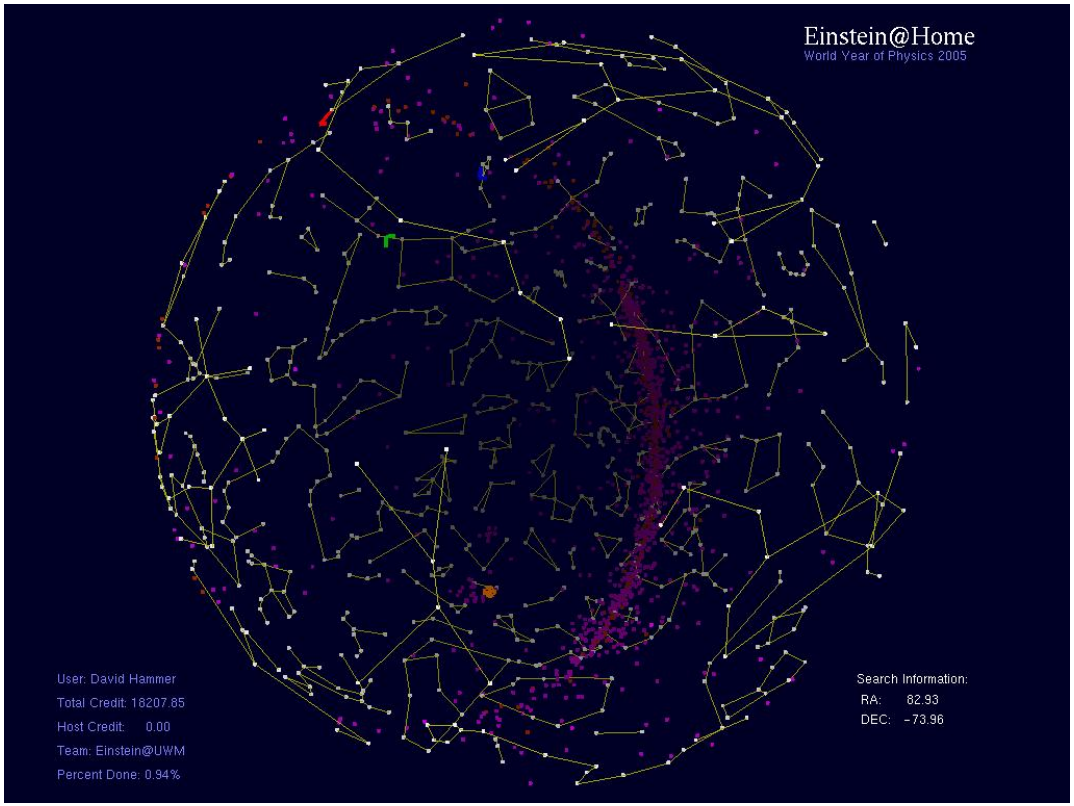
Managing scientific data in the Einstein@home volunteer computing project

Christian Beer & Avneet Singh
Albert Einstein Institute (AEI) Hannover

Einstein@Home in a nutshell

Volunteers contribute computing power to help search for
Pulsars and Gravitational Waves



HELMHOLTZ GEMEINSCHAFT
Open Science

Albert Einstein Institute

Einstein@Home GW searches

- 11 major GW searches performed on Einstein@Home
  - S3, S4, Early S5, S5 Hough, Full S5, S5 Galactic centre, S5GC1, S5GC1HF, S6Bucket, S6CasA, O1AS20-100* (*ongoing), looking for isolated compact objects (e.g. neutron stars etc)

- Why?
  - For e.g. S5GC1HF search:
    - without Einstein@Home = 5,000 years on standard laptop
    - with Einstein@Home = 6 months
  - For e.g. S6Bucket search:
    - without Einstein@Home = 10,000 years on standard laptop
    - with Einstein@Home = 1 year

Einstein@Home GW searches

- Trivia:
  - S5GC1HF searched for a whooping 10^{17} templates! This equates to 10^{17} comparisons performed over 6 months.

- Work division (S5GC1HF)
  - Searches usually search for frequency (1 parameter), spin-down (1 parameter) and locations (2 parameters) in sky
  - Divide the 4-parameter search ranges into very small tiles & compare all such tiles to the observed data.
  - Work-units sent to volunteers; each work-unit covers a certain range of the 4 parameters
  - Process 6.4 million work-units in parallel for 6 hours each!

Einstein@Home GW searches

- Results back: (S5GC1HF)
  - Bring back processed results from the volunteers.
  - Whooping quantity of data: 4 terabytes of data to be post-processed

- Post-processing (S5GC1HF)
  - Post-processing of 'candidates' required; roughly 10^{11} candidates released
  - We are looking for a few key candidates (1 or 2!) among the 10^{11} candidates
  - Several stages in 'post-processing'

Einstein@Home GW searches

- Did we find anything?
  - No? We are used to it! :-D
  - Yes?! Well, open the champagne! :-D

- Publish results in any case . . .
  - Publish signals, if any found. Let the chaos ensue!
  - Publish upper bounds, if nothing seen.
  - Estimate the loudness of the loudest candidate that we MOST LIKELY (with >90% confidence) would have detected, if it existed.
  - Current upper bounds put the GW strengths to be
  - $< 10^{-24}$ over 50-1500 Hz frequency range

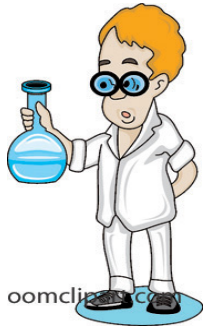
Einstein@Home GW searches

- Results till date . . .
  - No GW signals found!
  - We have searched all sky!
  - Frequencies searched between 50-1500 Hz
  - Spin-downs searched between 10^{-8} — 10^{-10}
  - Bounds of signal strengths < 10^{-24}

We keep searching . . .

The data problem

massive amount of data



High Performance Cluster takes **100 days** to process, **costs are high**

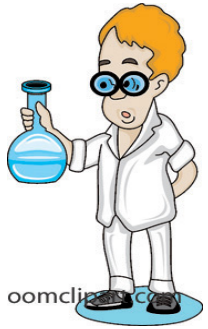
Internet-connected PCs take **100 days** to process, **costs are low**

HELMHOLTZ GEMEINSCHAFT
Open Science

Albert Einstein Institute

The data problem

massive amount of data



Splitting

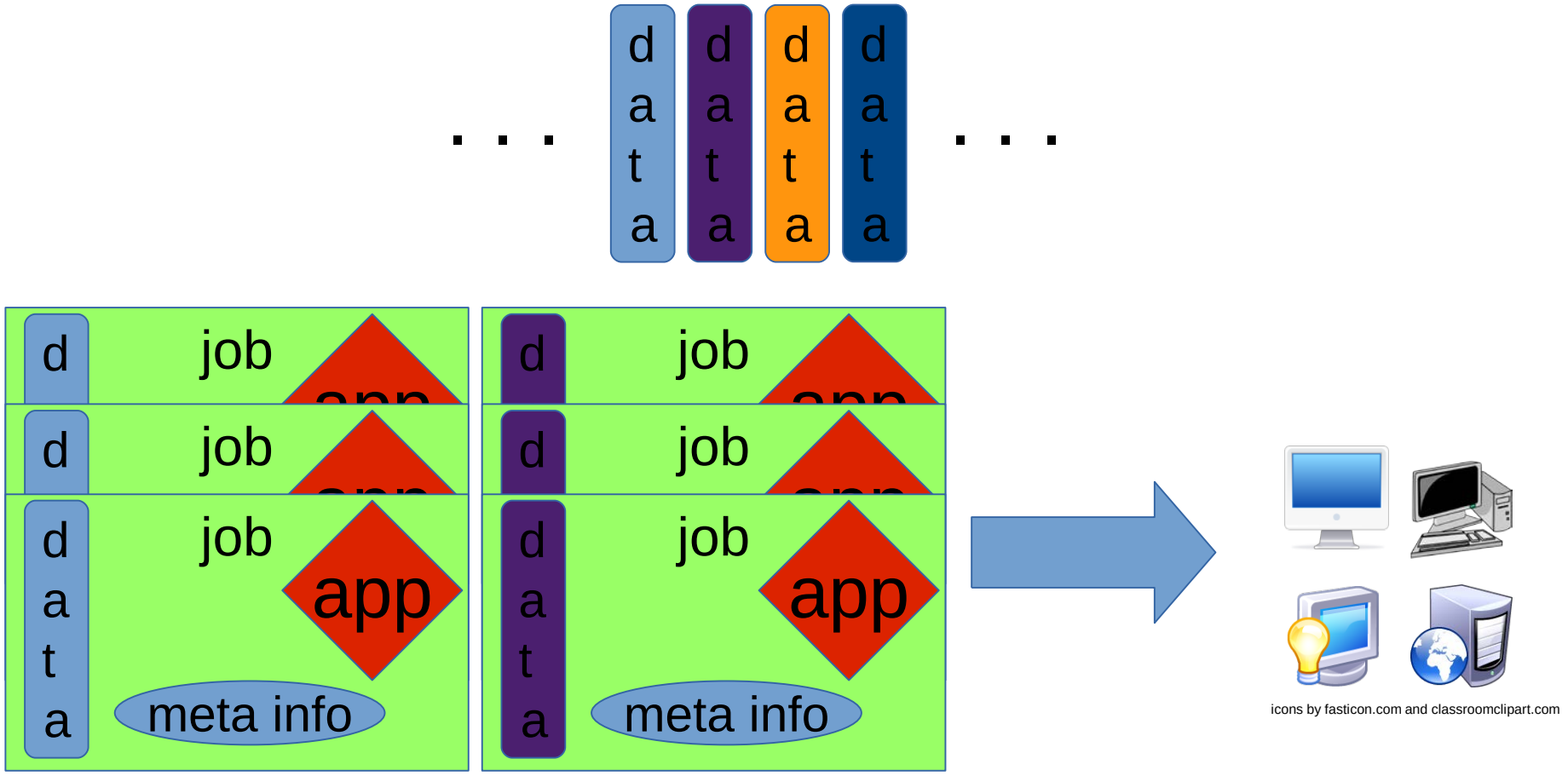
S6CasA: 190 GB
O1-All-Sky: 75 GB
(projected)

data data data data data data data data data data data data data data data data data data data data data

HELMHOLTZ GEMEINSCHAFT
Open Science

Albert Einstein Institute

The data problem



HELMHOLTZ GEMEINSCHAFT
Open Science

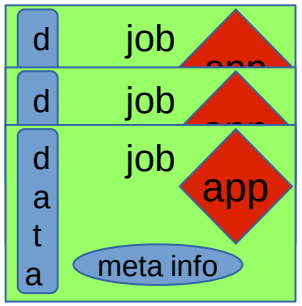
Albert Einstein Institute

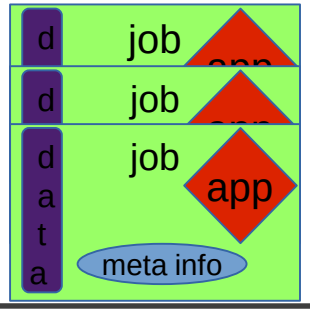
Sending data to volunteers

data

data

Einstein@home server

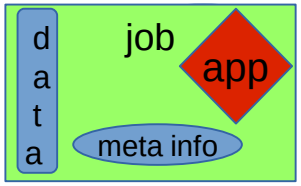




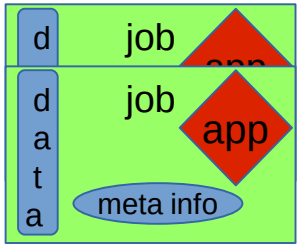
HELMHOLTZ GEMEINSCHAFT
Open Science

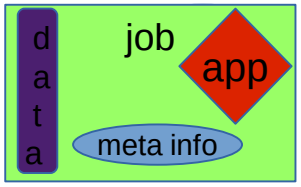
Albert Einstein Institute

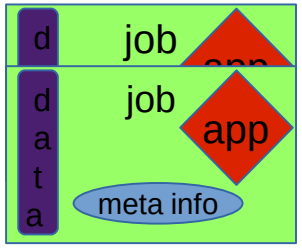
Sending data to volunteers



Einstein@home server







HELMHOLTZ GEMEINSCHAFT
Open Science

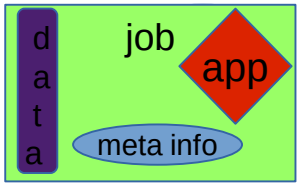
Albert Einstein Institute

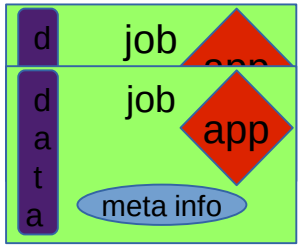
Sending data to volunteers

data

Einstein@home server

data





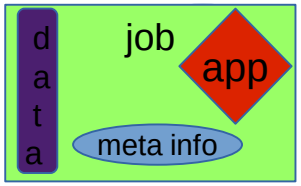
HELMHOLTZ GEMEINSCHAFT
Open Science

Albert Einstein Institute

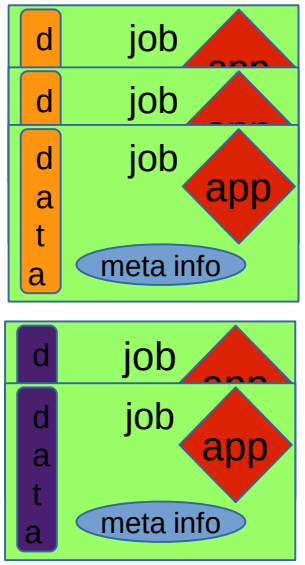
Sending data to volunteers

data

data



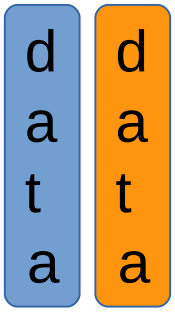
Einstein@home server

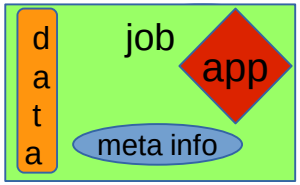


HELMHOLTZ GEMEINSCHAFT
Open Science

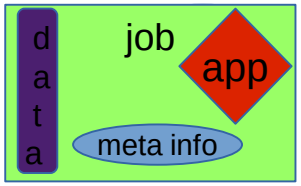
Albert Einstein Institute

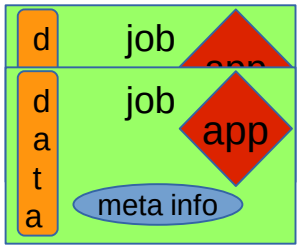
Sending data to volunteers

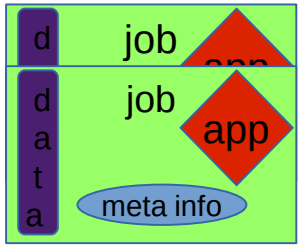




Einstein@home server







HELMHOLTZ GEMEINSCHAFT
Open Science

Albert Einstein Institute

Some Numbers

Data downloaded from AEI: 155 GB per day

Data uploaded to AEI: 16 GB per day

Metadata transfered: 94 GB per day

Storage Capacity at AEI: 50 TB (Einstein@home only)

Active Hosts: 60.000 (23.000 with a GPU)

Active Volunteers: 44.000

Detections:
55 Binary Pulsars, 18 Gamma-Ray Pulsars, 0 Gravitational Waves

Join Einstein@Home now

Go to:
https://**einsteinathome.org**
and sign-up

HELMHOLTZ GEMEINSCHAFT
Open Science

Albert Einstein Institute