Meta data information

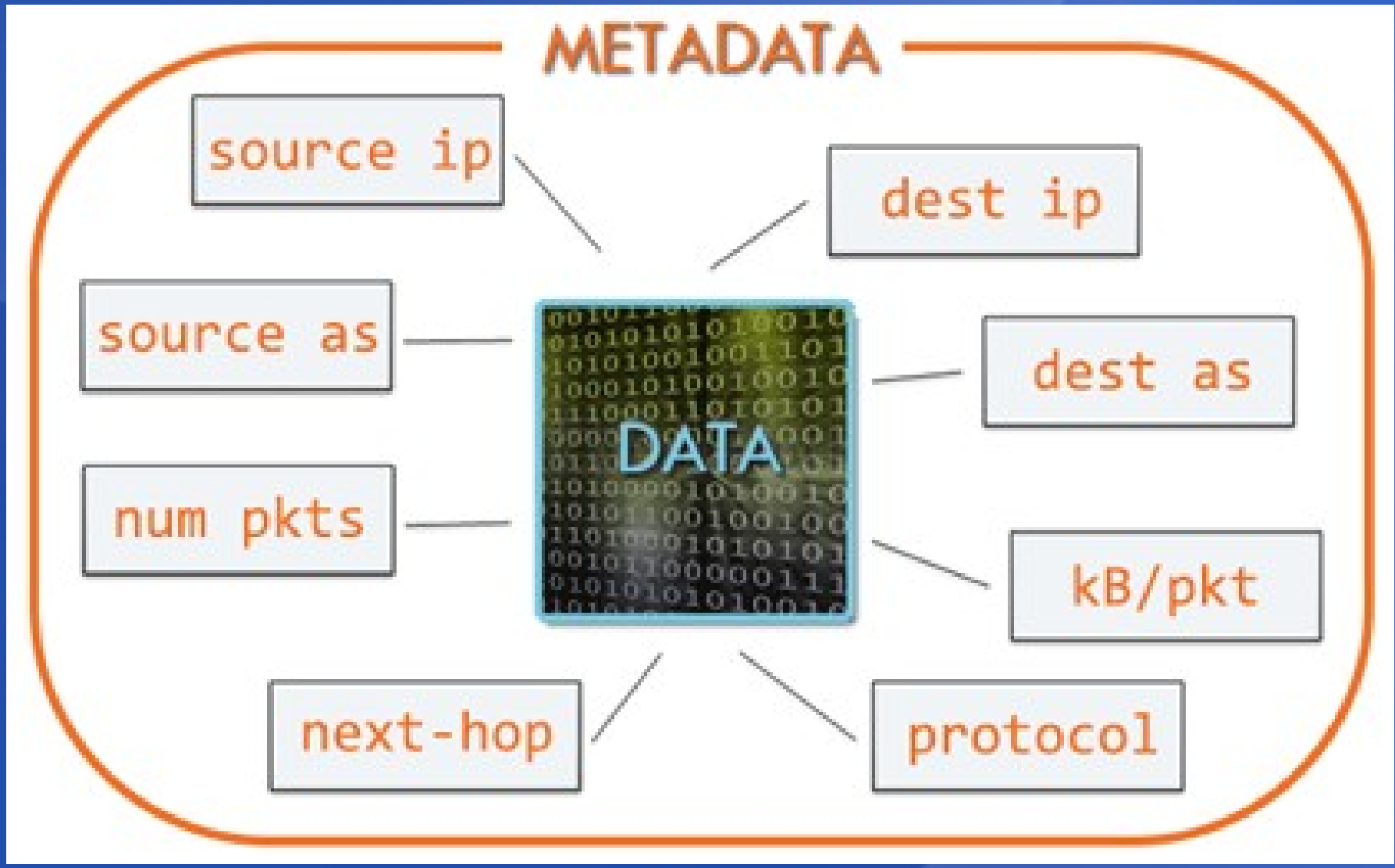
A.Kryukov, A.Demichev

SINP MSU

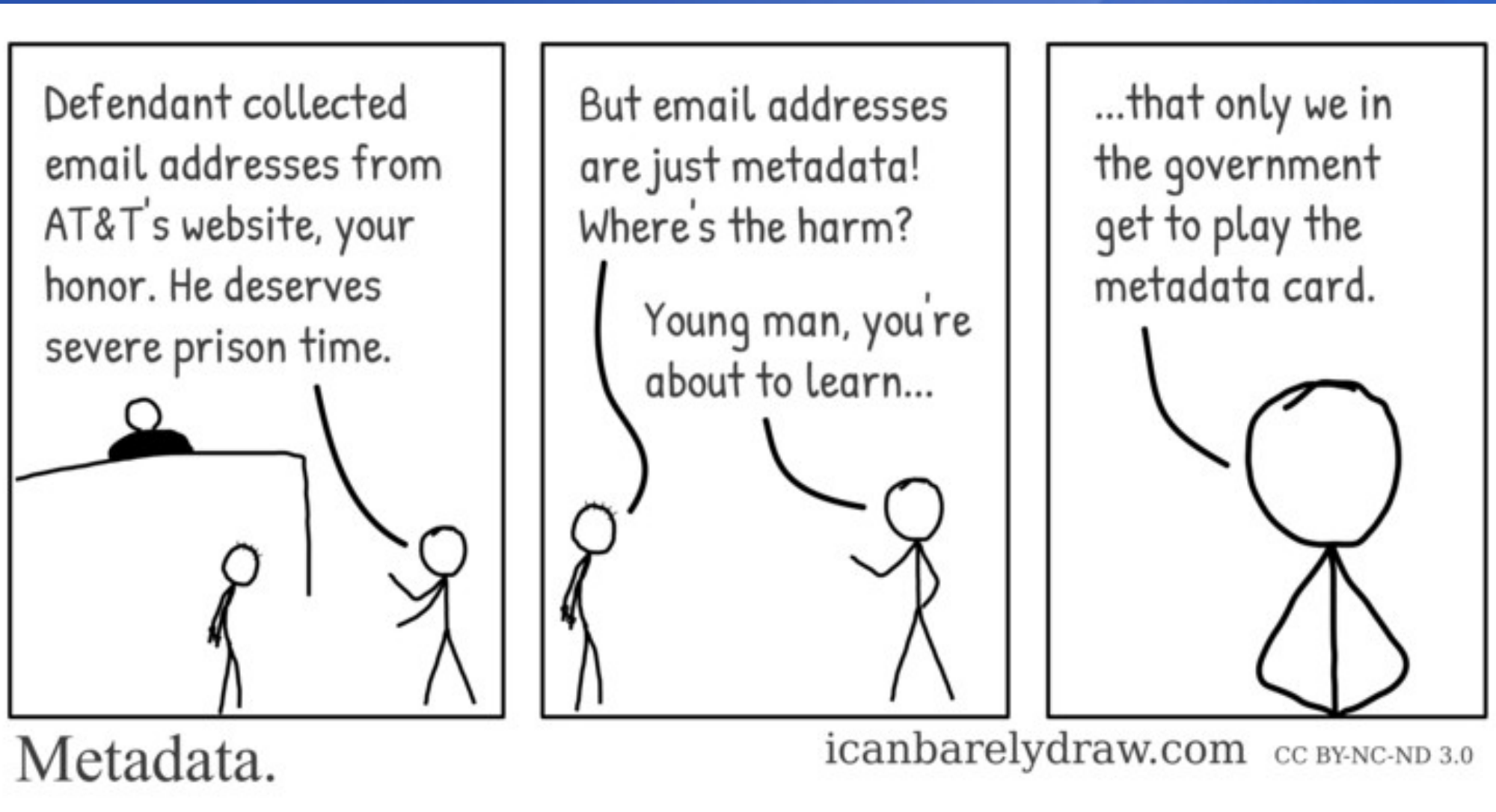
Main targets

- In modern astroparticle facilities that generate scientific data, there is no general approach to the formation and management of metadata. This makes it difficult to effectively and correctly organize a work with data throughout their life cycle, including data processing and analysis, to exchange data between different research groups and organizations, to integrate data and to ensure that analysis programs always comply with all characteristics of the data being processed, including methods, conditions and also the characteristics of the instruments with which they were obtained.
- Thus, metadata describing data, provide context and are vital for the accurate interpretation and use of data by both humans and machines.

Metadata


METADATA
source ip
dest ip
source as
dest as
DATA
num pkts
kB/pkt
next-hop
protocol

Metadata


Defendant collected email addresses from AT&T's website, your honor. He deserves severe prison time.
But email addresses are just metadata! Where's the harm?
Young man, you're about to learn...
...that only we in the government get to play the metadata card.
Metadata.
icanbarelydraw.com
CC BY-NC-ND 3.0

What is Metadata

- MD is a data which describes the data
  - Do not confuse with data format!

What is Metadata

- MD is a data which describes the data
  - Do not confuse with data format!
- Example1: the name of data files(IACT):
  - DATA_IACT/2016-17/dec16/031216/ BSM03/03126003.001

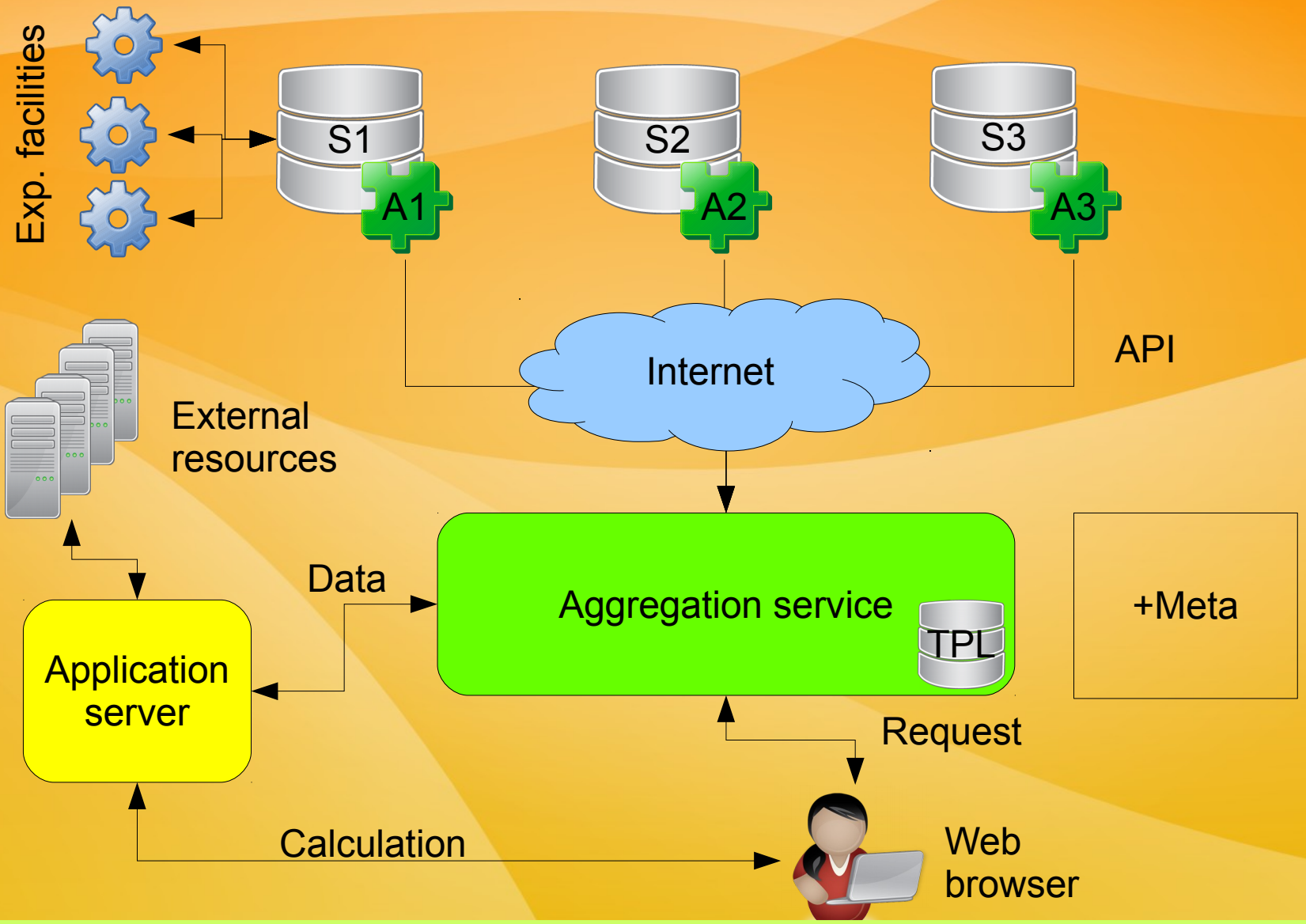
What is Metadata

- MD is a data which describes the data
  - Do not confuse with data format!
- Example1: the name of data files(IACT):
  - DATA_IACT/2016-17/dec16/031216/BSM03/03126003.001
- Example2: MC for IACT
  - Data file: ev-1-17239.txt
  - Event properties: pr-1-17239.txt
  - Metadata: md-1-17239.txt

Example (MC)

Data	1 0.000000 0.000000 0.000000 …. 0.000000 -1.662764 -2.795241
Proparties	[2] part=14 E[TeV]=4.817096 R[m]=172.847578 Theta[deg]=35.286495 Phi[deg]=79.211179 nnShower=1 nnScatter=8 SimulFileName=bpe2013_22_da5.0 MirrMisalign=0
Metadata	[coordinates] X[mm]=-15 0 -30 -45 -30 ... Y[mm]=-77.94228634059948 -103.92304845413263 ... [events] nn1=1 nnN=17239 date=03-Mar-2018

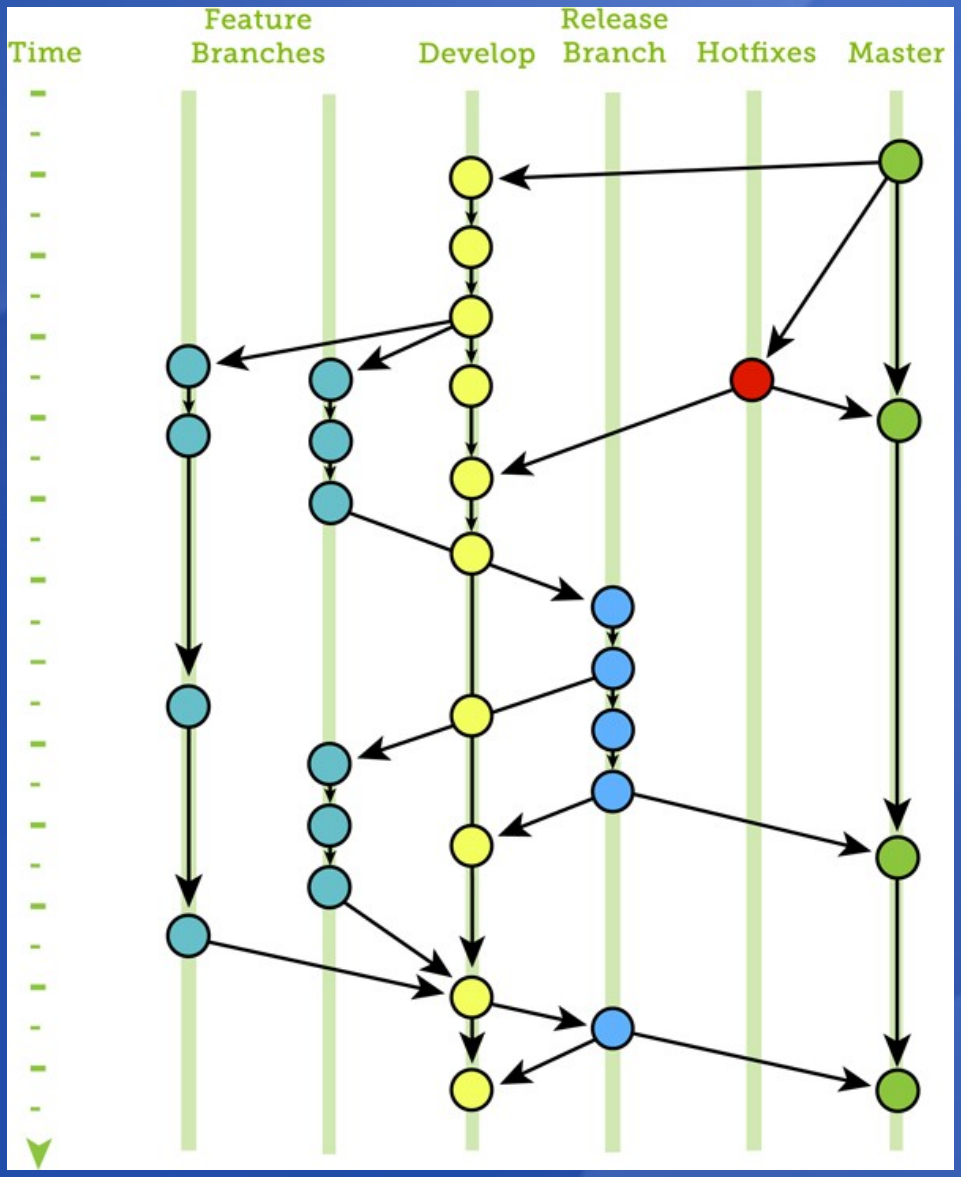
Architecture of APPDS (draft)



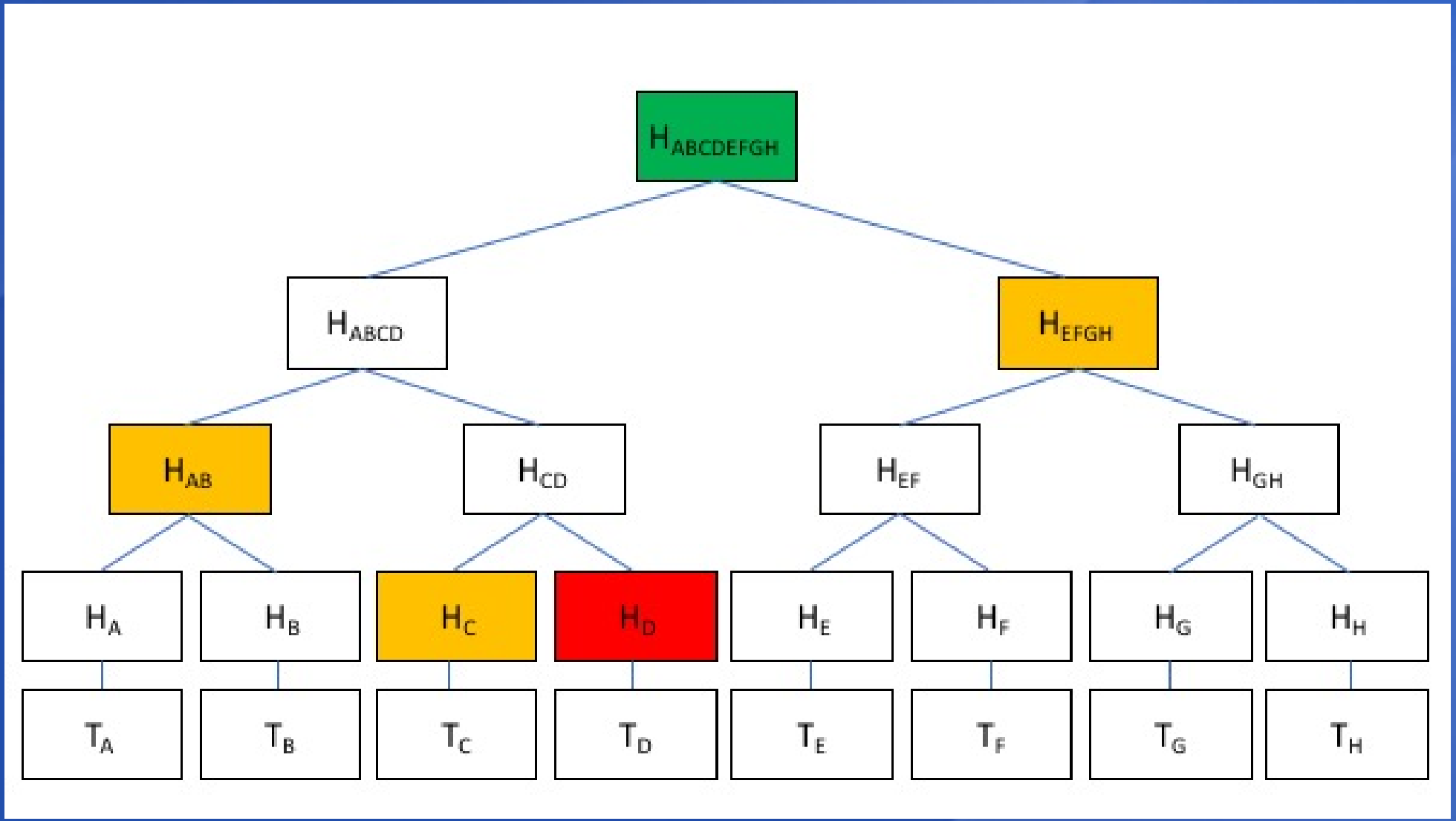
Provenance Data

- One of the important types of metadata is provenance metadata.
- Provenance is a meta-information related to the history of obtaining data, starting from the source. Metadata of this type is designed to track the steps at which data were obtained, their origin, their proper storage, reproduction, for interpretation and confirmation of the scientific results obtained on their basis.
- Thus, provenance metadata (PMD) are important for organizing a correct research workflow to obtain reliable results.

Provenance Data



Merkle (Hash) Tree



Reasons of provenance data

- Changes in the format of sensor data
- Accuracy and quality can occur for many reasons
  - changes in sensor design, refinements in calibration, or even physical displacement
- This can give rise to catastrophic errors in the processing and analysis of data. Similar consequences can have a "hidden" evolution of data processing and analysis algorithms, as well as code modification, change of versions and releases of corresponding computer programs

Sofisticated search

- Indecies over physical paramiters
  - Energy, axies
- Smart cuts
  - D.K. example:
    energy * sin(geomag_angle (71.7571, −2.7625,zenith , azimuth)) >=0.1;

Sofisticated search: problems

- File storage vs. Data base
  - NOSQL DB: MongoDB, SciDB
- Storage intermediate data
-

Thank you