MAX-PLANCK-GESELLSCHAFT

Max-Planck-Institut für Physik
(Werner-Heisenberg-Institut)

MC requests in JIRA

M Flowerdew (MPP Munich)

ATLAS MC tutorial, 29th September 2015

Why request MC?

- **Dumb answer:** Because ATLAS says so

ATL-GEN-INT-2015-001

Approved ATLAS results must be based on official MC production samples.

Preliminary material shown in group meetings can use unvalidated private data sets. However, analyses asking for approval must be based on fully validated official samples.

- This begs the question: **why?** Several reasons, including:
  - Validation
  - Reproduceability
  - Availability to the collaboration
  - Efficient use of grid resources
  - ...

- The next question is why you want **this** sample

Types of MC request (a little simplified)

- **Extension of an existing sample**
  - More statistics, with **no** changes to AMI tags
- **Use existing evgen with modified sim/reco**
  - Example: to compare AtlFast-II with fullsim
- **New evgen, based on an existing sample**
  - Examples:
    - To add an evgen filter
    - Systematic variations (different scales, PDF etc)
    - More points for your Higgs/BSM parameter scan
- **New model/process**
  - The **first** generation of a particular SM process or BSM model
    - Includes a change in generator and change in MC campaign (mc12, mc14, mc15, ...)

The what and why

https://twiki.cern.ch/twiki/bin/view/AtlasProtected/AtlasProductionGroup

- **Think through your request** as much as possible before you start

- **Discuss with**
  - **Conveners**, **subconveners**
  - **Generator experts** (for technical help)
  - Your group's **MC contact**

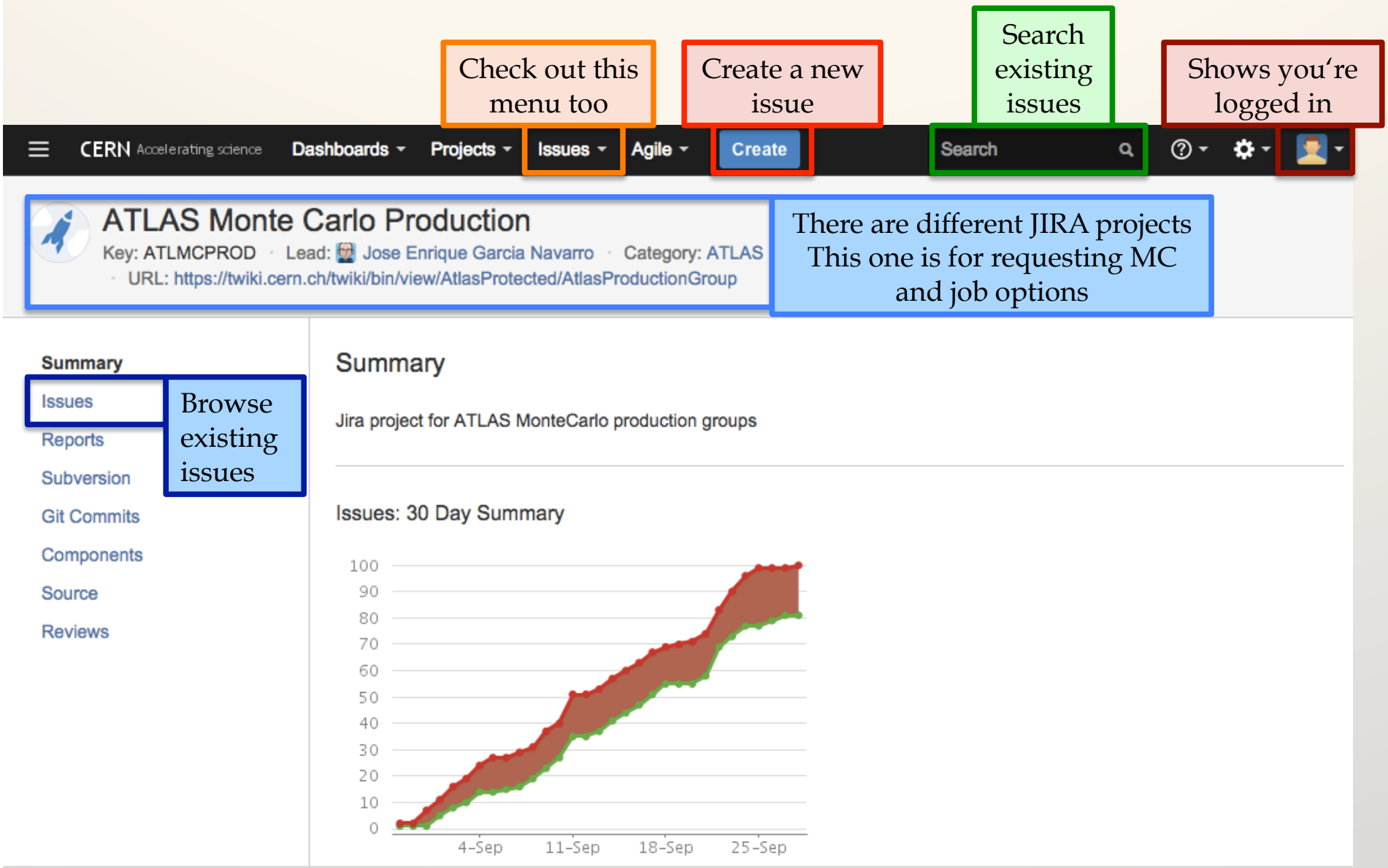
- **Questions to think about**
  - **How many datasets?**
  - **How many events?**
  - **What priority?**
  - **Justification**: for you? for ATLAS?
  - Remember to write clearly and that not all readers will be expert in your sub-channel
    - Example: How is your request different to other apparently similar requests?

- Don't panic if you don't know the answers to **all** of the questions before you start – JIRA is also great for discussions & feedback

JIRA front page

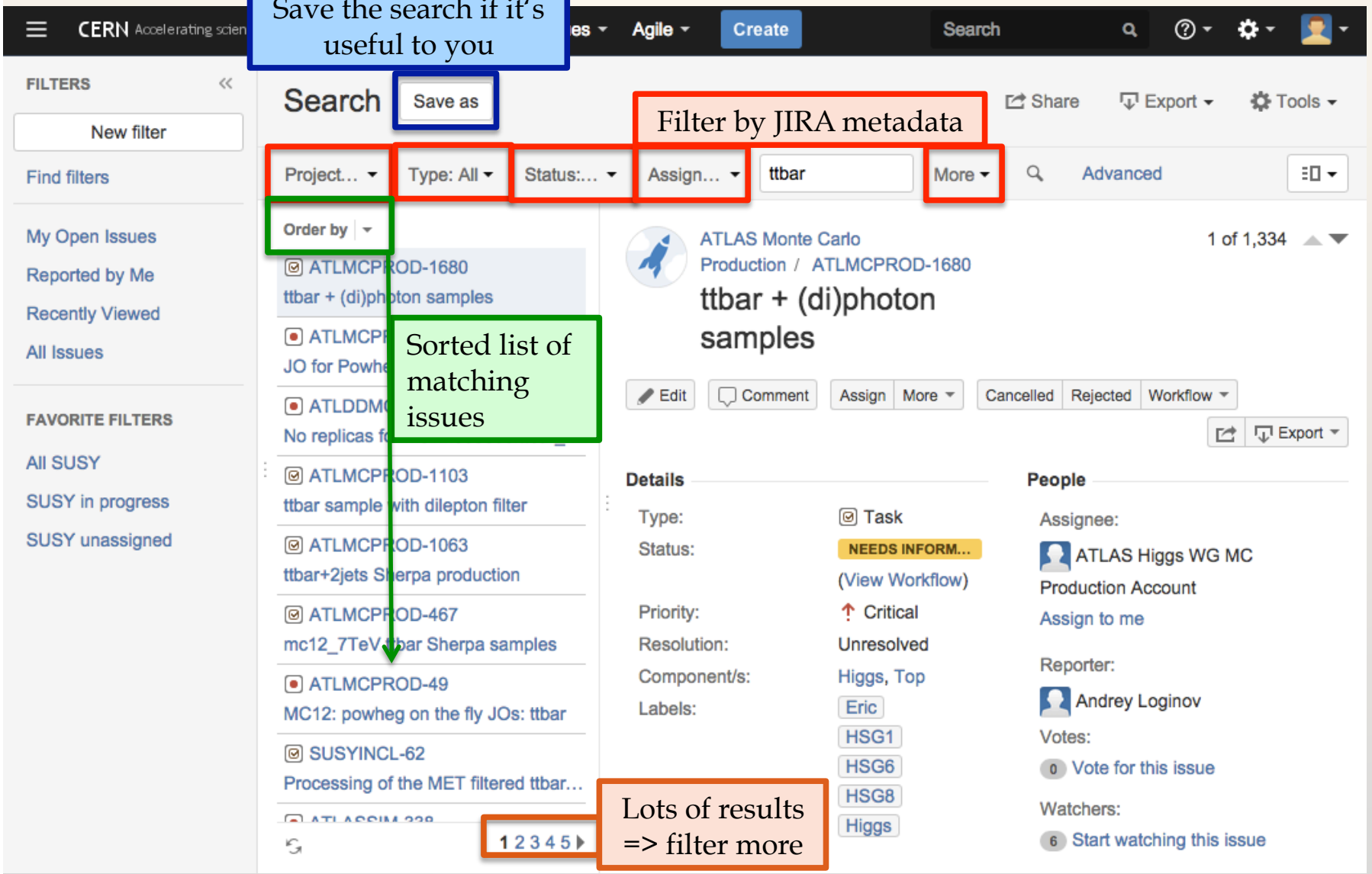
https://its.cern.ch/jira/browse/ATLMCPROD

JIRA front page

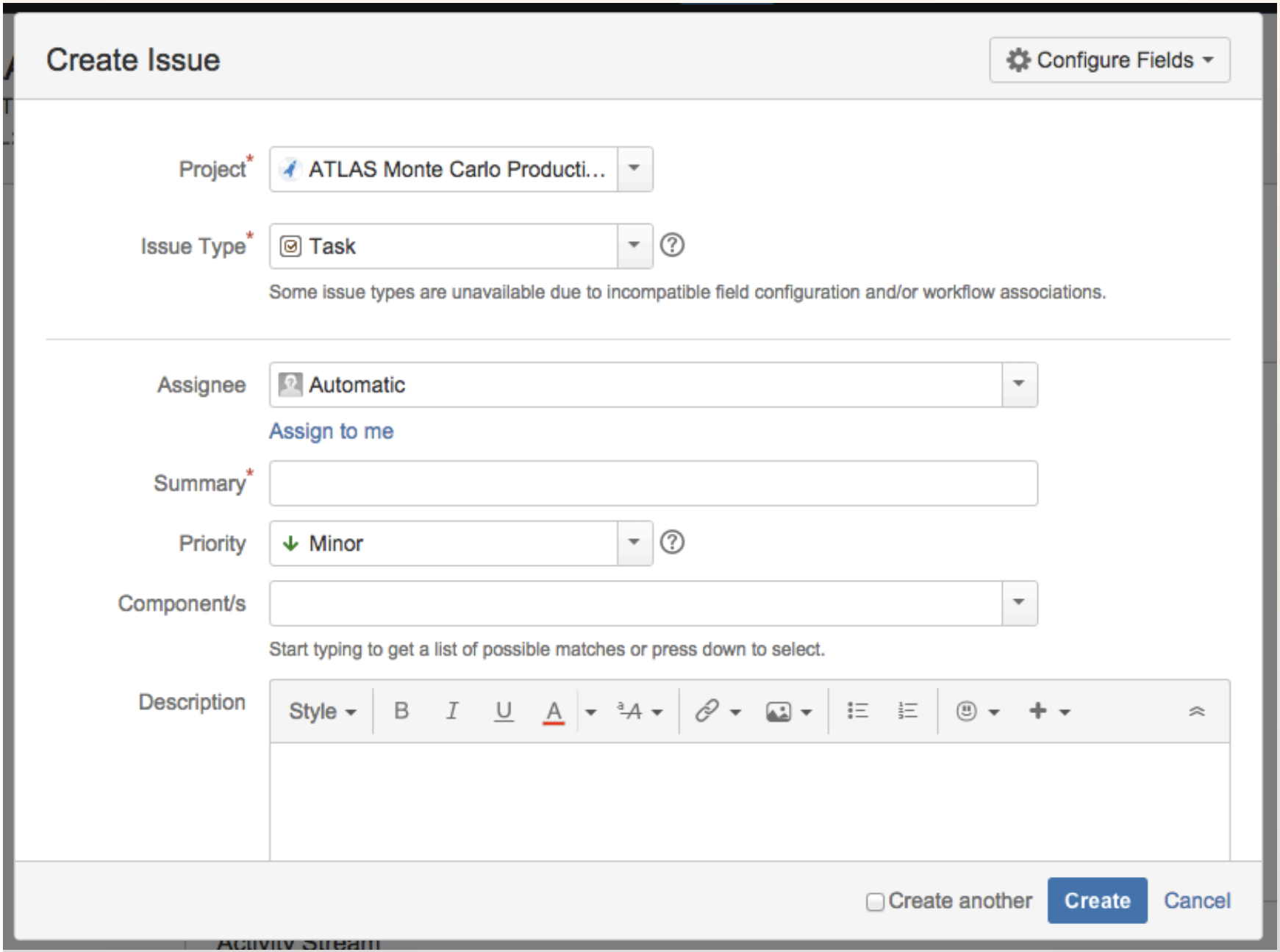


JIRA search

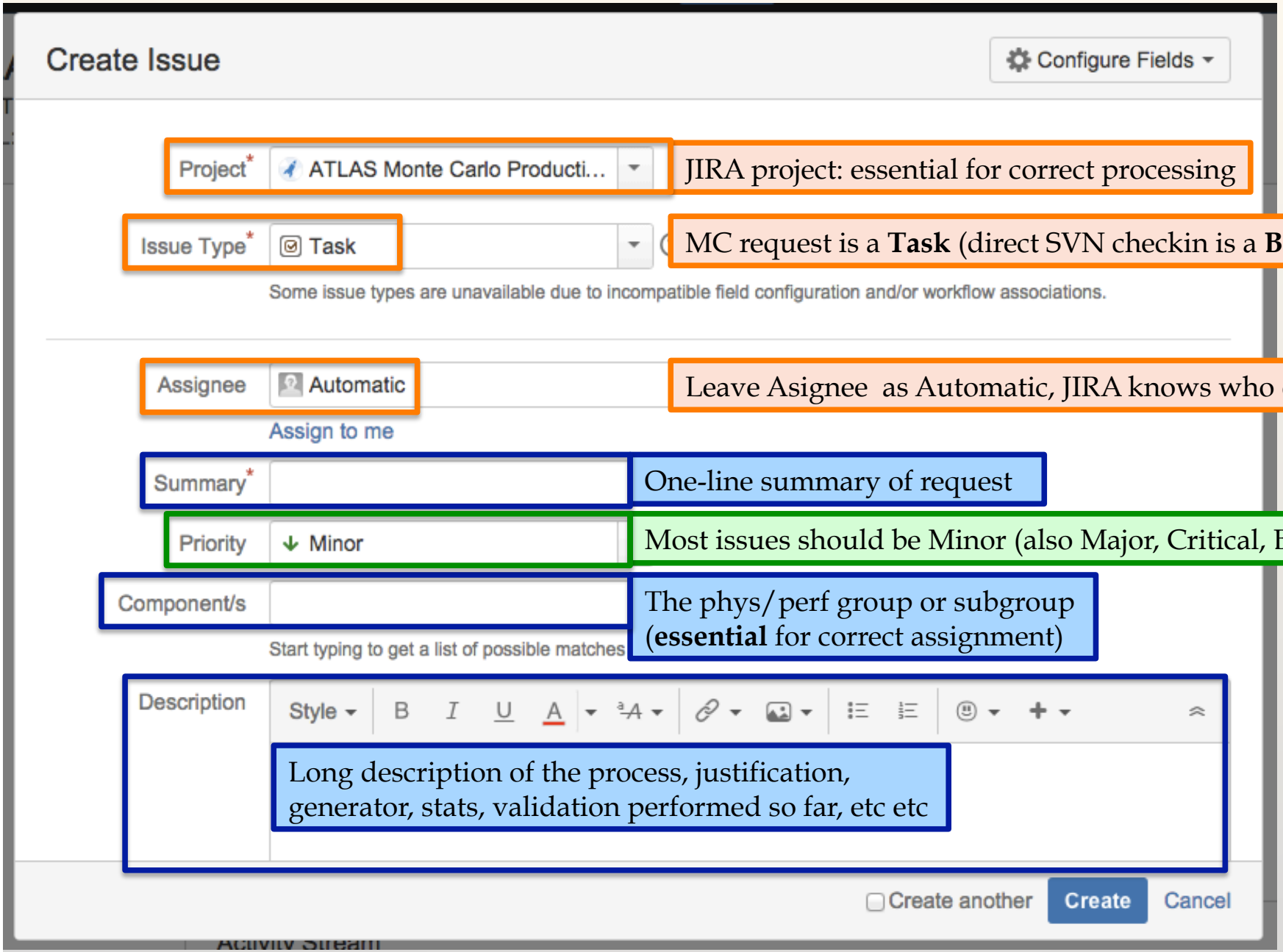
Example: typing „ttbar“ in the search box



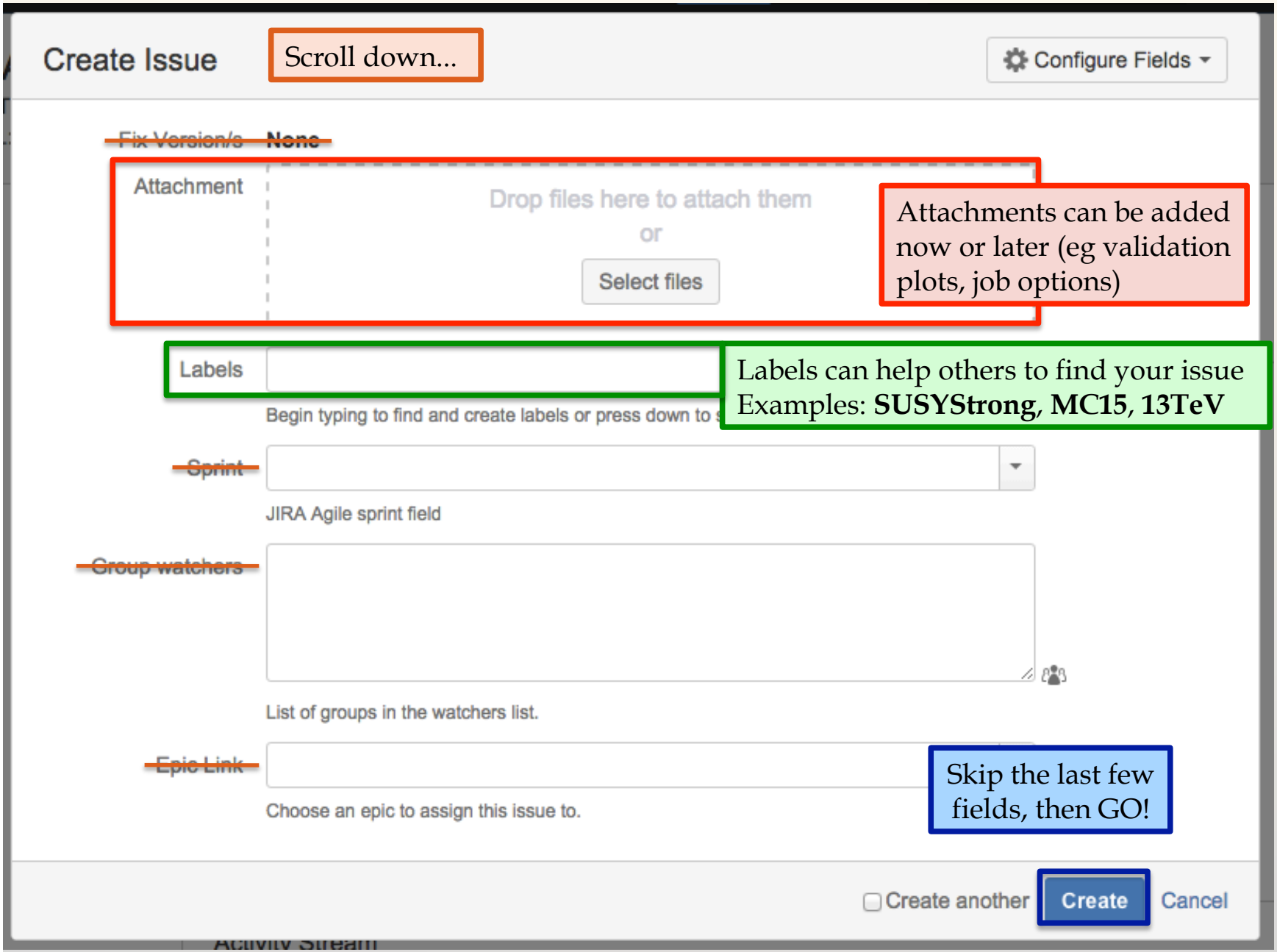
Create an issue



Create an issue

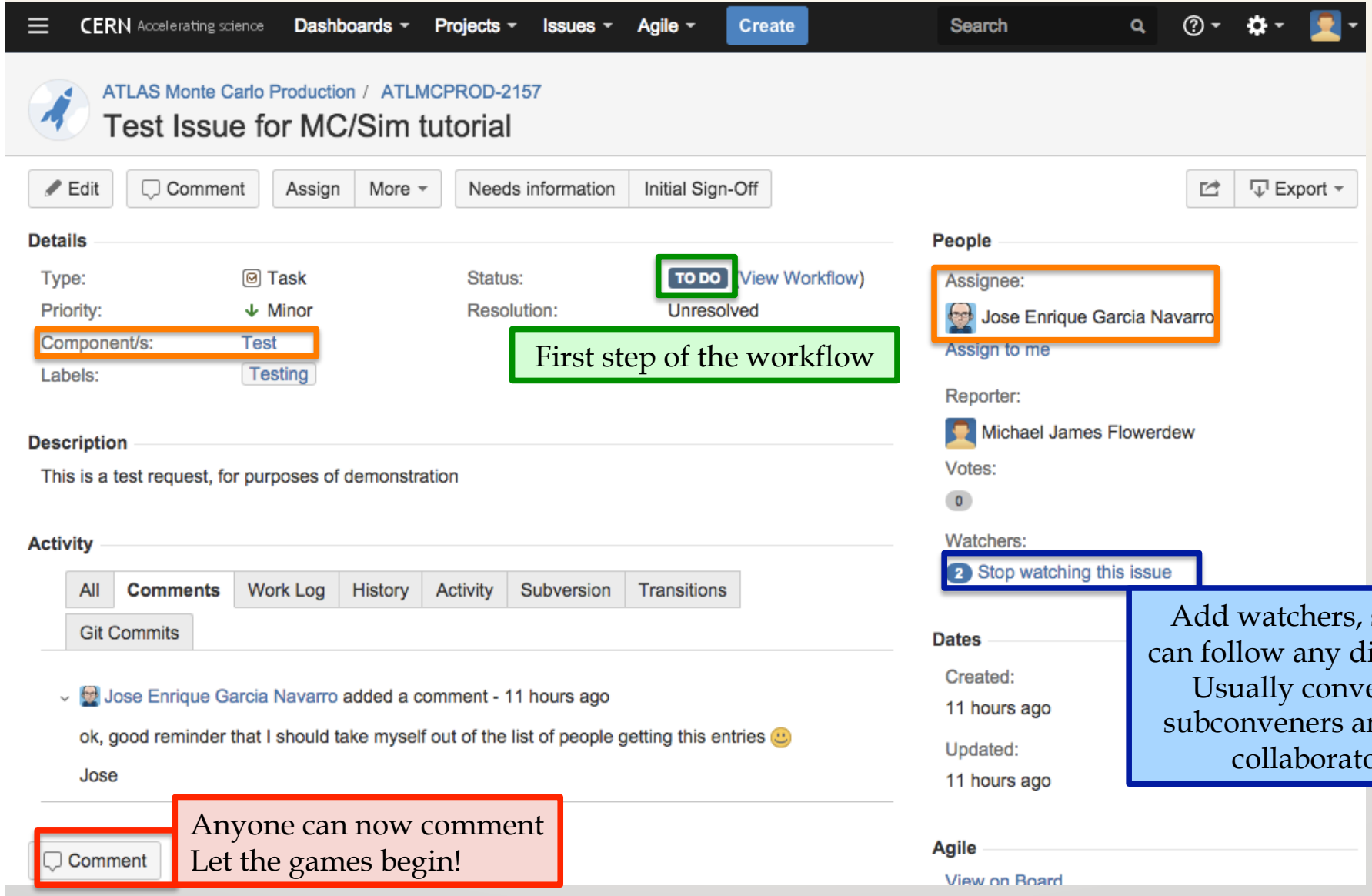


Create an issue



New request created!

I used the **Test** component, because I don't want this example to be confused with a real MC request. It has been automatically assigned to Jose (sorry Jose!).



Done by requester
Done by someone else

What next?

The issue then passes through a number of steps defined by the workflow

Issue created in JIRA

Initial signoff, assign DSIDs

Write JOs + spreadsheet + initial validation

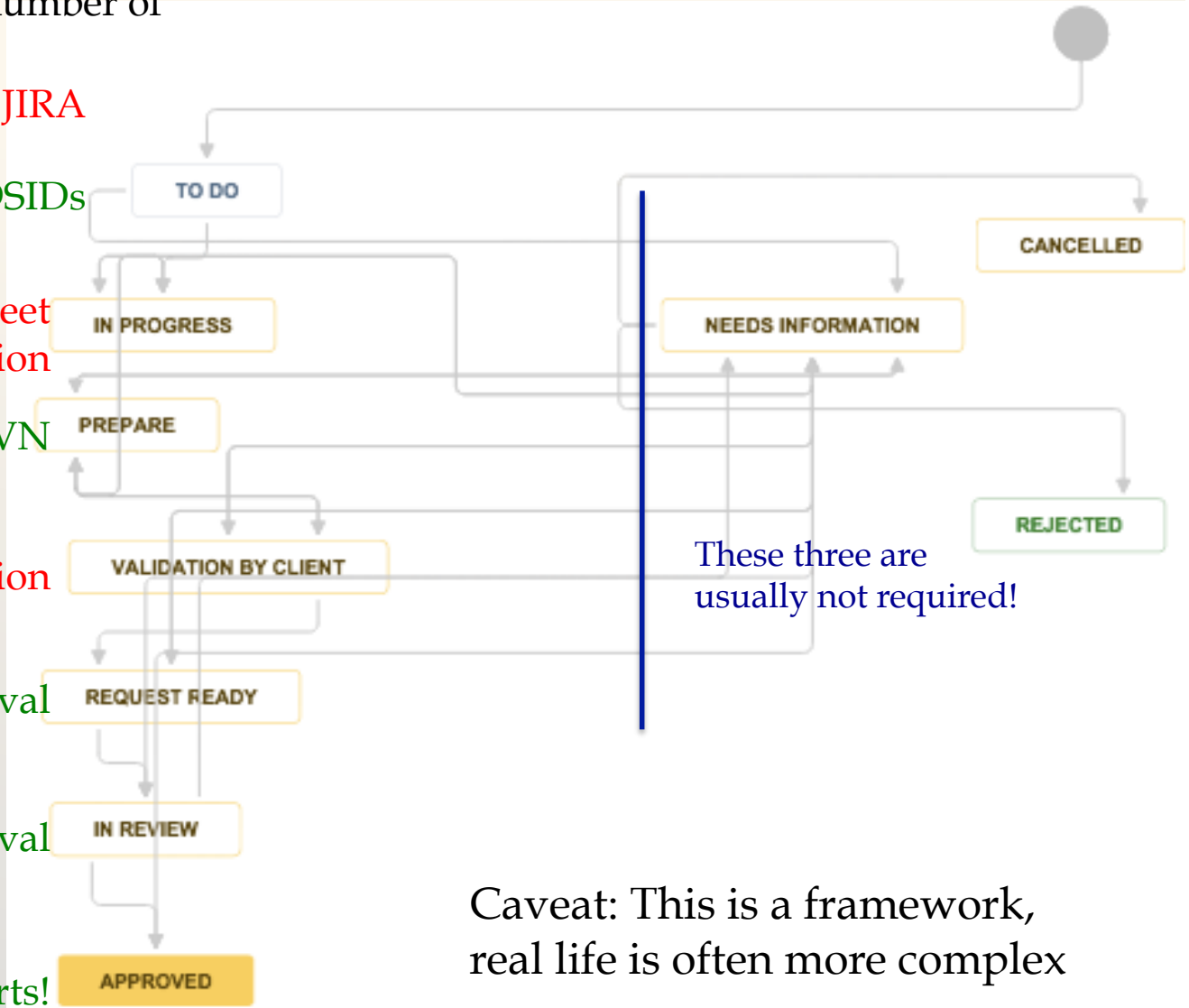
JOs go to SVN

Final tests + validation

Ready for convener approval

Ready for PMG/PC approval

Production starts!



These three are usually not required!

Caveat: This is a framework, real life is often more complex

Input for a request

- One must eventually follow words with actions

- Usually four kinds of input are required:
  - **Job options** + **associated files** needed to run the jobs
    - MC15JobOptions: easy & quick to update (see backup)
    - Other packages: requires time to tag a new release/p-cache
    - Additional files: next slide
  - **Validation plots**
  - **Athena log files** (log.generate)
  - **Spreadsheet**

Associated files

- Some generators need additional files to function or speed up production
  - **LHE files** for non-OTF production
  - **Integration grids** for Powheg, Sherpa and MadGraph5_aMC@NLO
    - Powheg: see https://twiki.cern.ch/twiki/bin/view/AtlasProtected/PowhegForATLAS#Re_using_integration_files
    - Sherpa: ask atlas-generators-sherpa@cern.ch for help
    - MadGraph5_aMC@NLO

  https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MadGraph5aMCatNLOForAtlas#Gridpack_generation

- The files themselves need to be prepared
  - MC contact & prodsys will upload them to the grid
- The file `MC15JobOptions/share/evgeninputfiles.csv` needs to be updated
  - Connects grid datasets with individual MC samples
  - Note the naming conventions when creating the files!

Line	
1	DSID,energy,inputeventfile,(inputconffile)
2	301535, 13000,, group.phys-gener.sherpa020100.301535.Sherpa_CT10_eegammaPt10_35_13TeV.TXT.mc15_v1
3	301536, 13000,, group.phys-gener.sherpa020100.301536.Sherpa_CT10_mumugammaPt10_35_13TeV.TXT.mc15_v1

- Add lines like these to your job options:

```
evgenConfig.inputfilecheck = "QBH_emu"
```

(LHE)

```
evgenConfig.inputconfcheck = "eegammaPt10_35"
```

(integration grids)

Validation, log files

- Everyone wants official production to make the thing they asked for

- Details and scope of validation vary from request to request
  - **Bare minimum:** log files using job options *in SVN using the final athena release*
    - Technical test, timings (**<24h**), filter efficiencies etc
    - Also useful to have logs early on (ie before SVN), *but this does not prove they will work in production!*
    - Only sufficient in some cases, eg extending the mass range for existing BSM model
  - **New processes/models need:** plots of kinematics *using the final athena release*
    - (Jets, leptons, MET, Z, SUSY, ...) x (p_T, η, ϕ, m, ...) + filter variables etc
    - Use previous campaigns, other generators etc as reference
  - Some cases need weeks/months of hard study
    - Now we‘re moving out of MC requests and into „real“ physics
    - Example: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TtbarBoson

Spreadsheet

- A spreadsheet provides information to prodsys in a systematic way
  - Example based on

https://docs.google.com/spreadsheets/d/1WYqoFey9BIVBiDP97ePOZDDRDRNwdIi1HNeca04RR1Y/edit#gid=0

	A	B	C	D	E	F	G
1	**Brief description**	**Dataset Number**	**Formats**	**Joboptions**	**Events (Fullsim)**	**Events (ATLFAST II)**	**Priority**
2	A sample	123456		MC15...		100000	2

H	I	J	K	L	M	N	O	P	Q
Evgen	**Simul**	**Merge**	**Digi**	**Reco**	**Rec Merge**	**Rec TAG**	**Atlfast**	**Atlf Merge**	**Atlf TAG**
AMI tags: important for some requests									

Core request information

R	S	T	U	V	W	X	Y
Cross-section (nb)	**Filter efficiency**	**NLO cross-section (nb)**	**Generator**	**ECM**	**EF**	**Comment**	**Response**
Additional info: useful but can be processed without				13000		25ns spacing	

Spreadsheet fields 1

	A	B	C	D	E	F	G
1	**Brief description**	**Dataset Number**	**Formats**	**Joboptions**	**Events (Fullsim)**	**Events (ATLFAST II)**	**Priority**
2	A sample	123456		MC15...		100000	2

Brief description: Exactly what it says, best if unique within the sheet
Dataset number: Self-explanatory

Formats: Say if you wish to keep RDO, ESD or DESD
Can say, eg, RDO-10 to keep 10% of the raw files

Job options: Full job option file name MC15.999999.Generator_TunePDF_Process.py

Events: Enter the number of events you want to produce
Do not enter numbers in both columns. If you want fullsim **and** AF-II, you need two rows
Evgen-only: Use the fullsim column and write „Evgen only“ in the comment column

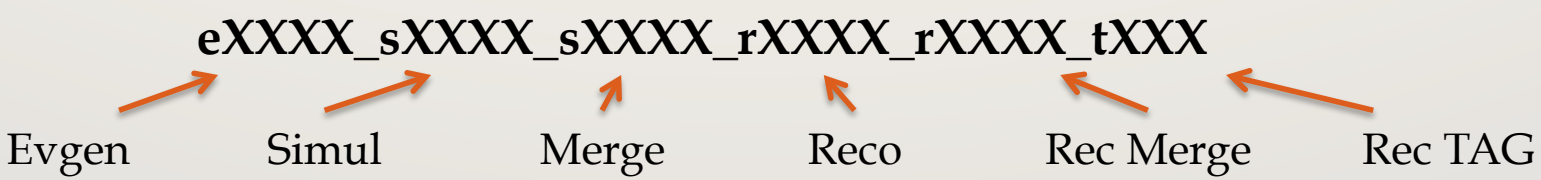
Priority: 0 = urgent, 1 = standard, 2 = long-term
Urgent requires justification (eg it‘s holding up a paper)

Spreadsheet fields 2

H	I	J	K	L	M	N	O	P	Q
Evgen	**Simul**	**Merge**	**Digi**	**Reco**	**Rec Merge**	**Rec TAG**	**Atlfast**	**Atlf Merge**	**Atlf TAG**

- Fill this bit if you have specific release/AMI tag requirements, for example
  - Consistency with a previous sample
  - Non-standard sim/digi/reco
  - Statistical extension of an existing sample (**please always fill the tags for this**)
- A lot of requests („standard MC15“) don‘t need anything here

- You should understand the tags now, but just in case:



Spreadsheet fields 3

R	S	T	U	V	W	X	Y
Cross-section (nb)	**Filter efficiency**	**NLO cross-section (nb)**	**Generator**	**ECM**	**EF**	**Comment**	**Response**
				13000		25ns spacing	

These columns are mostly for information only (eg might help during approval)

(NLO) Cross-section: Useful to judge if the number of events is appropriate
Filter efficiency: May have CPU/walltime issues if low... but you tested this, right?
Generator: Probably redundant, should be obvious from JO name

Please **always** fill the next two

ECM: In GeV
Comment: Essential info for prodsys
Should always say „Evgen only“ or give pileup bunch spacing
Any other info that varies from sample to sample and may be useful to prodsys
Eg will this sample require high CPU for evgen? (how many events/24h?)

EF and **Response** are basically obsolete now

What next?

http://bigpanda.cern.ch/prodsys/prodtask/request_table/

- Once approved by the working group, the request moves to panda for PMG/PC approval
  - Done by the MC responsible, you don't need to worry about the details

https://prodtask-dev.cern.ch/prodtask/inputlist_with_request/3757/

ATLAS PanDA | Requests | Tasks | Train | Datasets | Error retry | DDMDash | SSB | Quick search | Help | Login

ATLAS PanDA

	Request ID:	Description:	Reference:	Manager:	Physic group:	Project:	Status:
menu(≡)	3757	Simulation and Reconstruction of samples of gluino production with one step decays	ATLMCPROD-1976	jferrand Me	SUSY	mc15_13TeV	finished

0 + MC15.371581.MGPy8EG_A14N_GG_onestepCC_1705_1105_505.py mc15_13TeV:mc15_13TeV.371581.MGPy8EG_A14N_GG_onestepCC_1705_1105_505.evgen.EVNT.e4039/

(Atlfast)Evgen exists events: 20000

	e4039	a766						a777	r6282	t599	submitted	edit (saved)
T:		done						done	done	done		

1 + MC15.371582.MGPy8EG_A14N_GG_onestepCC_1705_1345_985.py mc15_13TeV:mc15_13TeV.371582.MGPy8EG_A14N_GG_onestepCC_1705_1345_985.evgen.EVNT.e4039/

What does this all mean? Next talk!

Exercise

- Create an issue in JIRA!
  - https://its.cern.ch/jira/browse/ATLMCPROD
  - **Important:** We don't want to confuse real requests with tests, so please
    - Use Component **Test**
    - Do **not** tag conveners etc or add them as watchers – use your neighbour instead
  - Use one of the samples from the other practical sessions, or use your own favourite process/model

- You don't have validation plots etc, but think about what you **would** provide
  - For example: if you look up the cross-section & filter efficiency (now you know how!), how many events do you think you would need to provide?

- **Try some JIRA functionality**
  - Add your neighbour as a watcher, add a comment and tag them (@username)
  - Add a comment with a bullet/numbered list, or some code
    - Can you find how to preview your post? Formatting guidelines are there too
  - Attach a file (hint: explore the **More** menu)
  - Link to another JIRA issue (how many ways can you find to do this?)

ADDITIONAL INFO

Job options + associated files

- Most MC requests require updates to **Generators/MC15JobOptions**
  - Described yesterday, let's assume you know what you need

- MC15JobOptions is unusual, it is not part of any release
  - A tarball with important files is made for production jobs
  - This arrangement is flexible: one can request arbitrary MC15JobOptions tag with any ATLAS release
  - But testing local modifications can be tricky, as CVMFS may override your copy
    - Checkout, add your modifications, then compile
    - Easiest to copy --jobConfig and any param cards etc to your local directory
    - Python includes should be found OK if compiled properly
    - More details (aimed at SUSY, but most suggestions can apply generally):

https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/SUSYMcRequestProcedure#2_1_Quick_EVGEN_running_start_up

- Depending on what you're requesting, you may also need to provide:
  - Updates to other packages (GeneratorFilters, XxxControl and so forth)
    - Allow time to tag, validate in the nightlies and to build new release/p-cache
  - Associated files (LHE files, integration grids aka gridpacks)
    - Generator-dependent, again allow time to produce them