



## IVOA ExecutionPlanner Data model and metadata schema

*“WhenCanIDoThis ?”*

IVOA interop, April 2022

Dave Morris, Edinburgh University

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.





The problem – different science platforms use different technologies

We end up having to understand all of them.



EUROPEAN OPEN  
SCIENCE CLOUD



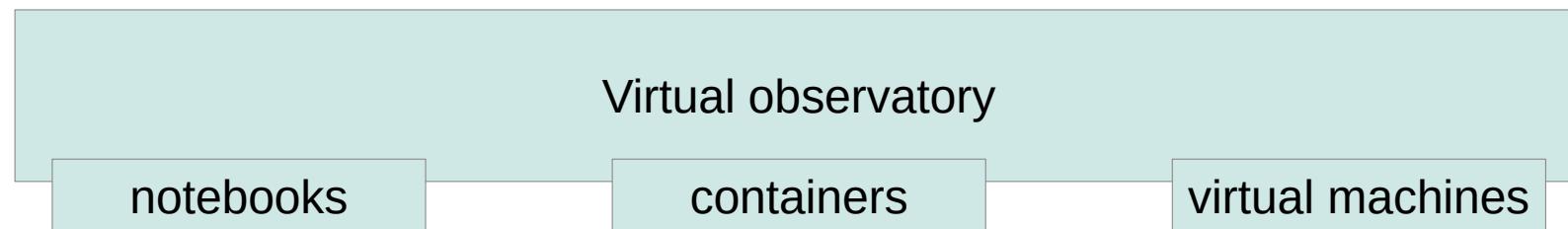
Which becomes more complex as the questions get more detailed.





The problem – different science platforms use different technologies

Plugin architecture helps



jupyter

binder

Apache  
Zeppelin



docker



kubernetes

DIRAC  
THE INTERWARE



EUROPEAN OPEN  
SCIENCE CLOUD

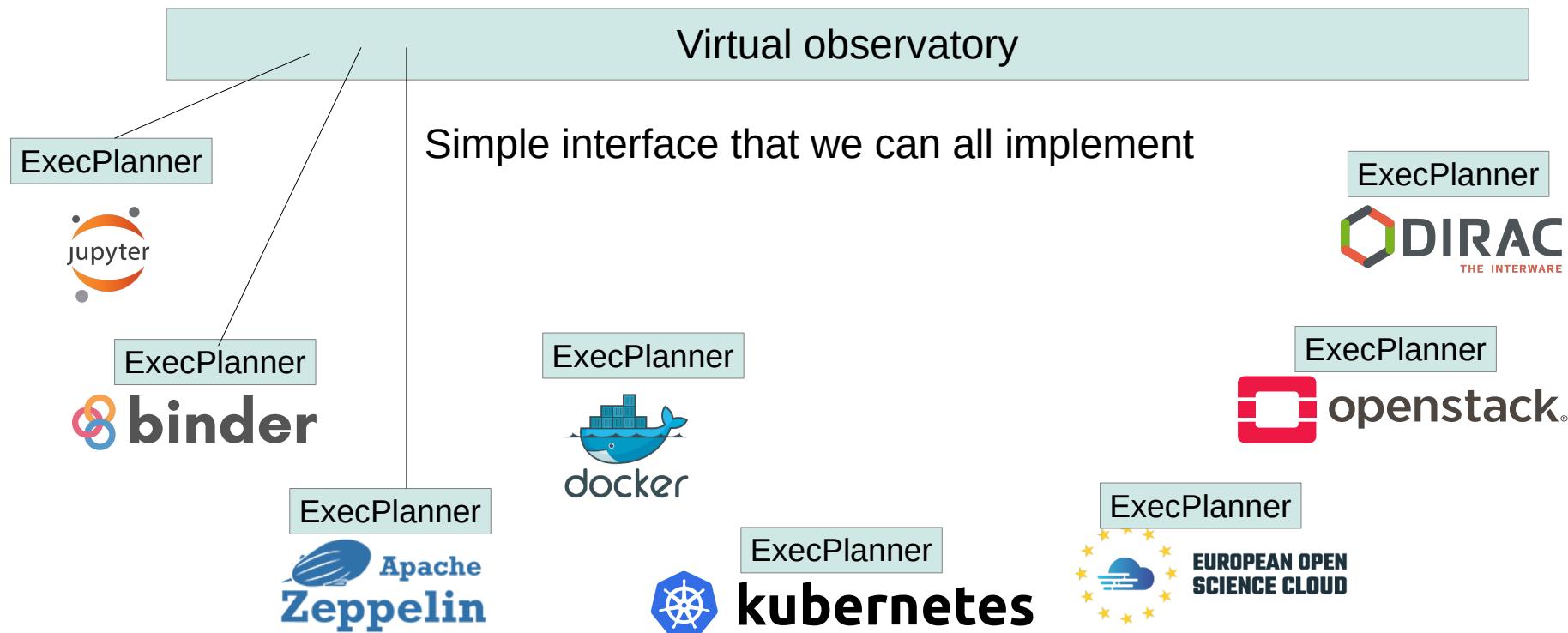


openstack.



## The solution

Delegation is the logical extension





## Current customers

ESCAPE Science Analysis Platform (ESAP)



Gaia Data Mining platform (Gaia DMP)



## Expressions of interest

CANFAR Science Platform

SciServer, JHU

D.Morris  
Institute for Astronomy,  
Edinburgh University



**ESCAPE**  
European Science Cluster of Astronomy &  
Particle physics ESFRI research Infrastructures



binder

docker



jupyter

DIRAC  
THE INTERWARE

EUROPEAN OPEN SCIENCE CLOUD

openstack.

openstack.

kubernetes

Apache Zeppelin Apache Spark™

IVOA interop  
April 2022



Simple interface that we can all implement

Common language for describing things

Meeting all the use cases adds complexity

```
{
  "type": "uri://docker-container",
  "task": {
    "image": "docker.io/example:1.0"
  }
  "data-resources": [
    ....
  ]
  "compute-resources": [
    ....
  ]
  "storage-resources": [
    ....
  ]
}
```

New sections added to the schema

Based on ESAP and GaiaDMP use cases





Simple interface that we can all implement

Common language for describing things

Meeting all the use cases adds complexity

Lower the barrier to entry by making details optional

simple

```
{
  "type": "uri://docker-container",
  "task": {
    "image": "docker.io/example:1.0"
  }
}
```

complex

```
{
  "type": "uri://docker-container",
  "task": {
    "image": "docker.io/example:1.0"
  },
  "data-resources": [ .... ],
  "compute-resources": [ .... ],
  "storage-resources": [ .... ]
}
```



Simple interface that we can all implement

Common language for describing things

Meeting all the use cases adds complexity

Lower the barrier to entry by making details optional

Flexible vocabulary should not be a problem

Customer : Can I buy an <apple>

Shop : No, we are a <bakery>, we only sell <bread>.

Customer : Can I buy an <apple>

Shop : Yes, we can offer <5> different varieties of <apple>.





Simple interface that we can all implement

Common language for describing things

Meeting all the use cases adds complexity

Lower the barrier to entry by making details optional

Adding more detail should not be a problem

Customer : Can I buy an <iPad> with <64G memory> and <5G network>

Shop : No, we are a <bakery>, we only sell <bread>.

Customer : Can I buy an <iPad> with <64G memory> and <5G network>

Shop : Yes, we can offer 2 different payment contracts.



Working with limited resources

Small task, large cloud

Simple answer

YES



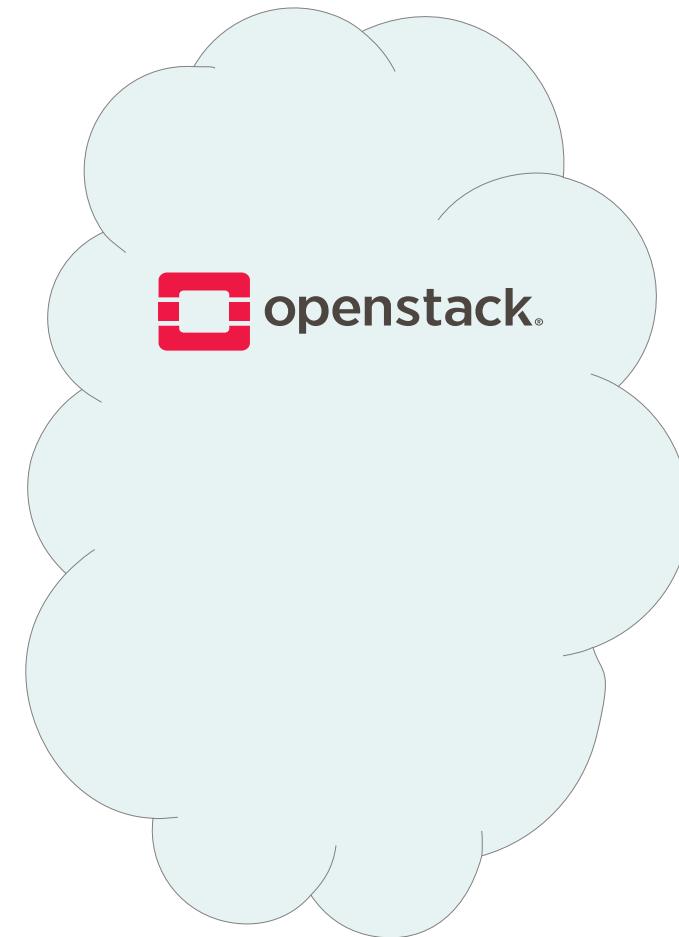


## Working with limited resources

Large task, limited cloud

Not so simple

Depends .....





Working with limited resources

Big data, complex analysis



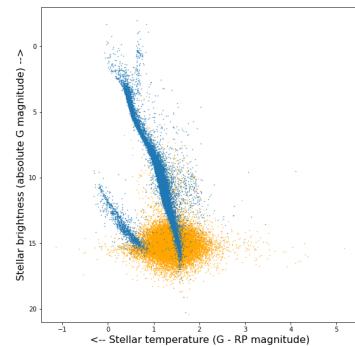
Gaia Data Mining Platform (Gaia DMP)

3.7 Tbytes of numerical data

Zeppelin - 54 cores, 86G memory

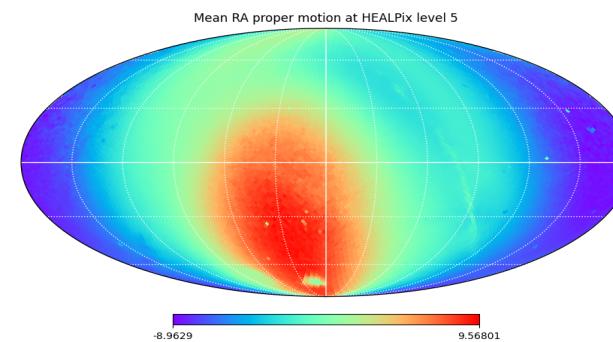
Spark - 6x 26 cores, 43G memory

> 9hrs for a complex analysis

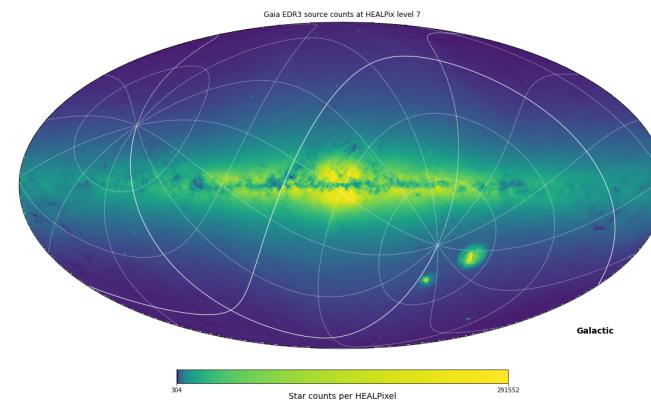


Stellar classifier N. Hambly, D. Crake 2022

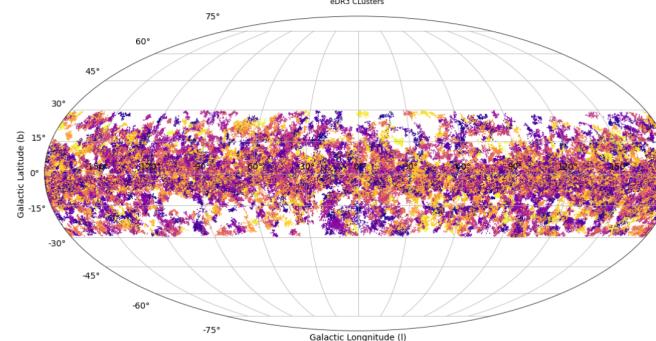
D.Morris  
Institute for Astronomy,  
Edinburgh University



Mean proper motions, N. Hambly, 2022



Mean proper motions, N. Hambly, 2022



HDBSCAN Clustering, D. Crake, 2022

**ESCAPE**  
European Science Cluster of Astronomy &  
Particle physics ESFRI research Infrastructures



IVOA interop  
April 2022



## Working with limited resources

Gaia DMP



gaia

3.7 Tbytes of numerical data

Zeppelin - 54 cores, 86G memory

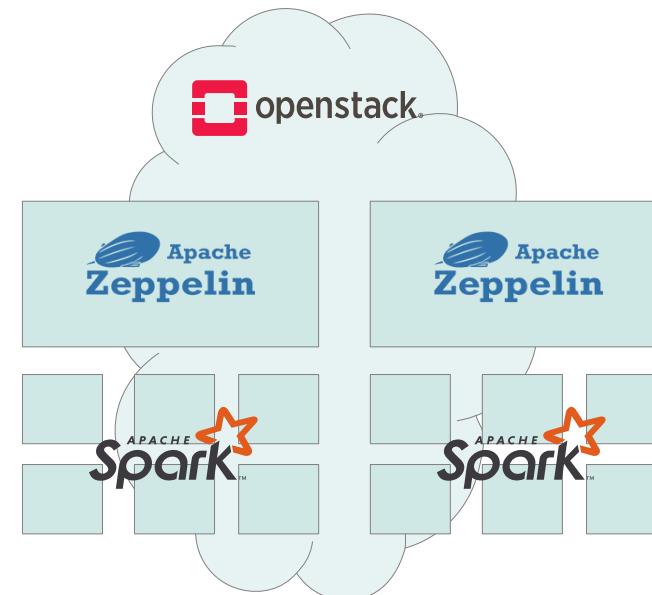
Spark - 6x 26 cores, 43G memory

> 9hrs for a complex analysis

## Limited cloud

> 50% of the available resources

We can't run more than one at a time





## ***When*** can I do this ?

Gaia DMP



3.7 Tbytes of numerical data

Zeppelin - 54 cores, 86G memory

Spark - 6x 26 cores, 43G memory

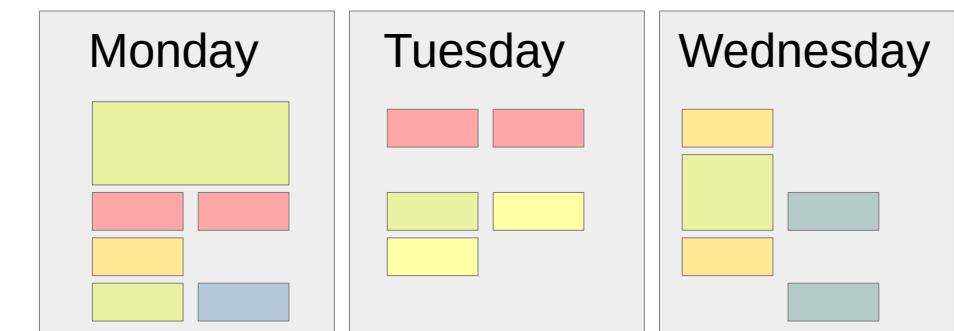
> 9hrs for a complex analysis

Booking system – resource leases

Offer : today at 14:00

Offer : today at 21:00

Offer : tomorrow at 8:00



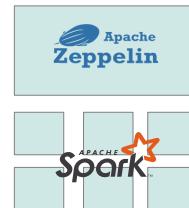


3 levels of detail

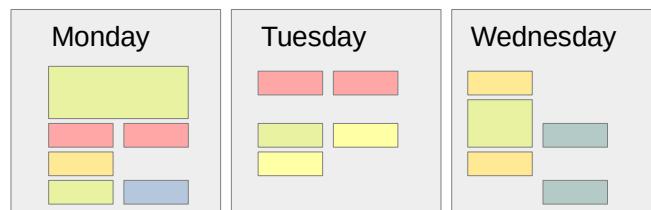
Can I do this *simple thing* ?



Can I do this *complex thing*?



**When** can I do this complex thing ?





3 levels of detail

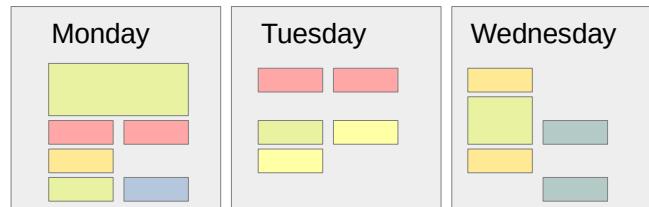
Can I do this *simple thing* ?



Can I do this *complex thing* ?



*When* can I do this complex thing ?



ESAP is here



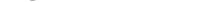
Gaia DMP needs all three levels  
Implemented in stages  
2022 - 2023





3 levels of detail

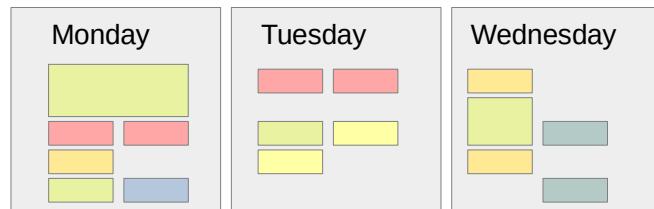
Can I do this ***simple thing*** ?



Can I do this ***complex thing*** ?



***When*** can I do this complex thing ?



Are you interested in joining us ?

What level matches your use case ?

Hackathon session



dmr@roe.ac.uk



Dave Morris

