Rediscover the known Universe with NASA dataset

@LostInBritanny
@AurrelH95
@PierreZ
@moyowi

Pierre Zemb

@PierreZ

Infrastructure Engineer working on distributed systems

OVH

Finist Devs

STARTUP WEEKEND BREST 26-27-28 JANVIER 2018 NO TALK ALL ACTION! SUP' JAVOUHEY

DEVOXX 4KIDS

WARP 10 MEETUP

Aurélien Hébert

@AurrelH95

Software Engineer

and data lover

Horacio Gonzalez

@LostInBrittany

Spaniard lost in Brittany, developer,
dreamer and all-around geek

OVH
Team DevRel

Breizh C@mp
Mix de technologies

Finist Devs

WARP 10 MEETUP

Google Developers Experts 2015
Web Technologies GDE

Emmanuel

we ♥ you

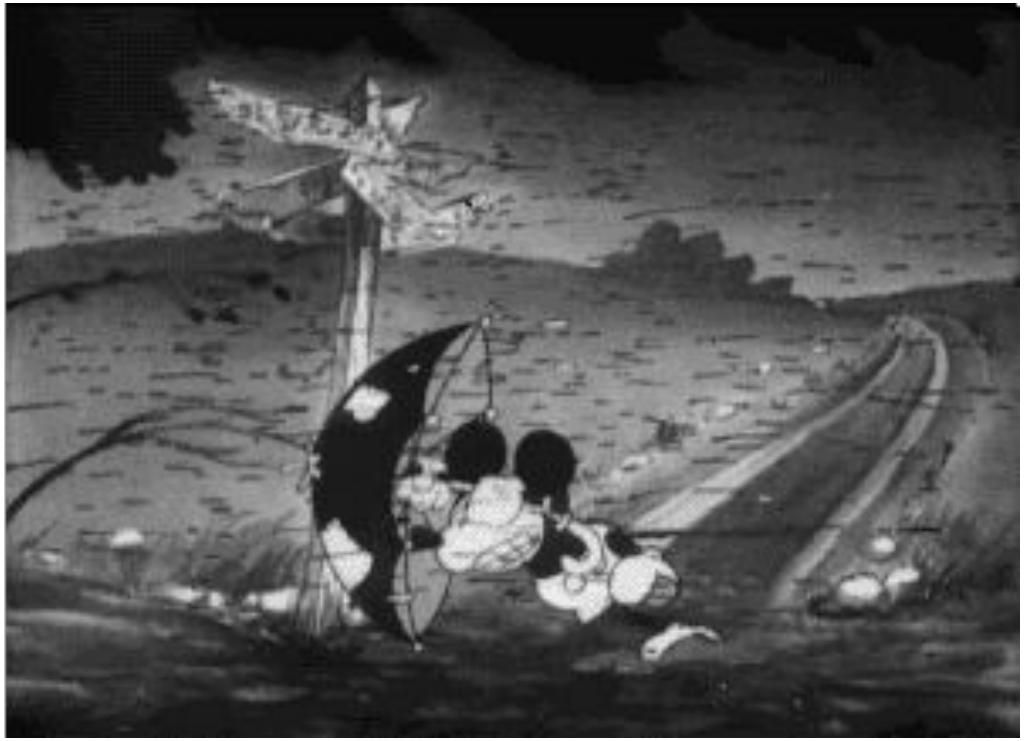
HelloExoWorld

Looking for exoplanets in NASA datasets

HelloExoWorld

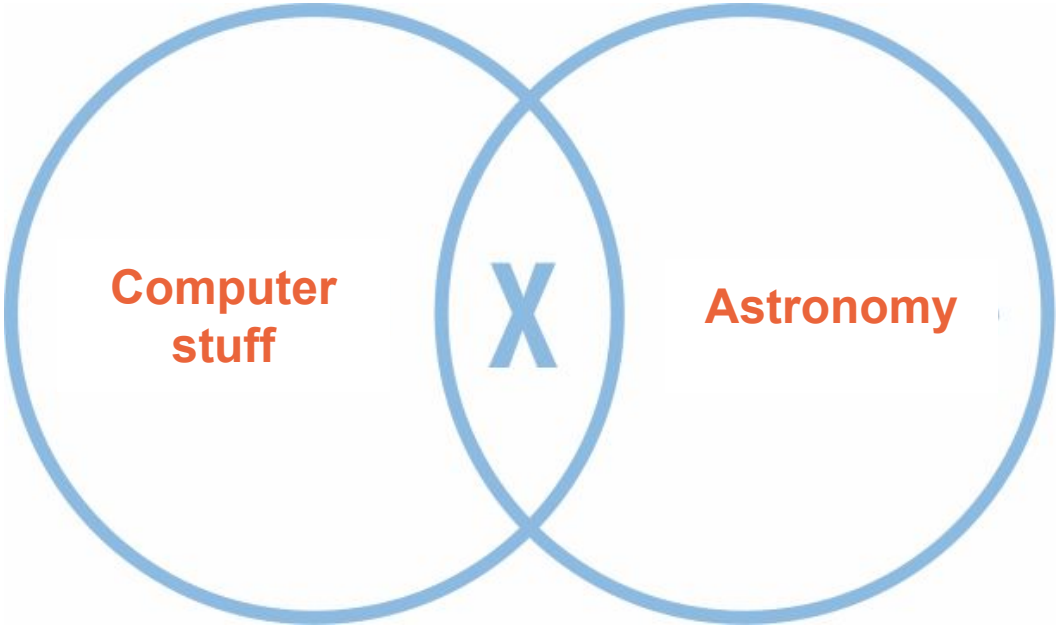
Once upon a time...

What not to do if you love astronomy



Live in Brest

Looking for solutions

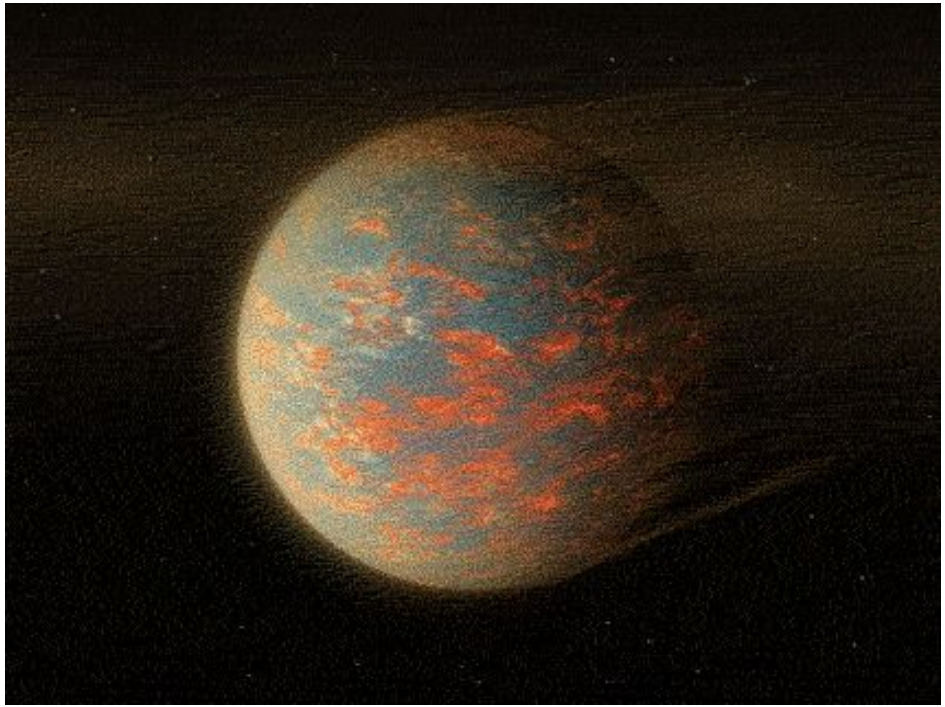


Mixing passions

Google is your friend...

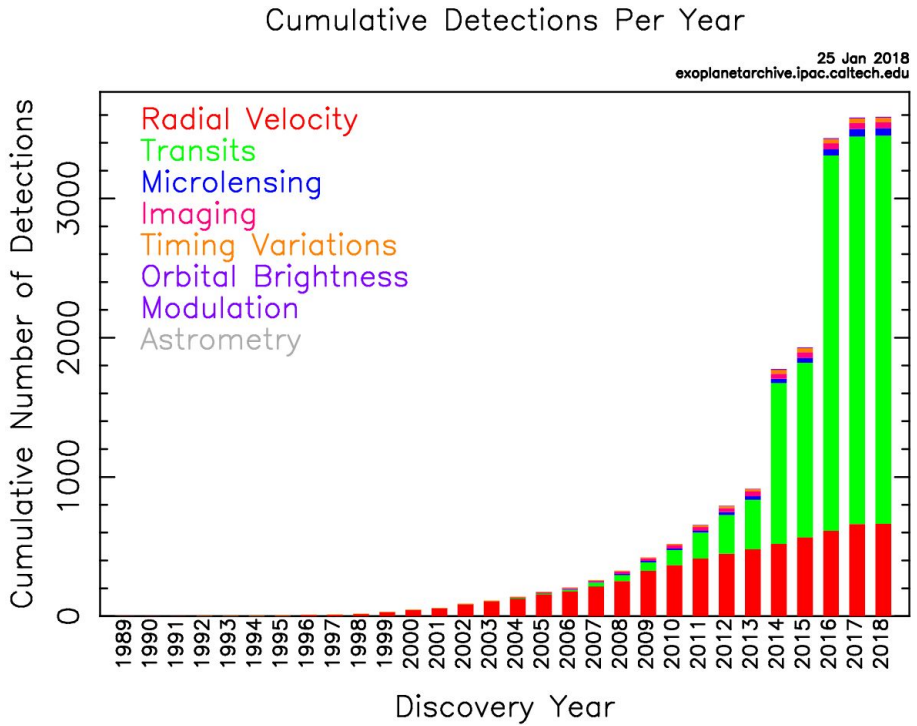
Let's find a project

Exoplanets?



Planets orbiting stars far away

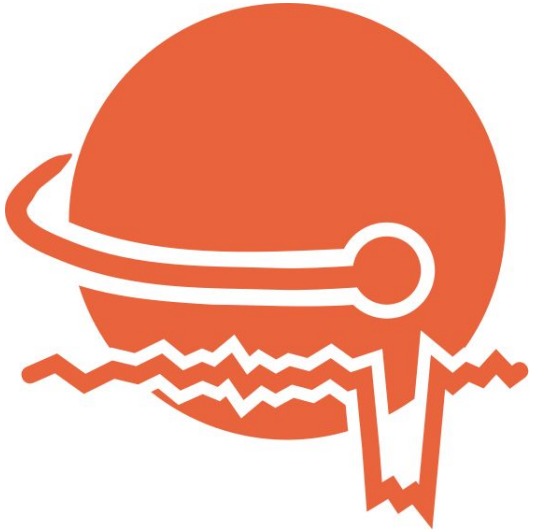
How do we find them?



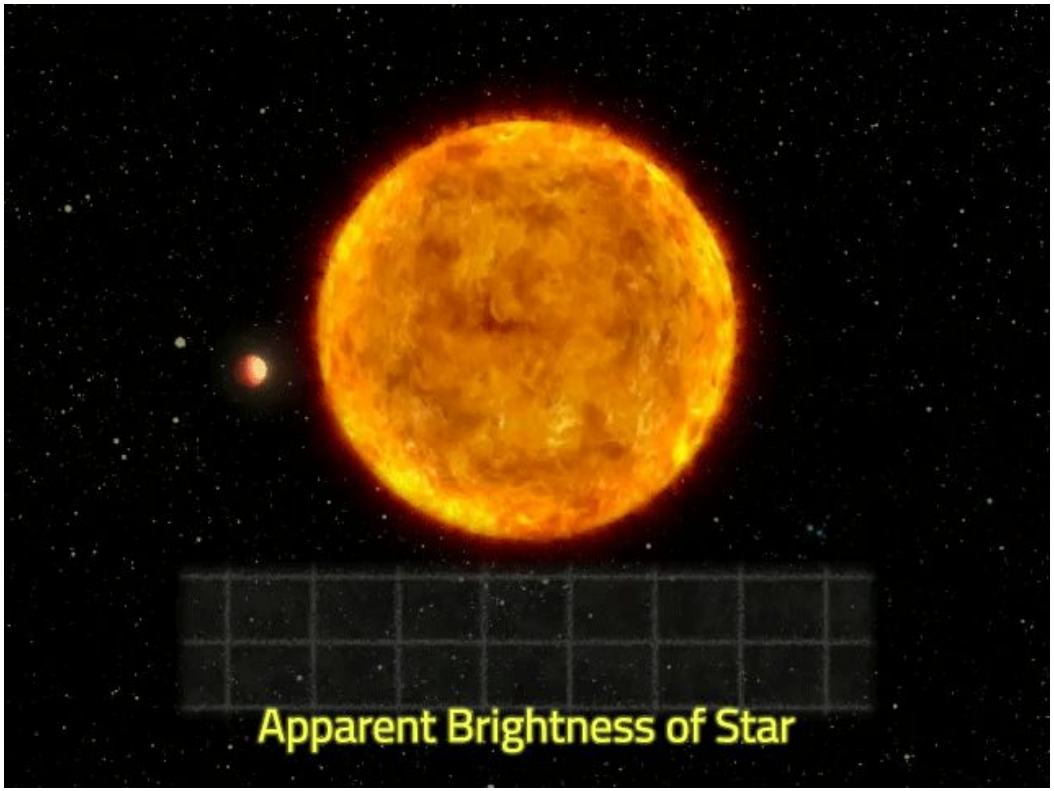
The transit method seems the best

Exoplanets detection

From theory to practice



The transit method



Credits: NASA's Goddard Space Flight Center

How do we look for transits?

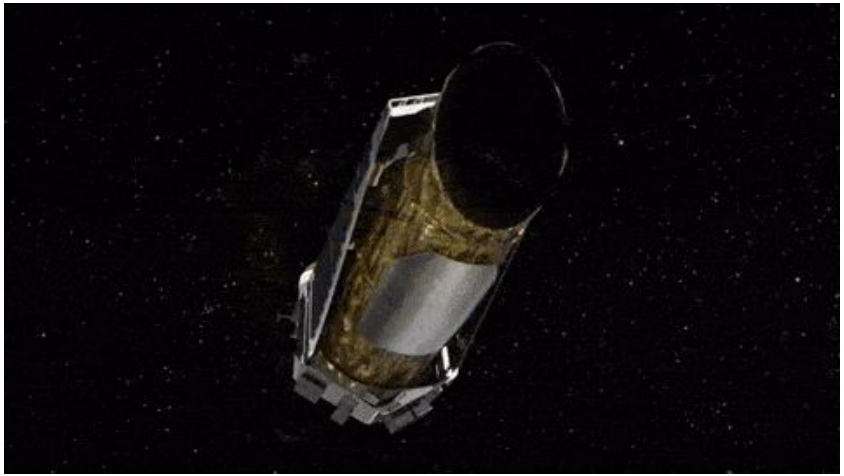


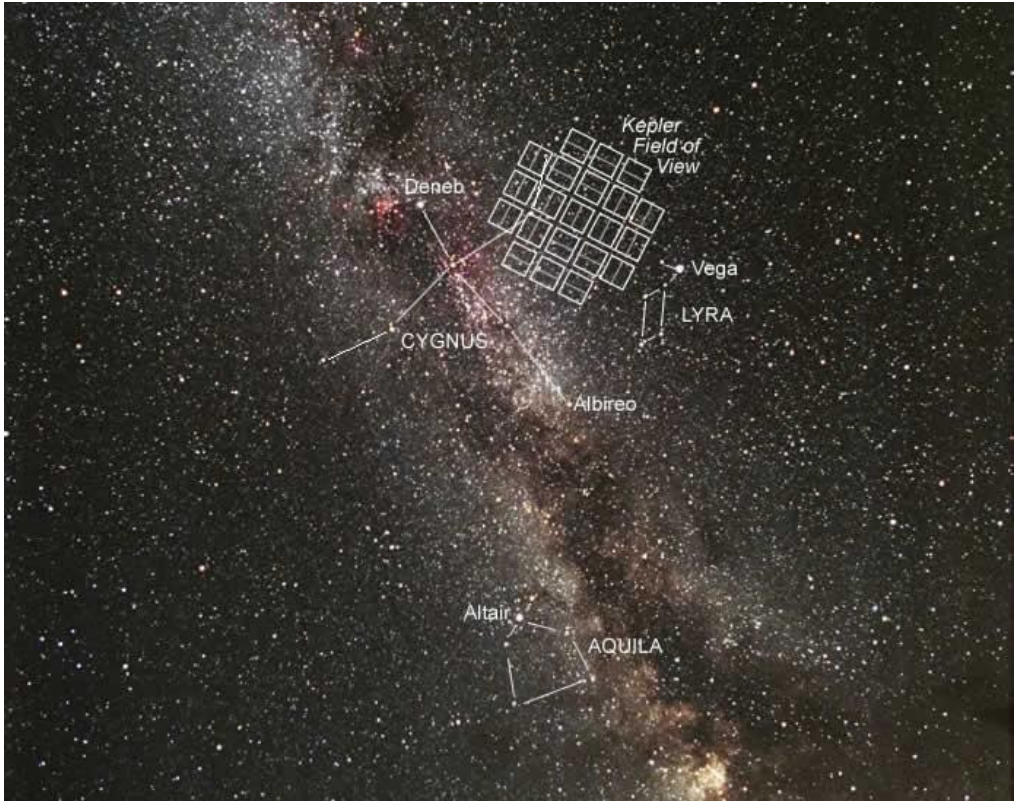
Image credits : NASA

Kepler

Image credits : NASA

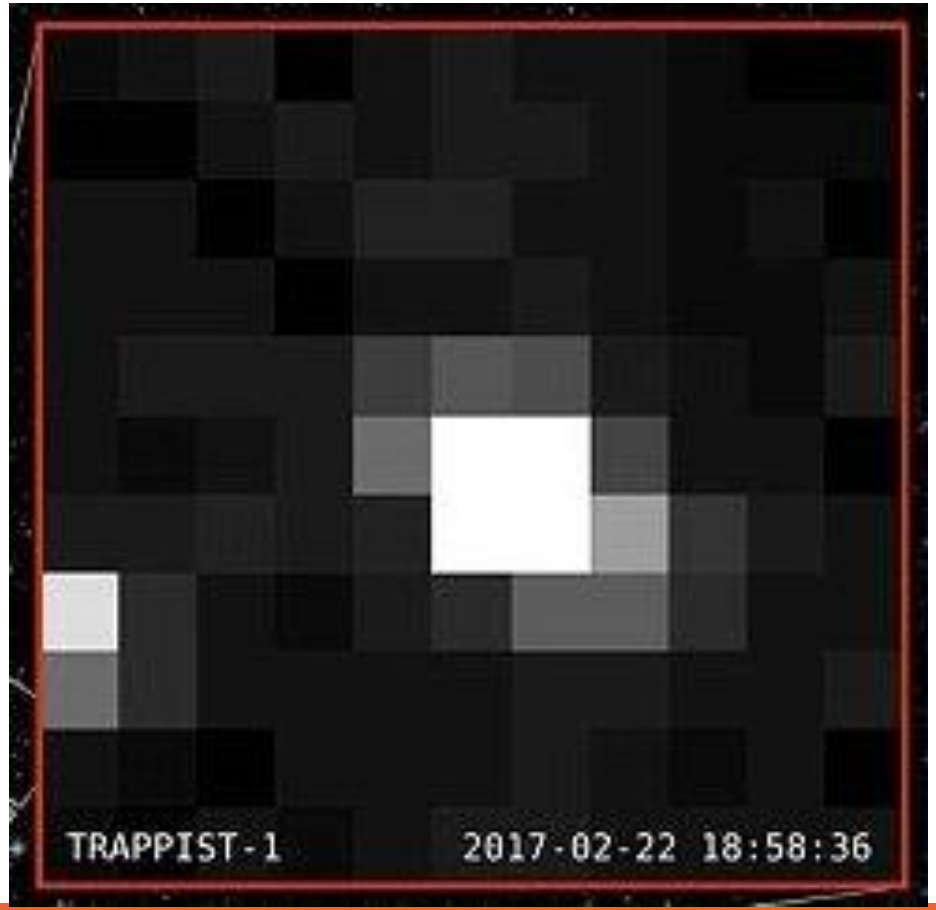
Tess

Watching the sky

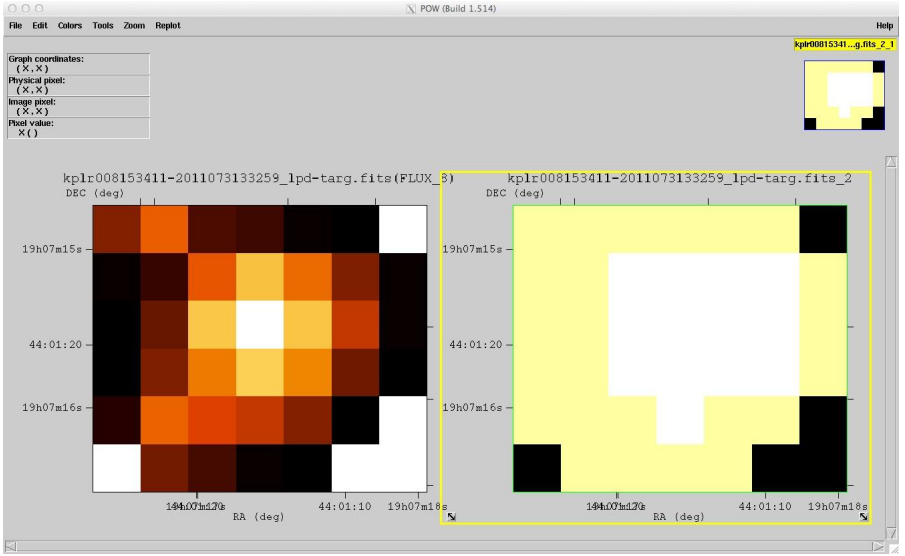


By Carter Roberts [Public domain], via Wikimedia Commons

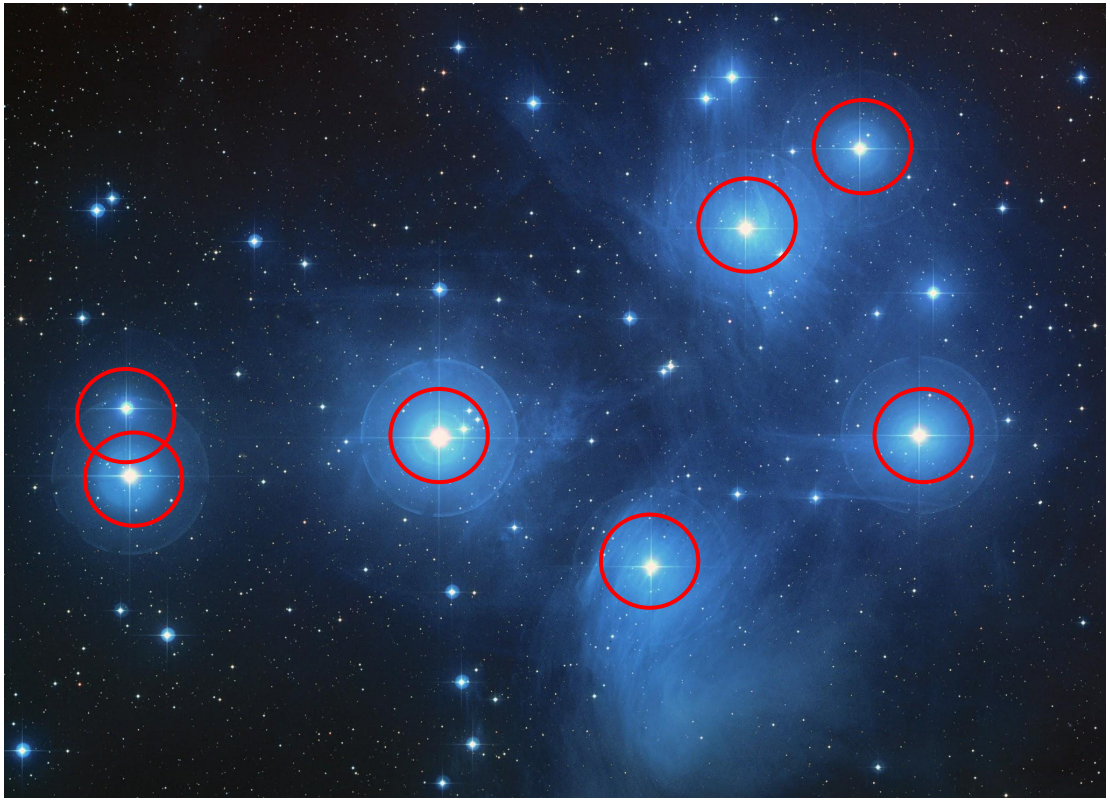
Kepler image



A star : 12*12px

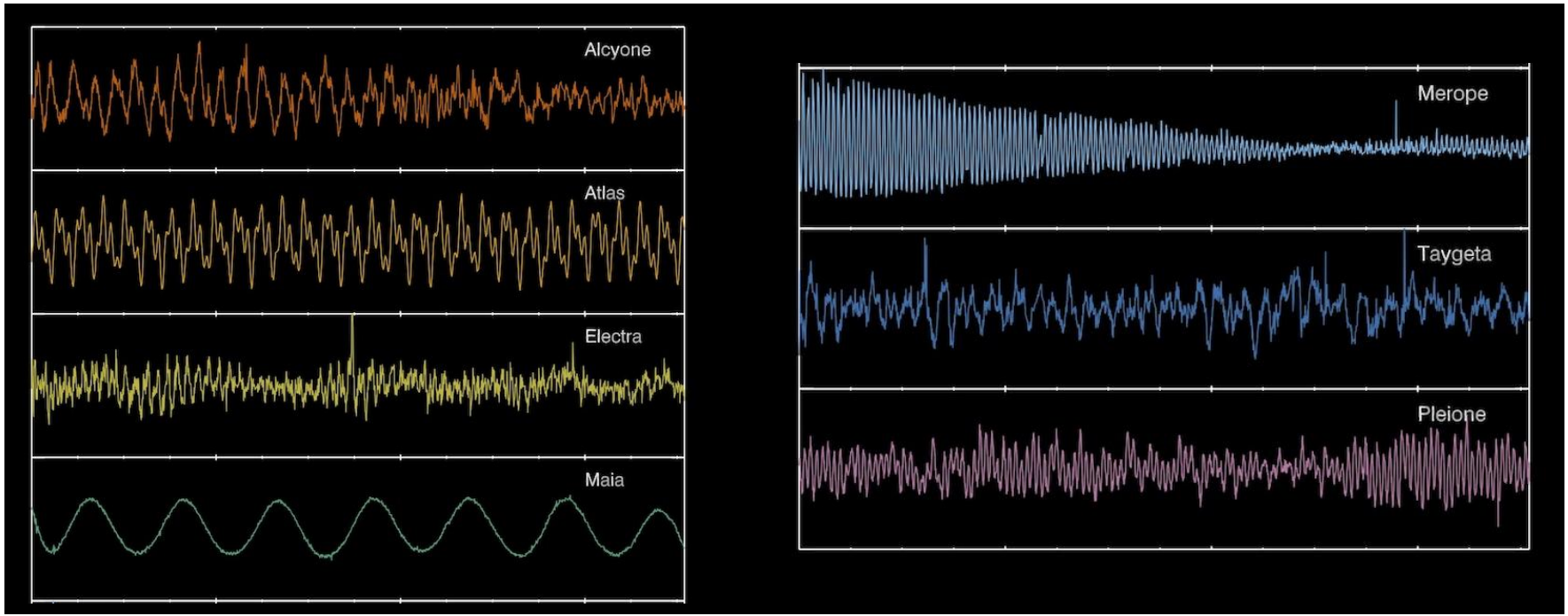


And what kind of data we get?



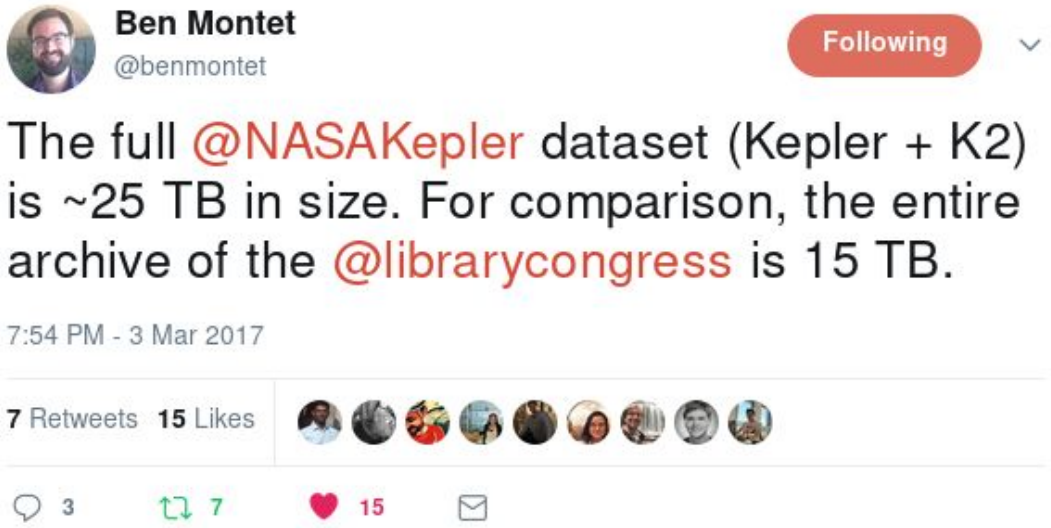
Pleiades By NASA, ESA, AURA/Caltech, Palomar Observatory. Via Wikimedia Common

Well, that's the problem



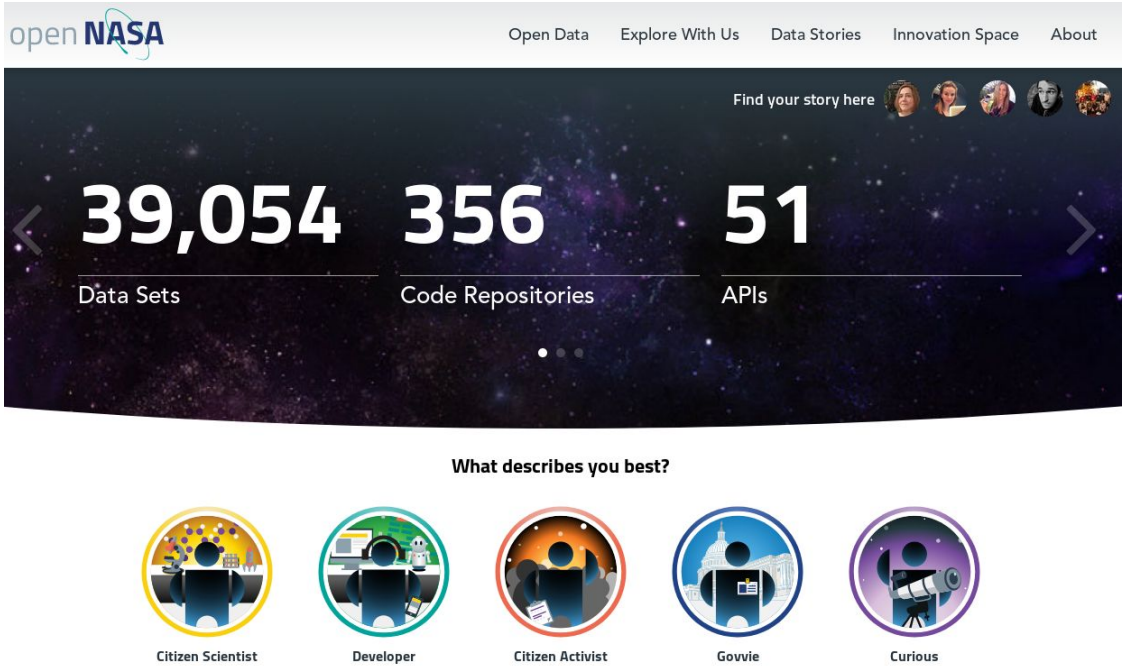
Seven stars, seven different profiles

Kinda big data



Over 40 million light curves

Big AND open data



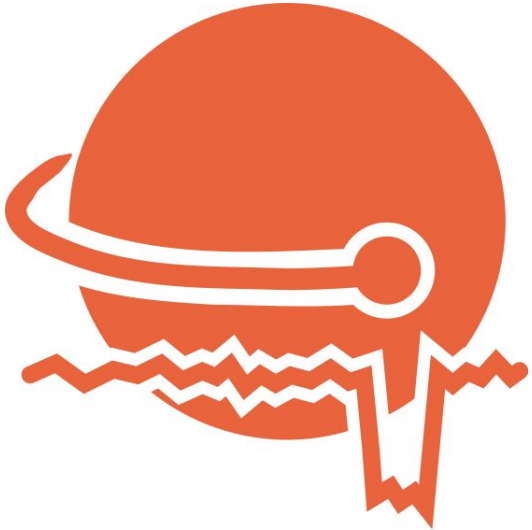
Lots of datasets in #opendata

And we can help with that!

Let's use our tools to analyse the data

Time Series

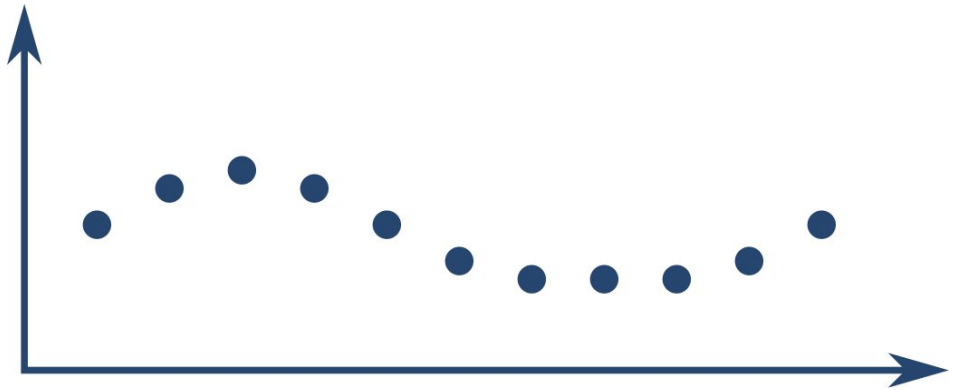
To analyse Kepler datasets



Kepler: spatial Time Series

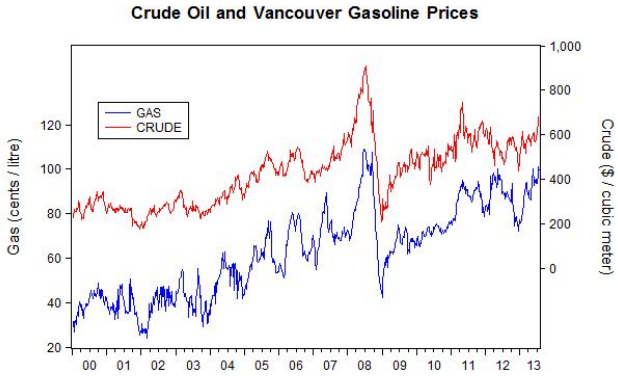
Definition of Time Series:

A series of data points indexed in time order



Time Series

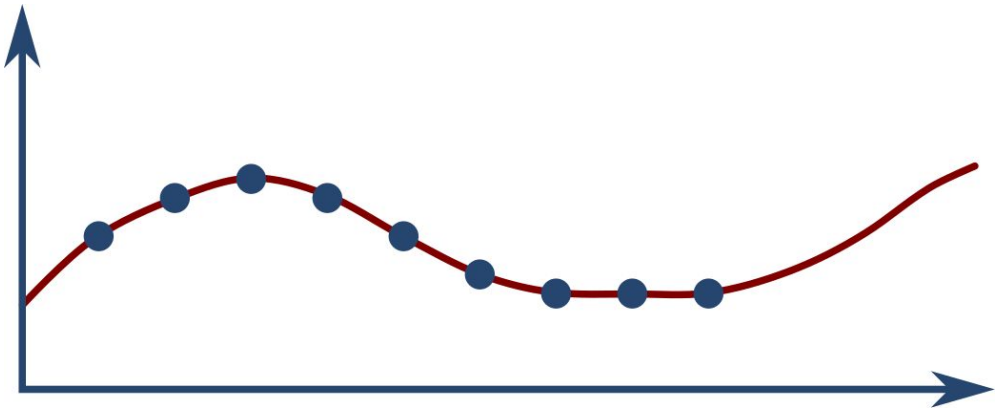
- Stock Market Analysis
- Economic Forecasting
- Budgetary Analysis
- Process and Quality Control
- Workload Projections
- Census Analysis
- ...



Time Series

Applications:

- Understanding the data
- Fit a model
  - Monitoring
  - Forecasting



Time Series

Stock market Analytics

Economic Forecasting

$$$

Study & Research

Time Series

Many specific analytical tools:

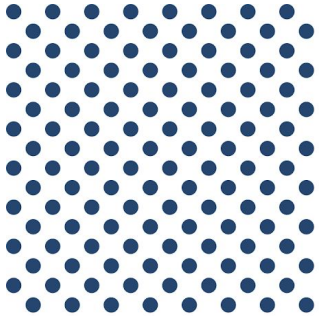
- Moving average
- ARMA (AutoRegressive Moving Average)
- Multivariate ARMA models
- ARCH (AutoRegressive Conditional Heteroscedasticity)
- Dynamic time warping
- ...

Time Series

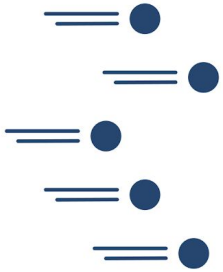
Specific application of general tools

- Artificial neural networks
- Hidden Markov model
- Fourier & Wavelets transforms
- Entropy encoding
- ...

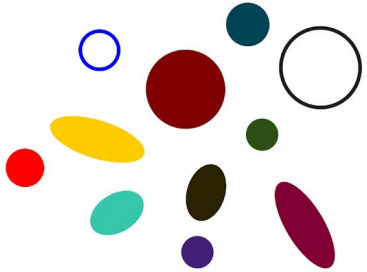
Dealing with Time Series



Volume



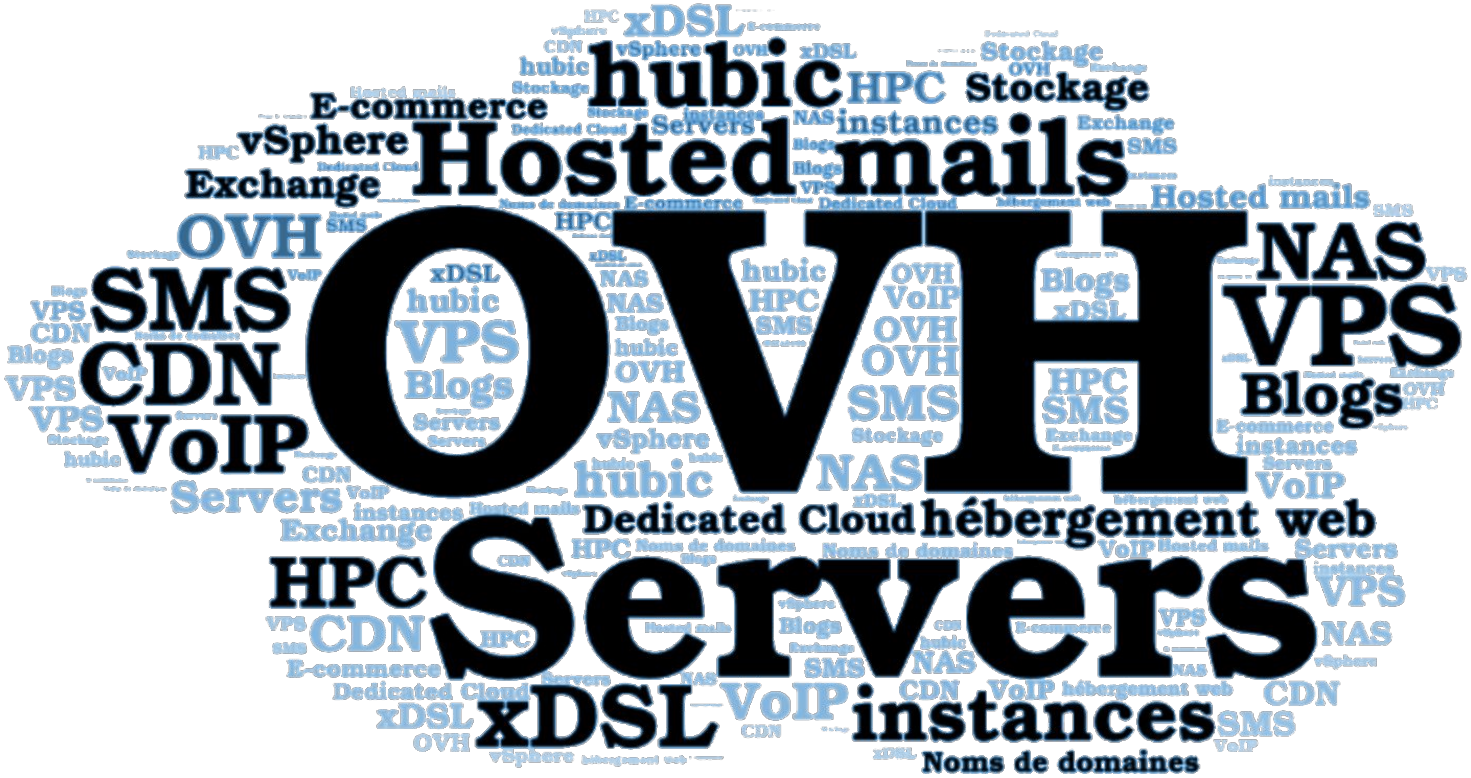
Velocity



Variety

The 3 'v'

Monitoring OVH with Time Series



OVH Metrics

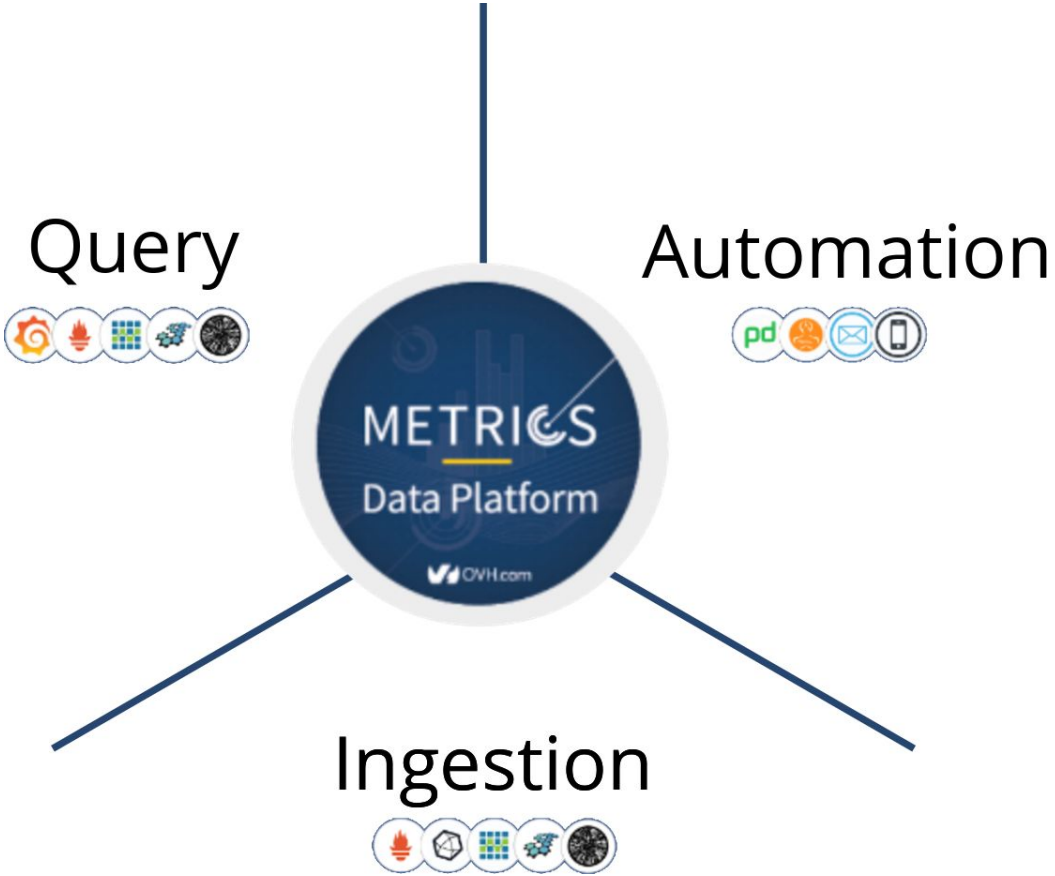
A metrics data platform

METRICS

Tools to deal with Time Series

Many options

Metrics Data Platform



Metrics' metrics

- 1.5M datapoints/s, 24/7
- Peaks at ~10M datapoints/s
- 500M unique series

Metrics Data Platform

Why Warp 10?

Warp 10 is a software platform that

- Ingests and stores time series
- Manipulates and analyzes time series

Analytics is the key to success

Fetching data is only the tip of the iceberg

Manipulating Time Series with Warp 10

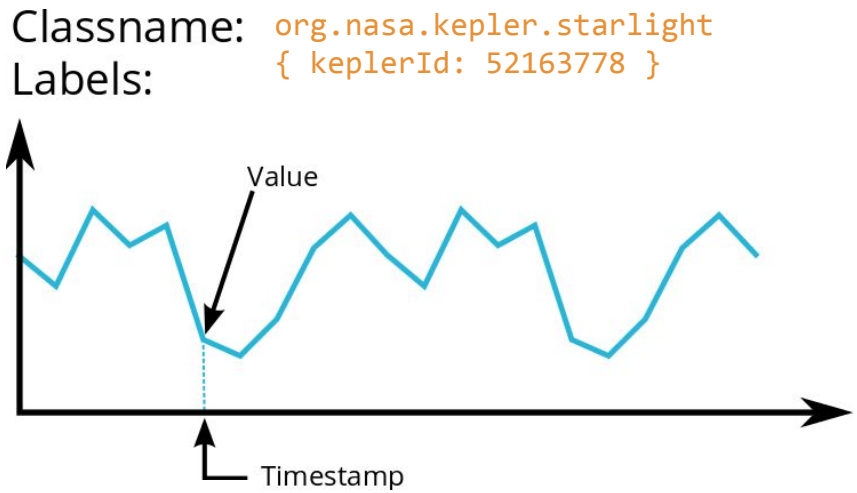
A true Time Series analysis toolbox

- Hundreds of functions
- Manipulation frameworks
- Analysis workflow

Anatomy of a time series

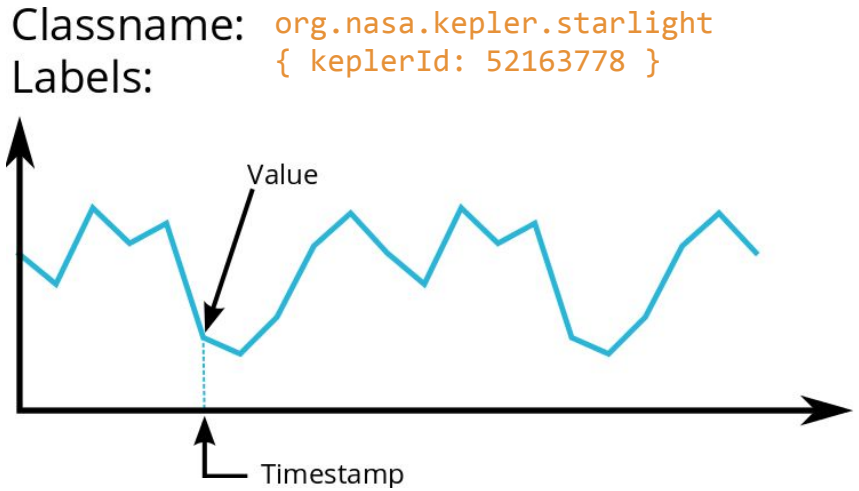
Each time series is composed of:

- Metadata
  - Class name
  - Labels
- Datapoints
  - Timestamp
  - Value



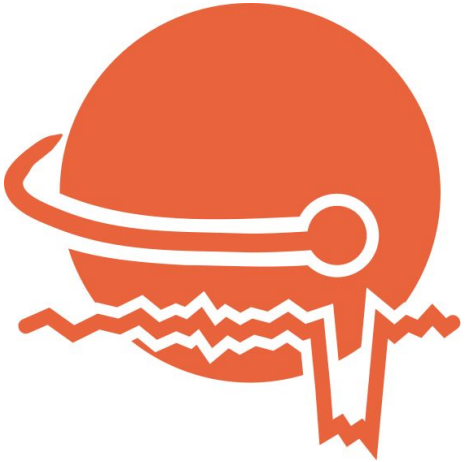
Class names and labels

- Class names define the kind of measure
  - Starlight, heart rate, speed…
- Labels define particular traits of a TS
  - Device Id, Device Type, Private User Id...



A match made in heaven

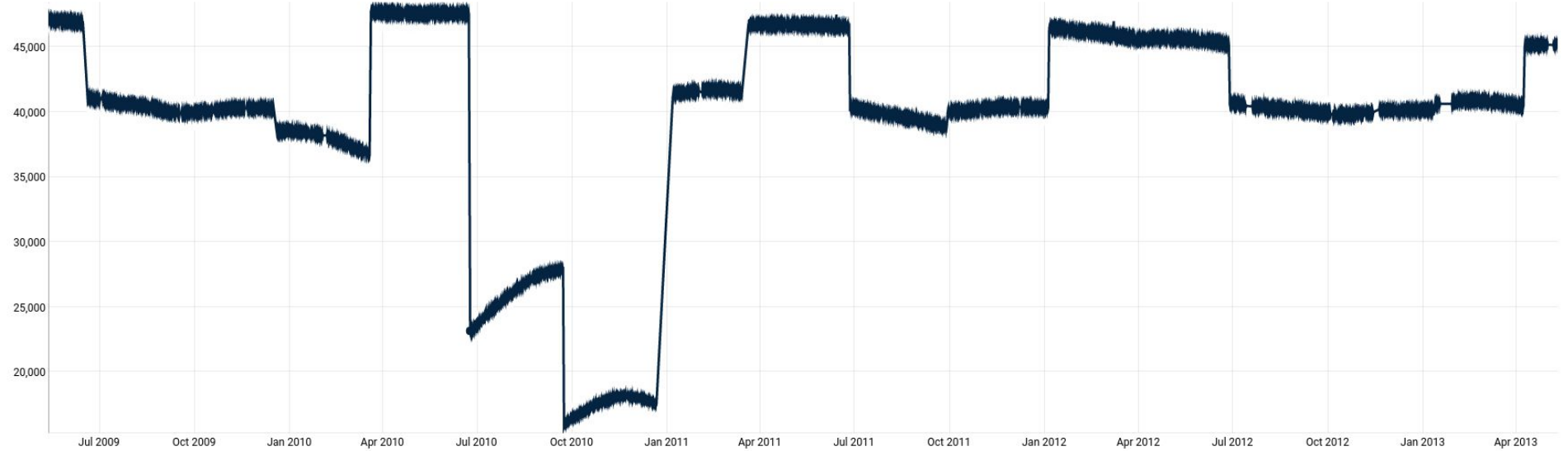
Warp 10, OVH Metrics and HelloExoWorld



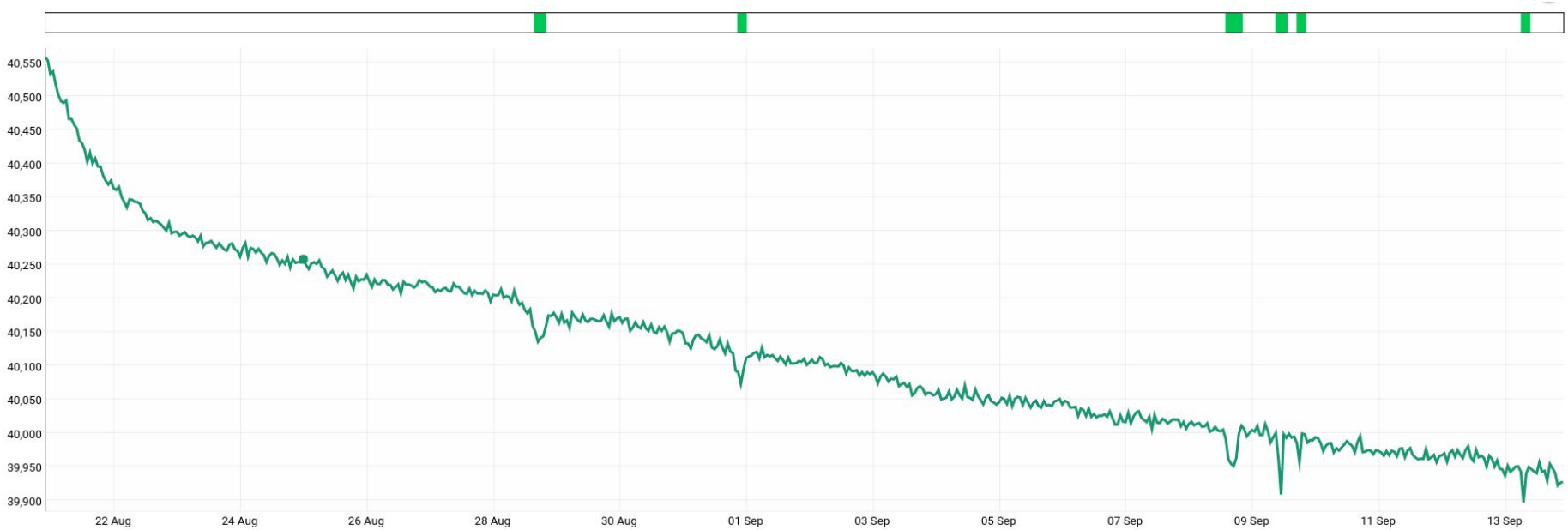
What we have done

- Downloaded and parsed 40 millions of FITS files
- Pushed it to OVH Metrics
- Select a cool subset as training set
- Verified we could find the same planets as NASA

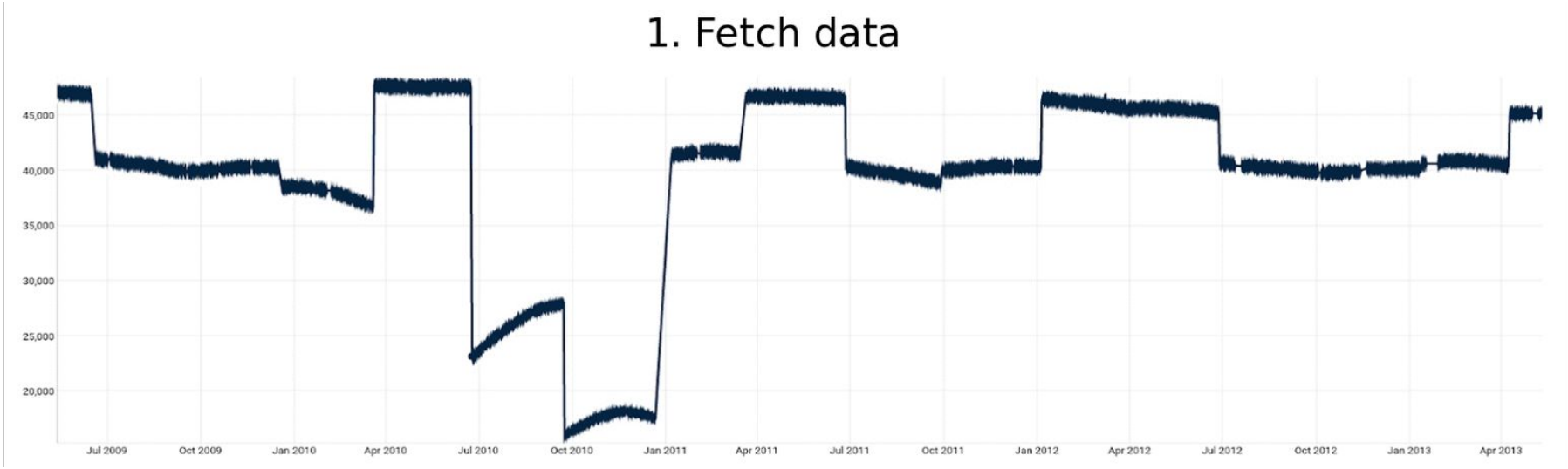
From kepler-11 raw data



To (candidates) exoplanets



Your job


1. Fetch data
45,000
40,000
35,000
30,000
25,000
20,000
Jul 2009
Oct 2009
Jan 2010
Apr 2010
Jul 2010
Oct 2010
Jan 2011
Apr 2011
Jul 2011
Oct 2011
Jan 2012
Apr 2012
Jul 2012
Oct 2012
Jan 2013
Apr 2013

Let's get started!

1. Connect to https://bit.ly/2H7Z5b3

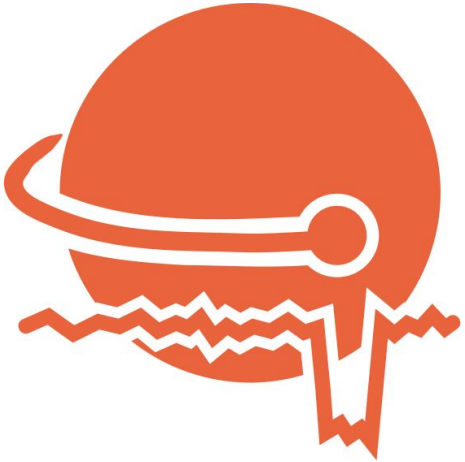
   or

   Connect to WIFI HEW-5G (or HEW)
2. Password is helloexoworld
3. Click on cancel on user password window
4. Open chrome/chromium on 192.168.1.2

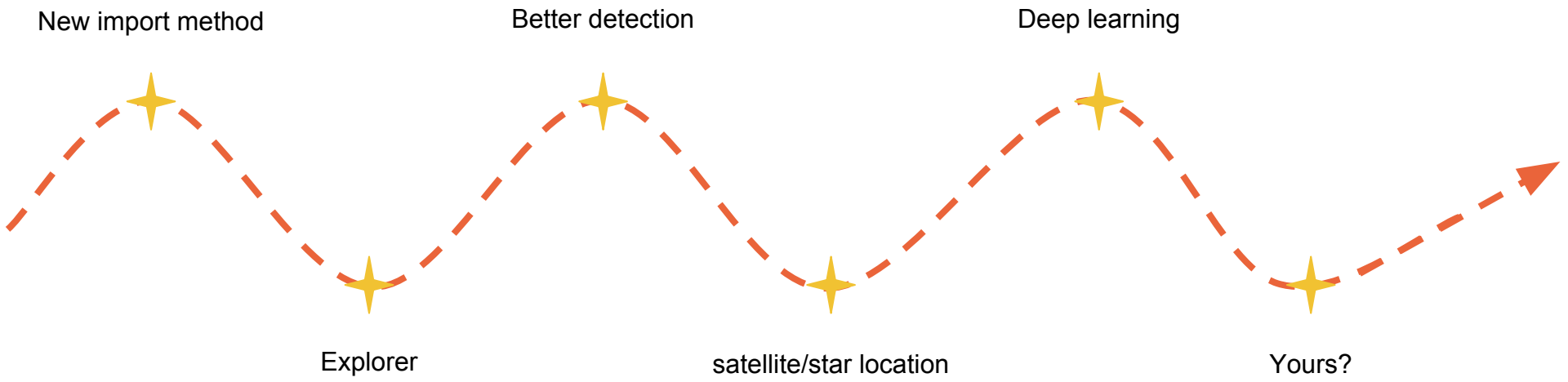
Reach step 3.2 and enjoy!

What's next?

Where do we go from here?

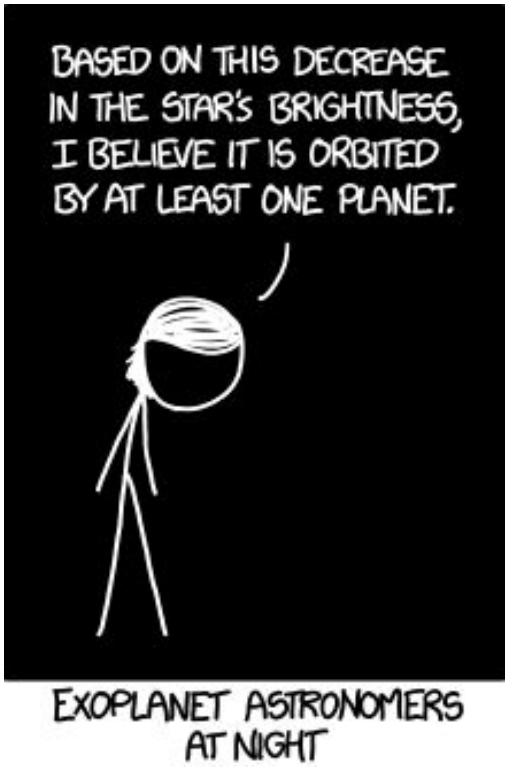


Only the beginning



A growing team

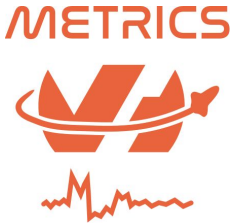
And you!



https://xkcd.com/1371/

Join us!

https://helloexo.world



OVH Metrics

Come speak with us about your time-series or monitoring projects and OVH Metrics

Thank you, dear sponsors!

OVH

clever cloud

HELLO EXO.WORLD

Thank you!


thank you
danke
рахмат
謝謝
ngiyabonga
شكراً جزيلاً
teşekkür ederim
gracias
merci
obrigado
dziękuję
bedankt
спасибо
Баярлалаа
go raibh maith agat
mochchakkeram
sukriya
kop khun krap
grazie
arigatô
dakujem
terima kasih
감사합니다
ありがとう
ευχαριστώ
мерси
tapadh leat
dank je
mersi
kia ora
barka
welalin
tack
spas
vinaka
спасибі
blagodaram
misaotra
matondo
paldies
grazzi
mahalo
хвала
asante
manana
obrigada
tenki
murakoze
djiere dieuf
tau
дякую
mamnun
chokrane
faafetai lava
kiitos
dankie
nanni
nandri
dhanyavad
bayarlalaa
gracje
hvala
maururu
köszönöm
enkosi
chnorakaloutioun
gratias ago
gràcies
sulpáy
akun dankon
taiku
sobodi
děkuji
mèsi
sagolun
didi madloba
najis tuke
kam sah hamnida
rahmat
tanemirt
rahmet
xièxie
diolch
dhanyavadagalu
shukriya
takk
trugarez
mercé