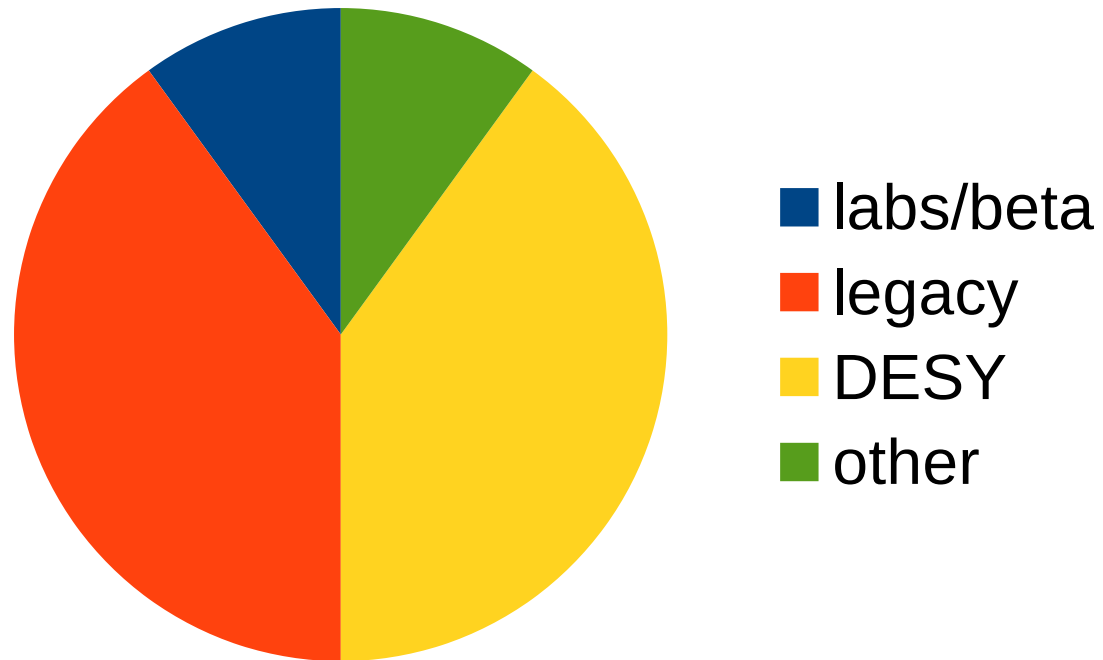


# Status HEP curation tools



# labs / beta

## user submissions + arXiv harvest

harvest

matching

automatic merging

refextract

**selection:** core guesser, references, keywords

Adding/changing CORE & subject (unless auto-ingested)

## editor

exceptional tasks, single records

resolving errors/conflicts from harvest

refextract

authorextract

# legacy / production

## BibEdit

curation of new records

**manual merging** of new (holdingpen1) records  
(creating new records)

## MultiEdit

curation - cleanup

## RecordMerger

## API +

## BatchUpload

**all DESY workflows** (see next slide)  
many local scripts at DESY and FERMI:  
adding CNUM  
cleaning collaborations, adding experiments  
long author lists  
general metadata cleanup



....

## Harvest @CERN

APS  
Elsevier  
PoS

# DESY\* - stand-alone

## journal workflow

**harvest** (apart from above)

**matching** (BibMatch - stand-alone)

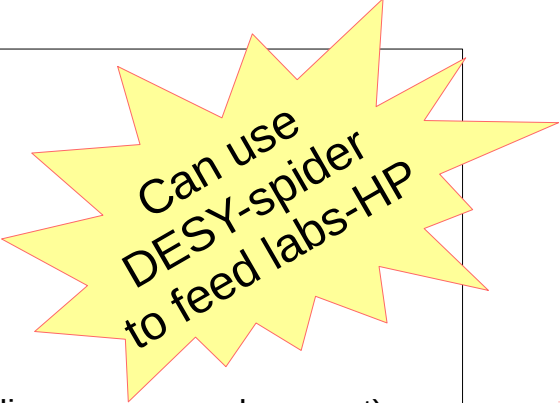
**affiliation guesser**

**add INSPIRE-ID from email**

**refextract on structured references from feeds**

**semi-automatic merging** (preparing reduced records for holdingpen+append+correct)

**selection: keywords, references, authors, source**



Can use  
DESY-spider  
to feed labs-HP

## arXiv harvest

**add core (and subject) to auto-ingest articles** (gr-qc, nucl, ...)

GROBID@CERN: affiliations, PACS, keywords

affiliation guesser

## book-keeping of curation

assigning tickets/records (not sure if only RT is enough)

## adding keywords

**assign article to external specialist**

**collect and check keywords, upload**

\* mostly (based on) tools hacked together to bridge a few month  
in the transition from SPIRES to INSPIRE

## For tools in use:

### bug-list

Improving workflow logic – when a WF matches an active WF  
When workflows match only one should be active (match candidates stay active)  
Auto-ingest records should be accepted independent of status of a matched WF  
If WFs are merged (one stopped by another), records should be merged  
...

### needed-list

Status-indication for completed workflows (now only for manually selected WFs)  
Improving GUI / formats, information is missing  
Improving matcher, validator  
Show match candidates scrolling, but new workflow record static  
Auto-ingest user submissions from catalogers  
Editor: add field in specific place (e.g. author, reference)  
Editor: split fields (e.g. keywords)  
....

### nice-to-have-list

Auto-accept based on CORE guesser  
Same coloring for decisions for HEP and author  
Same brief format in list, references, holdingpen  
.....  
...  
.

# Summary HEP curation tools

