# KDIG Related Sessions

## KDIG Session

## Monday May 13 - Salle Danjon - 17:00 - 18:30

| Speaker | Title | Duration | Materials |
| --- | --- | --- | --- |
| Kai Polsterer | Introduction | 3' | |
| Petr Skoda | deep learning methods on LAMOST DR2 | 20' + 6' | |
| Antonio D'Isanto | Is VO ready for machine learning? | 20' + 6' | |
| All | Open Discussion: are we ready for data-science? | 35' | |

# Active Learning
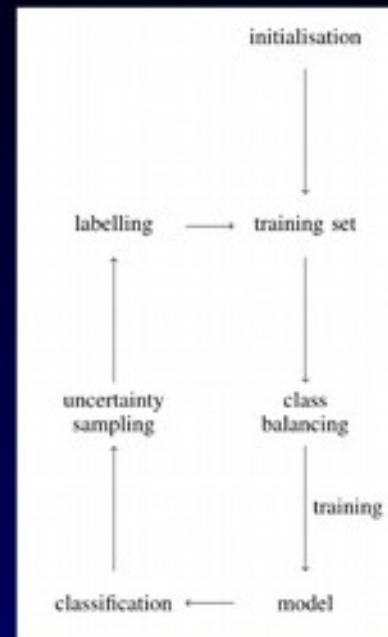
# Active Learning

Task: searching for rare stellar spectra

- Complicated data retrieval for massive data-set
  → just possible with the help by colleagues of China-VO

Machine Learning: refined model interactively by asking the user / domain expert for input

- Using VO-tools for the work-flow
  → manual orchestration through super-VO-user / very complicated

- Improving the data-set interactively
  → preserving provenance / reproducibility

Is the VO providing the right tools for massive data-exploration?

- orchestration / work-flow management / scripting

- how to find outlier / rare objects?

# Data retrieval



## Some "simple" tasks...

**Heidelberg Institute for Theoretical Studies**

1. Given the coordinates, download 28x28 pixel$^2$ images for all the quasars in SDSS.

2. Download some hundreds of thousands of images from FIRST/UKIDSS.

3. Download all the HARPS spectra from ESO archive.

Obtaining data products can be a not easy task

# Data retrieval

3 simple tasks to retrieve many optical image cutouts / radio+NIR images / all the spectra of a single instrument

- no SIA implemented / no public & documented access

- limitations on number of files / timeout

- poor documentation / too technically complex for astronomy

- no bulk processing

all problems got solved by asking friends in the VO directly

- in many cases the solution was not the VO solution
  → ftp-transfer of a whole survey
  → creating own HiPS to access the data
  → scripts that use non-standard undocumented functions

# Are we ready for data-science?

Discussion: Is the VO a digital plate archive or an observatory? Machine learning analysis can be seen as building telescopes for a virtual sky. But where do we place those telescopes? In it's current status Single Source Science (S$_3$) is well supported by many services.

- are science platforms a solution?

- how do we standardize those platforms and exchange code?

- who is paying for the compute?

For the upcoming IVOA interop in Groningen:

- there will be a KDIG session / reserve a room!

- define and present some data-science use cases
  to see how to solve those with the current VO landscape

- tests continue for bringing code to the data