

Yandex
Data Factory



SCHOOL OF DATA ANALYSIS

Imperial College
London



Improving Reproducibility of Data Science Research

Tatiana Likhomanenko, Alexey Rogozhnikov,
Andrey Ustyuzhanin, Alexander Baranov,
Egor Khairullin

11th July 2015

Irreproducibility Indicators

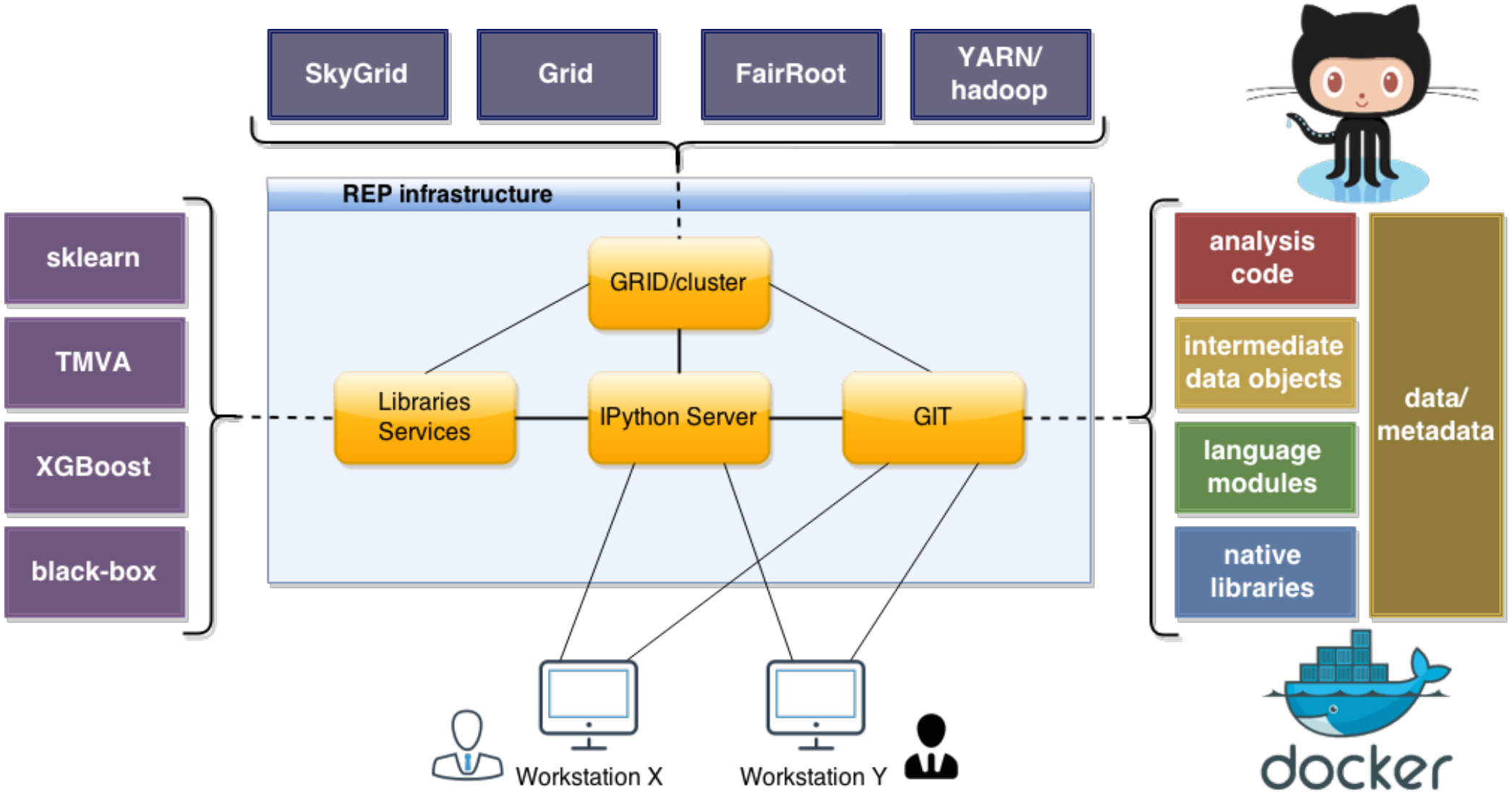
- ‘I can’t remember which version of the code I used to generate figure 13’
- ‘The new student wants to reuse that model I published three years ago but he can’t reproduce the figures’
- ‘I thought I used the same parameters but I’m getting different results!?’
- ‘On what dataset did I do the comparison exactly?’
- ‘It worked yesterday!’
- ‘Why did I do that?’

Irreproducibility Indicators

- ‘I can’ … the … enerate figure …
- ‘The ne… … ago but …
- ‘I though… results!?’
- ‘On what …
- ‘It worked …
- ‘Why did I do that?’



Reproducible Experiment Platform



Andrey Ustyuzhanin anaderi@yandex-team.ru, https://github.com/yandex/rep

REP for research & education

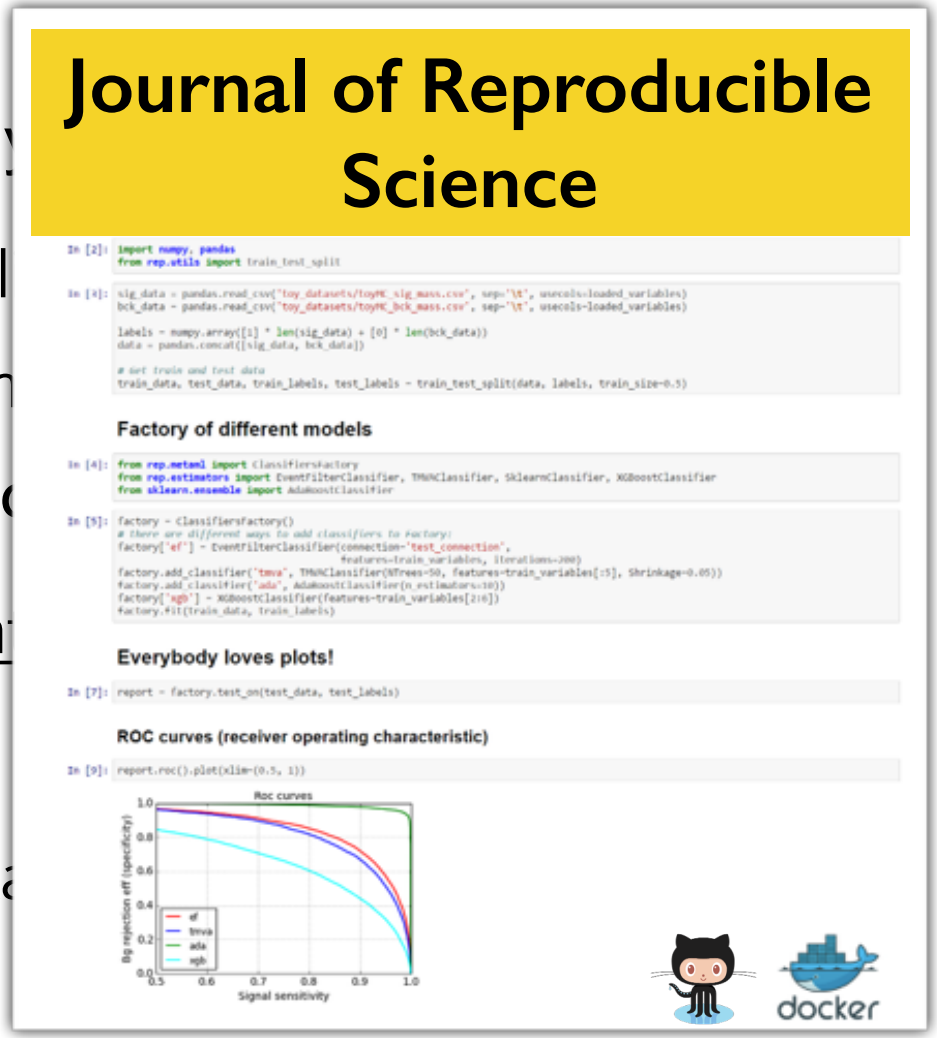
- **High Energy Physics**
  - online & offline data analysis at CERN
  - optimization of disk storage
- **AstroPhysics**
  - CRAYFIS (http://crayfis.io)
- **Business**
  - Yandex Data Factory
- **Education**
  - MLHEP summer school (http://www.hse.ru/mlhep2015/)

Yandex Data Factory

SCHOOL OF DATA ANALYSIS

REP for research & education

- High Energy
  - online & offl
  - optimization
- AstroPhysi
  - CRAYFIS (h
- Business
  - Yandex Data
- Education
  - MLHEP summer school (http://www.hse.ru/mlhep2015/)



Yandex Data Factory

SCHOOL OF DATA ANALYSIS