashenko, Experience with Snakemake in LHCb





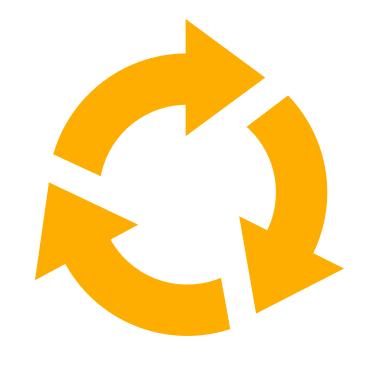
No time to train the newcomer

4

Archeology of others code



The difference of big analysis - it is not just me!



Huge people turn

over

V. Lukashenko, Experience with Snakemake in LHCb



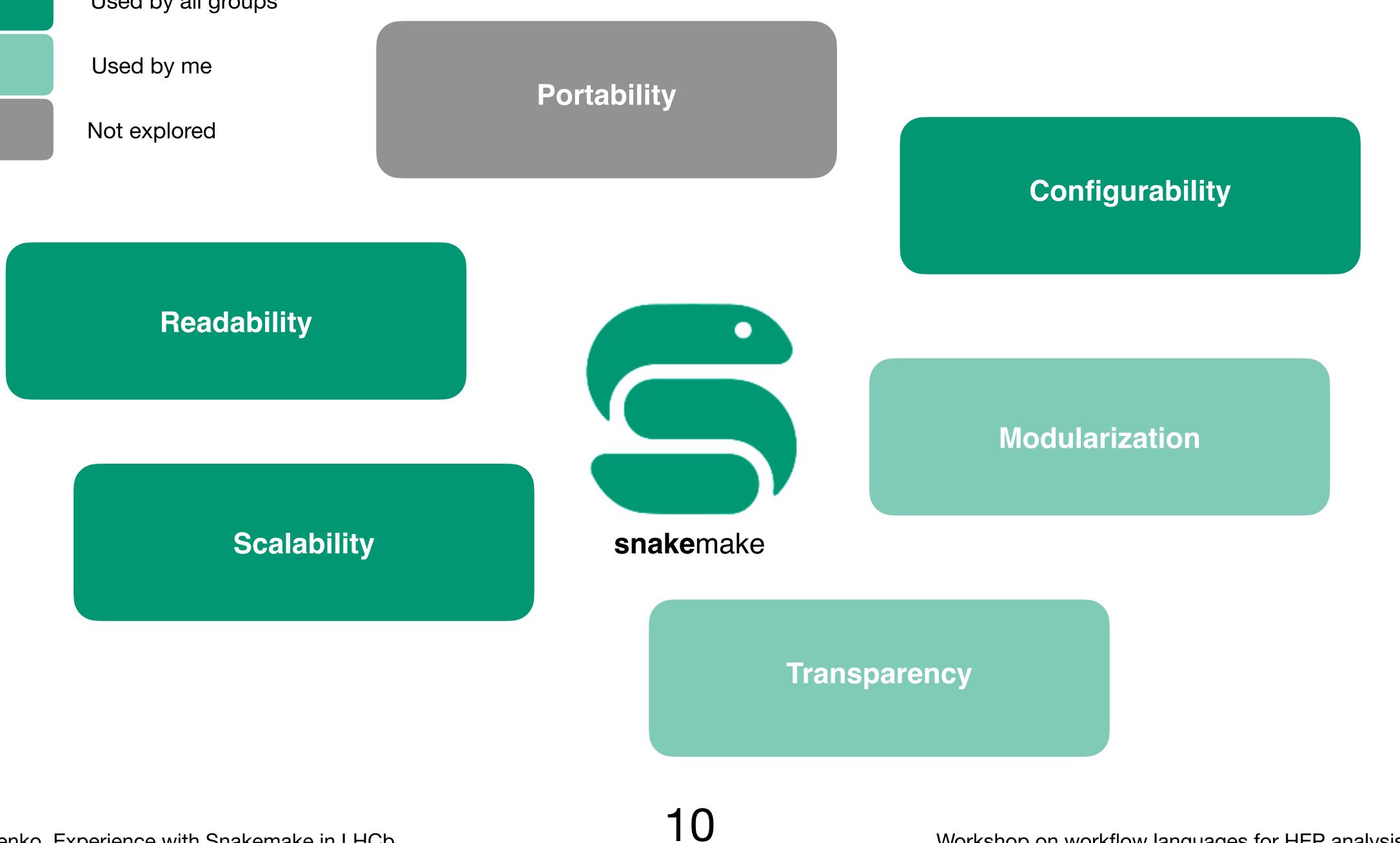
No time to train the newcomer

Archeology of others code

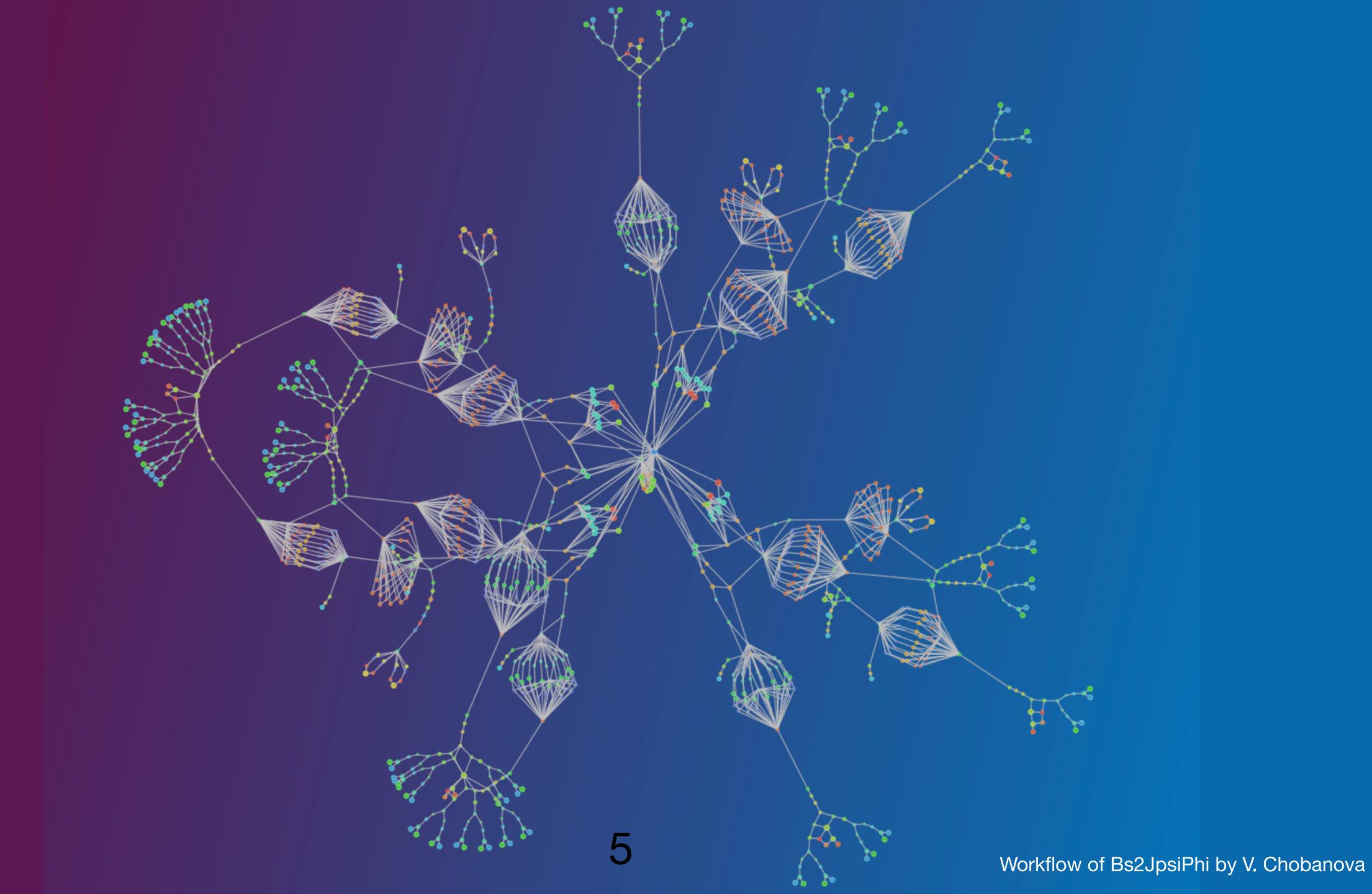




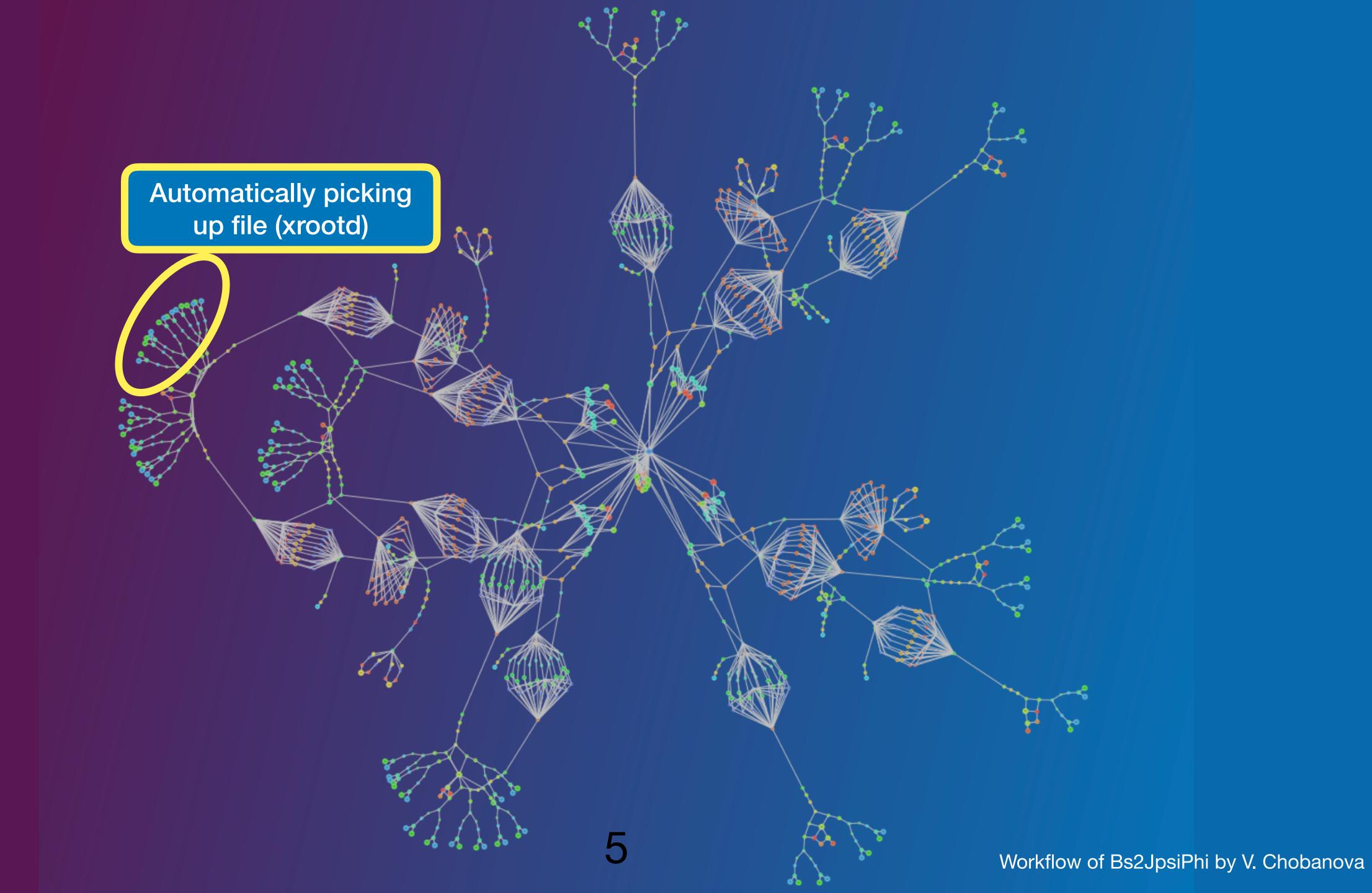
Used by all groups



V. Lukashenko, Experience with Snakemake in LHCb



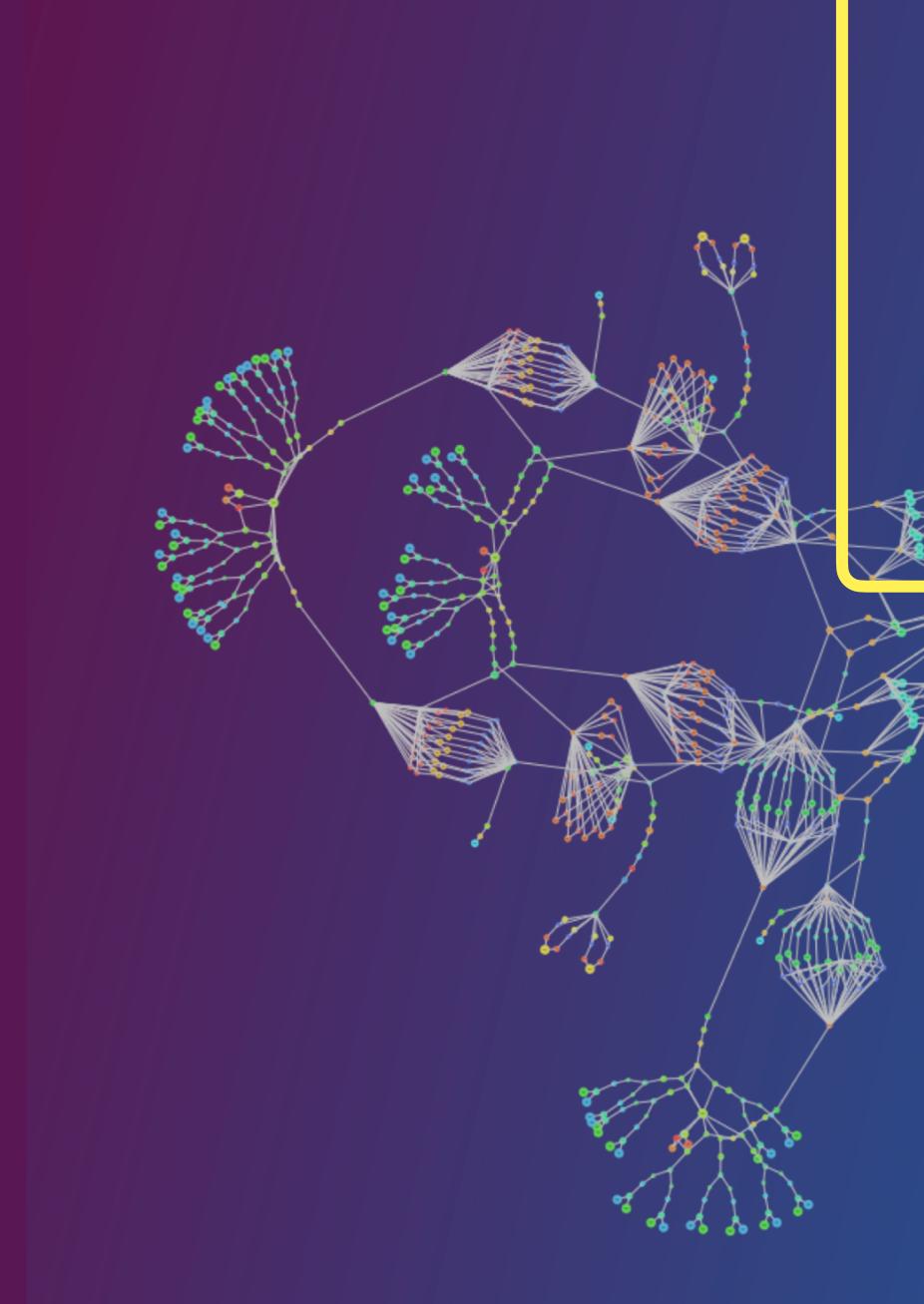












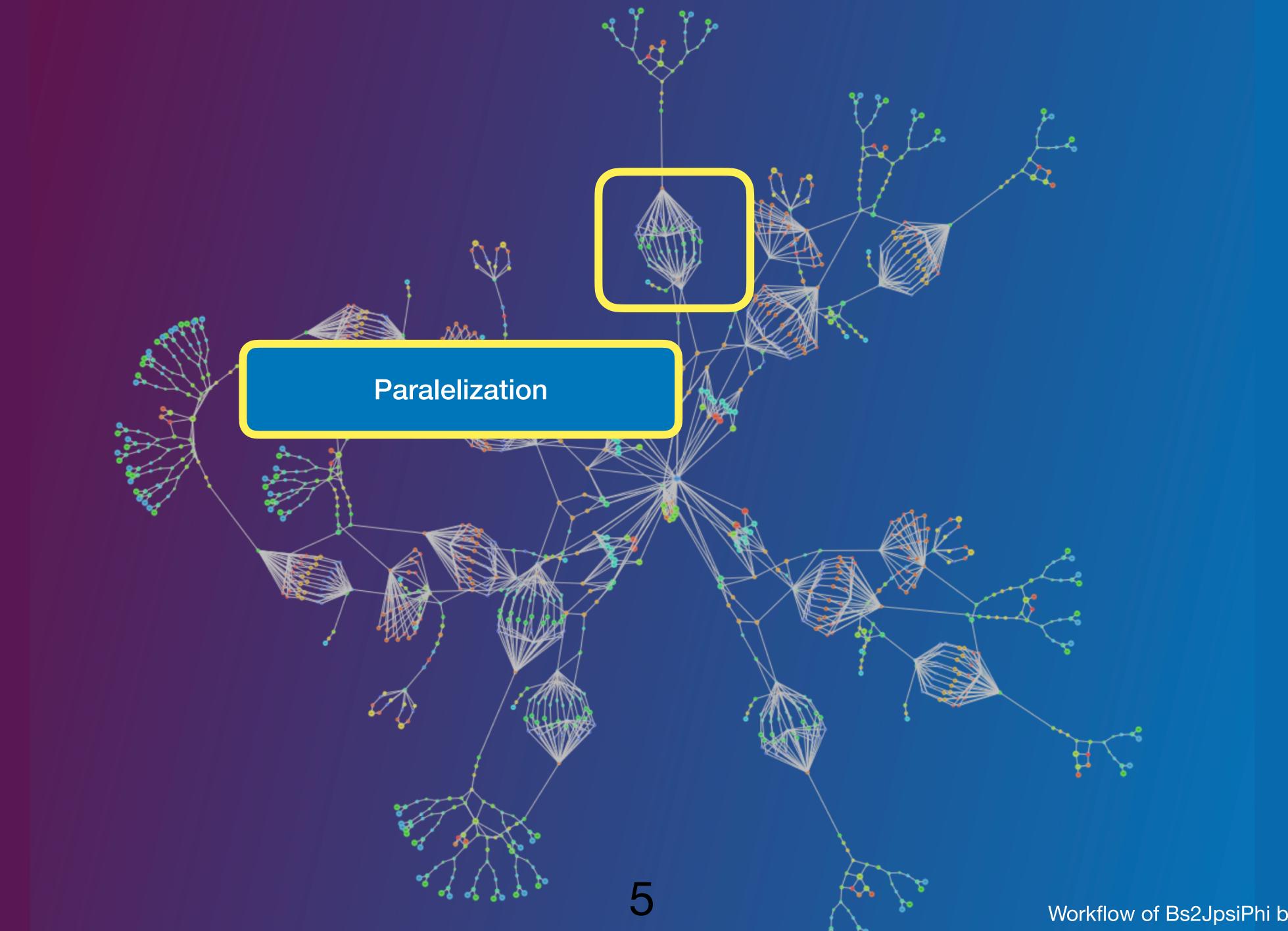
5



All workflow files are collected in a big master snakefile

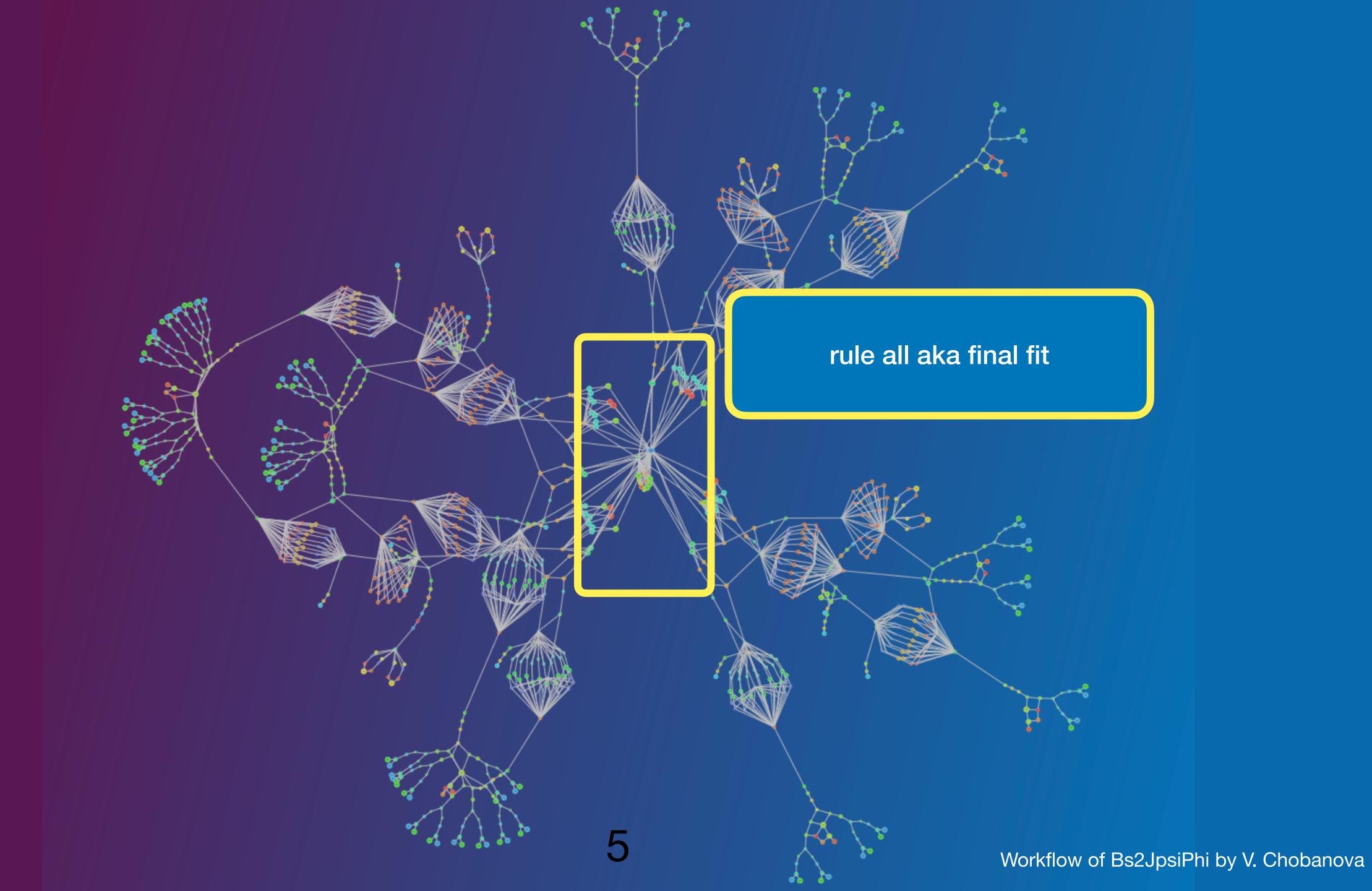
Workflow of Bs2JpsiPhi by V. Chobanova



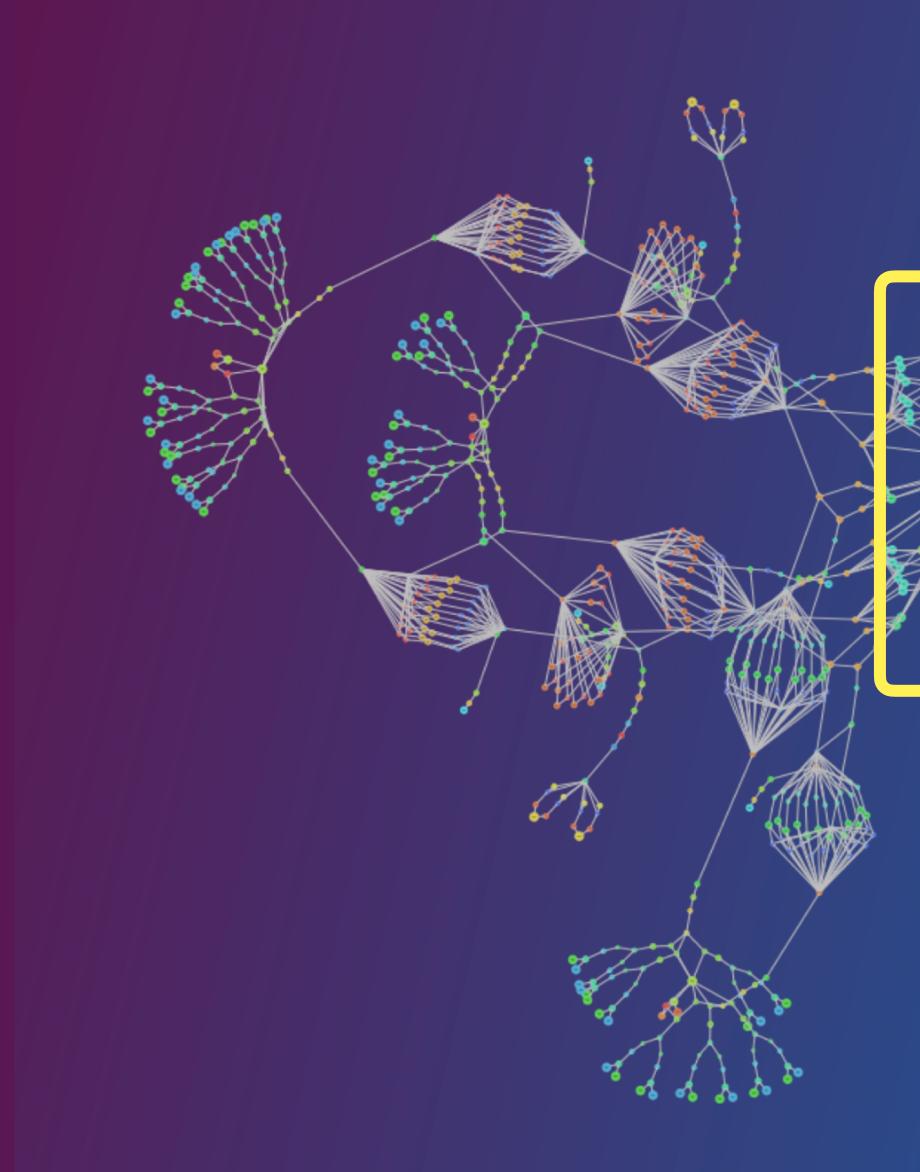


Workflow of Bs2JpsiPhi by V. Chobanova









5

rule all aka final fit

Systematics:

 Snakemake Paramspace for looping over constant params, like conservative +/- σ Modularisation: reuse slightly modified rules from other workflows, like mass fit with a different shape

Workflow of Bs2JpsiPhi by V. Chobanova





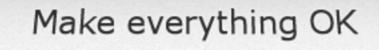
Danger



It is dangerous to rerun the workflow blindly, especially inherited

Combat with gitlab-ci and automated snakemake unit tests

V. Lukashenko, Experience with Snakemake in LHCb



make everything ok button

How to increase the number of users?



Promote:

training more and more people

V. Lukashenko, Experience with Snakemake in LHCb

21 September 2011

LHCb Publication Procedure

This document describes the steps that should be followed for the publication of analyses using data from the LHCb experiment.

Enforce:

no analysis is published without Snakemake

Starterkit: on-boarding training

Analysis essentials			
HEP Software Foundation			
Search docs			
CONTENTS:			
An introduction to Python			
Advanced Python Tutorial			
Introducing the Shell			
UNIX shell			

* » Analysis automation with Snakemake

G Edit on GitHub

Analysis automation with Snakemake

Learning Objectives

- Learn what analysis automation is and how it helps with analysis preservation
- Learn how to create a pipeline with Snakemake

Documentation and environments

V. Lukashenko, Experience with Snakemake in LHCb



Attendance: ~40 in person ~100 online

Snakemake lesson is one of the most well received

Starterkit advice : write workflow as soon as you start analysis

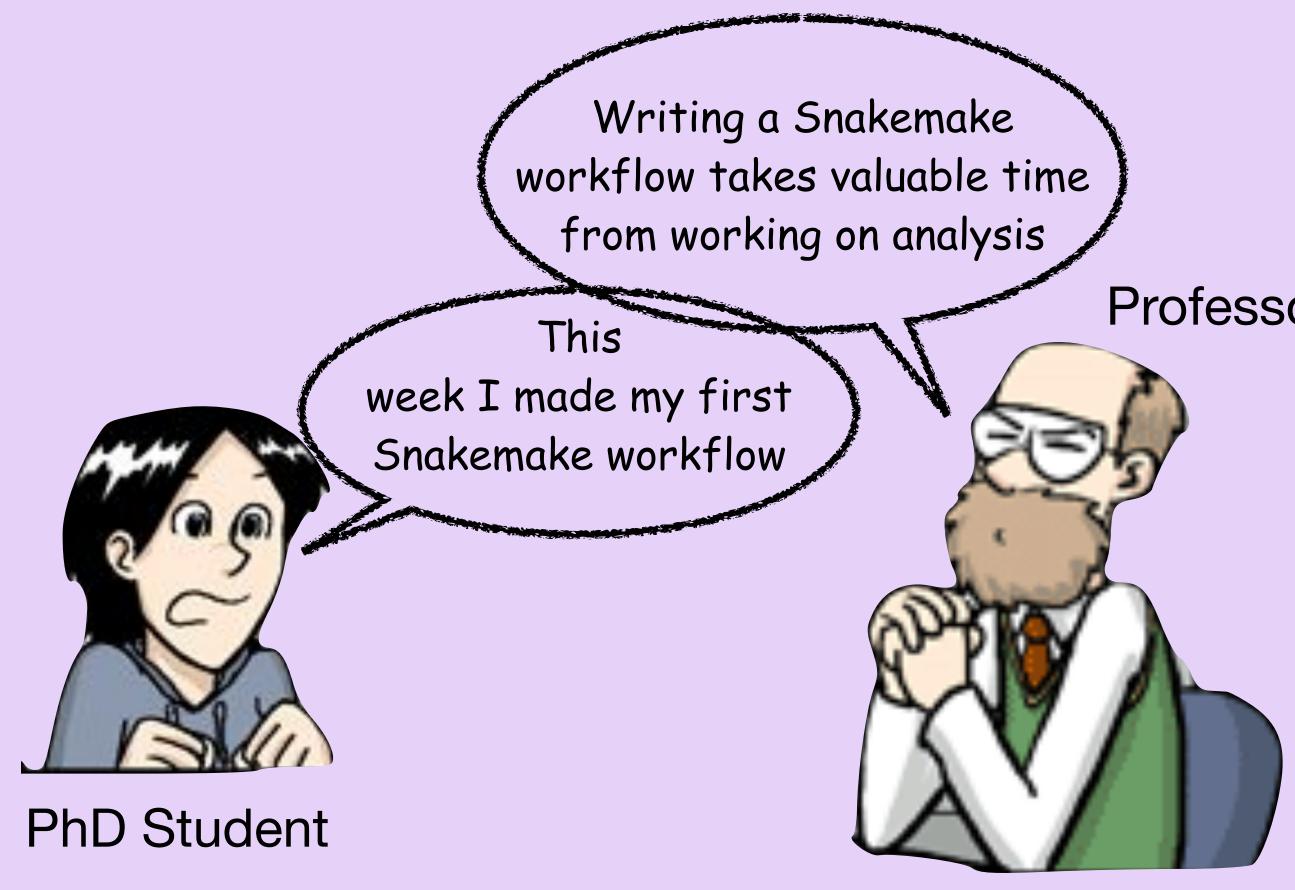
The LHCb snakemake training lives under HEP Software Foundation training umbrella - available for anyone to use

8





A common problem after Starterkit



JORGE CHAM @ WWW.phdcomics.com

V. Lukashenko, Experience with Snakemake in LHCb

Professor





Workshop on workflow languages for HEP analysis

Α

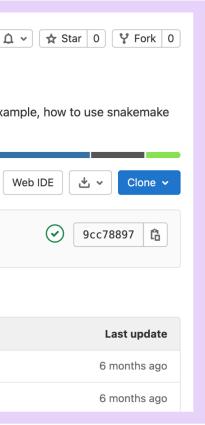
Analysis Workflow Template Project ID: 11583

🖓 0 Tags 🛛 🗈 184 KB Files 🛛 🗔 290 KB Storage

lemonstrates best practices for designing a user-analysis workflow. It contains an example, how to use snakemake can serve as a template for new analyses

master	analysis-workflow-template / +	History Find file Web IDE Clone ~
Merge branch 'add-independent-paramspace-workflow' into 'master' ••• Sebastian Neubert authored 3 weeks ago		
README	CI/CD configuration	
Name	Last commit	Last update
Scripts	rearrange and small improvement	6 months ago
snakefiles	rearrange and small improvement	6 months ago

LHCb snakemake template



Discussion points

- from small ones?
- There is a huge danger when the workflow is considered "working" it is less likely that in addition to workflow? How to incorporate unit tests in the best way?
- How to change this?
- workflows for the early measurement/data and use it as a monitoring tool?

• It is more than just saving workflow for the purposes of analysis preservation. For big analysis (> X people) having a defined workflow is a necessity to make sure the things are up-to-date and do not get lost. What should be guidelines for the big analysis groups? Are they different

individual jobs outputs get checked regularly. I myself relied too much on "this is an automatic procedure and therefore trivial", which is a logical mistake. How do we promote more testing

• Promotion among the older generation: this is a waste of time, when you could do physics.

• In a view of upcoming upgrades (and the just finished upgrade of LHCb) - should we prepare