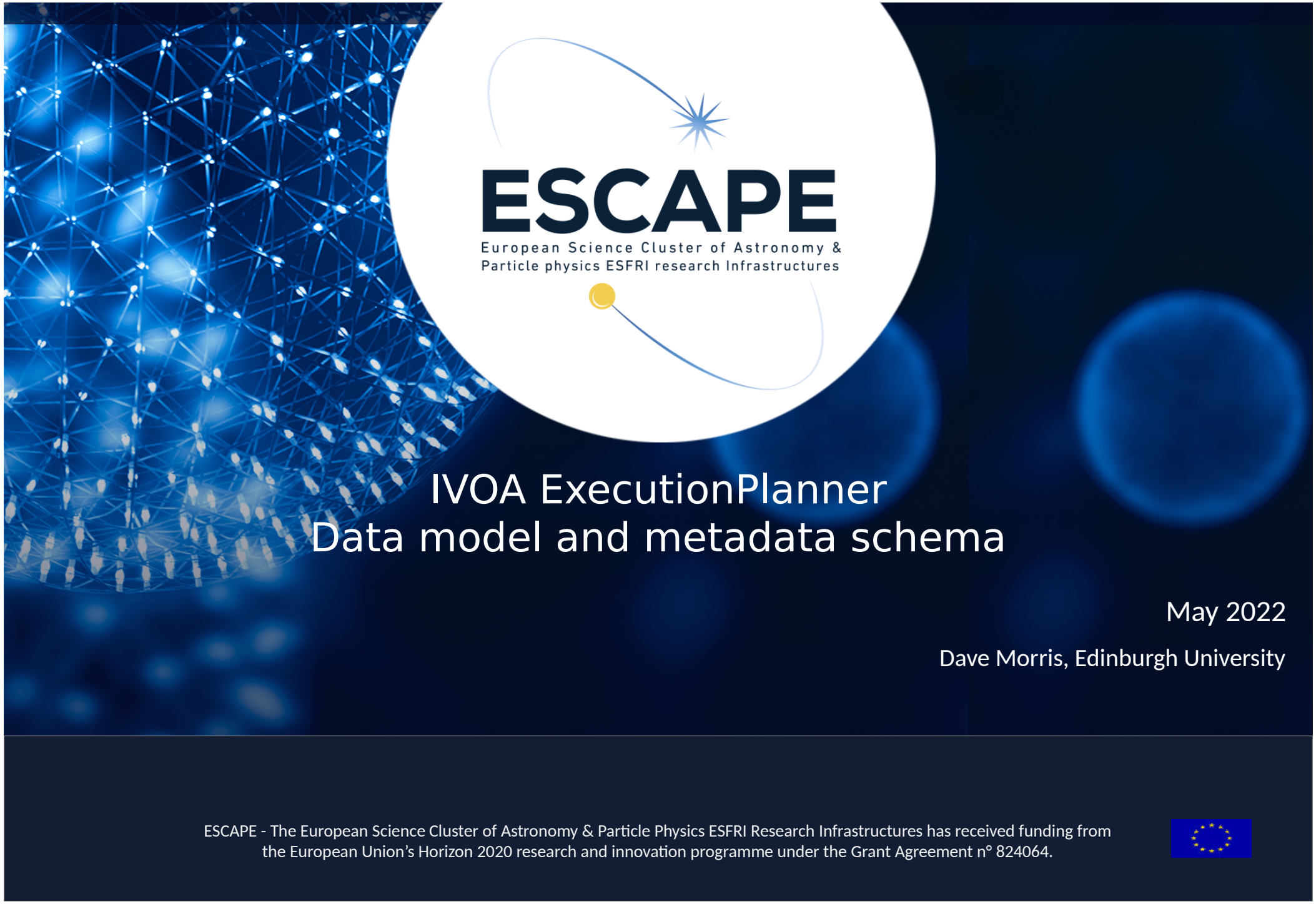


IVOA ExecutionPlanner
Data model and metadata schema

May 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 compute platforms use different technologies

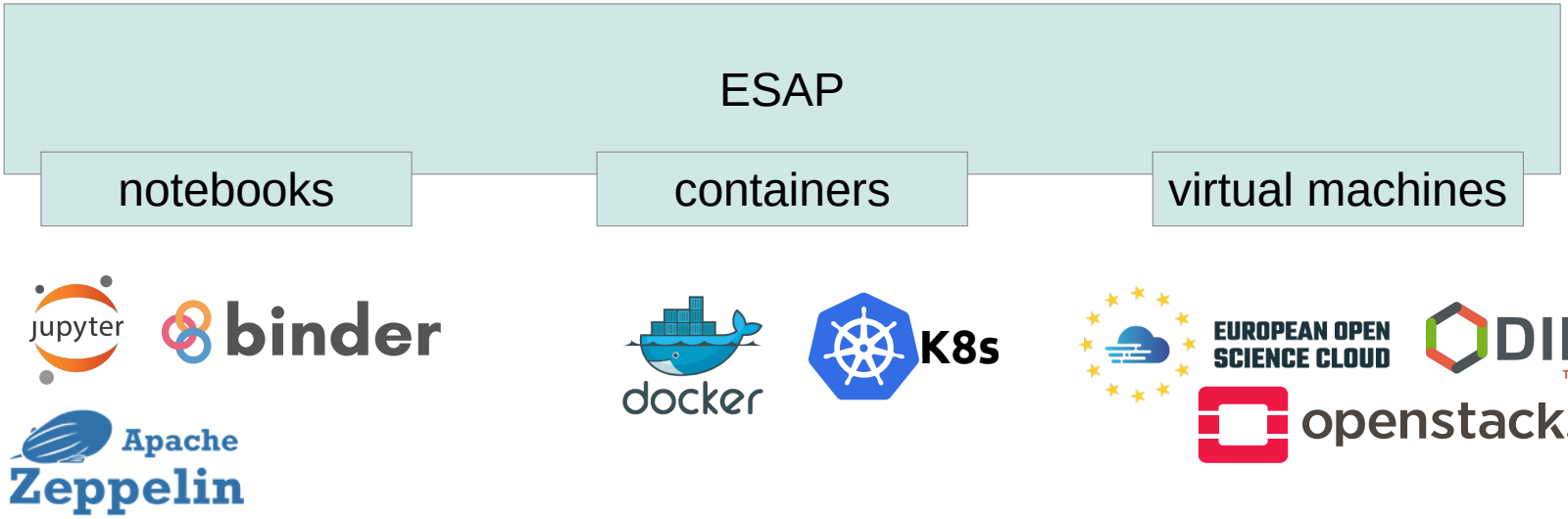
We end up having to understand all of them.

Which becomes more complex as the questions get more detailed.

D.Morris
Institute for Astronomy,
Edinburgh University

The problem – different compute platforms use different technologies

Plugin architecture helps



D.Morris
Institute for Astronomy,
Edinburgh University

Working with limited resources

Small task, large cloud

Simple answer

YES

D.Morris
Institute for Astronomy,
Edinburgh University

ESAP tech discussion
May 2022

ROYAL OBSERVATORY EDINBURGH

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

IVOA

Working with limited resources

Big data, complex analysis

gaia

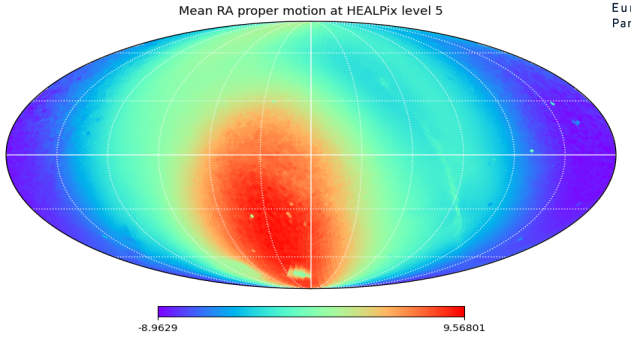
Gaia Data Mining Platform (Gaia DMp)

3.7 Tbytes of numerical data

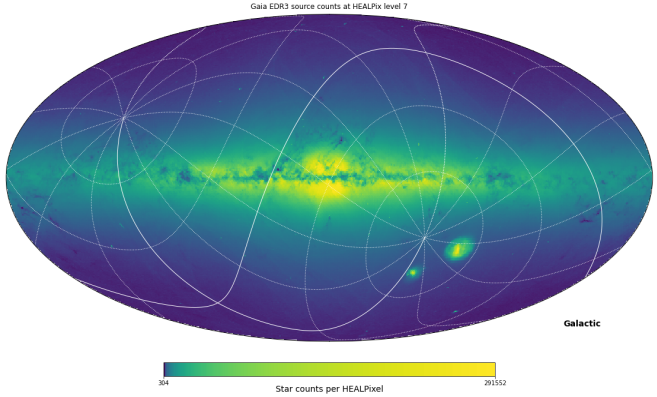
Zeppelin - 54 cores, 86G memory

Spark - 6x 26 cores, 43G memory

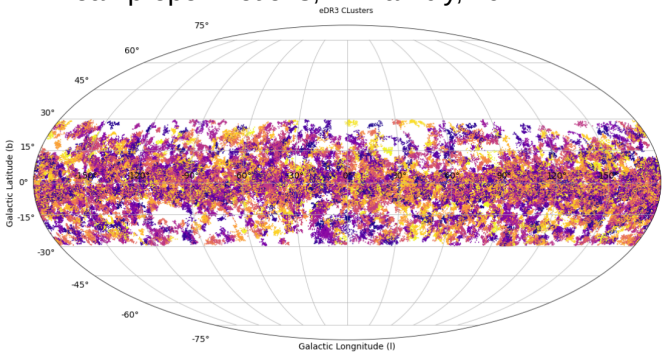
> 9hrs for a complex analysis



Mean proper motions, N. Hambly, 2022



Mean proper motions, N. Hambly, 2022



HDBSCAN Clustering, D. Crake, 2022

D.Morris
Institute for Astronomy,
Edinburgh University

THE UNIVERSITY OF EDINBURGH

ESAP tech discussion
May 2022

ROYAL OBSERVATORY EDINBURGH

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

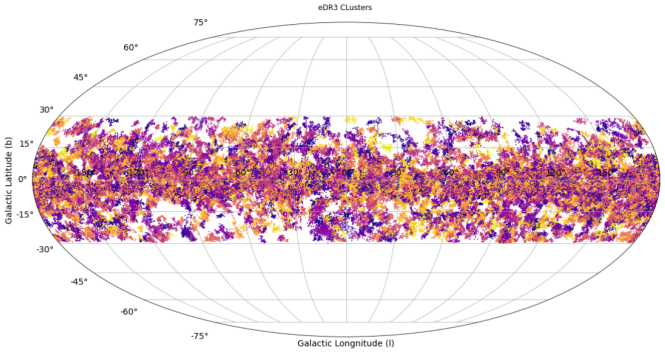
IVOA

EUROPEAN OPEN SCIENCE CLOUD

Working with limited resources

Large task, small allocation

Gaia DMp clustering notebook



3.7 Tbytes of numerical data

54 cores, 86G memory

> 9hrs for a complex analysis

Site : CESGA

VO : vo.access.egi.eu

docker

EOSC resource allocation depends on the site and virtual-organization (VO)

This notebook would not fit on this platform

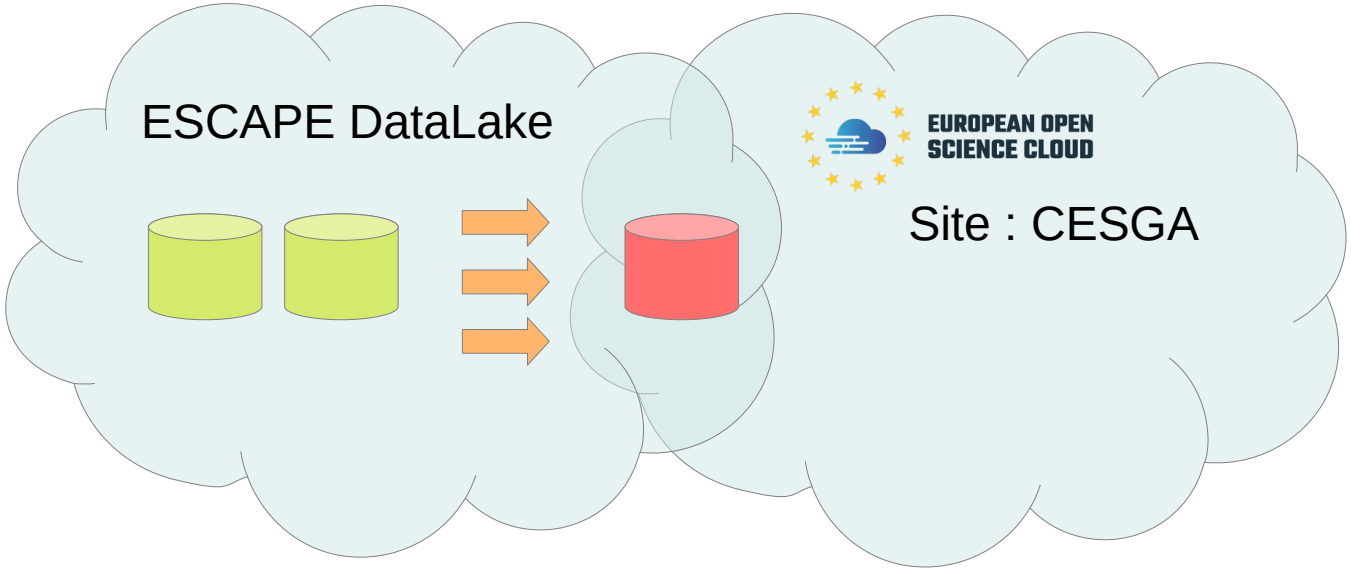
This notebook would fail ~2hrs into the task

D.Morris
Institute for Astronomy,
Edinburgh University

THE UNIVERSITY OF EDINBURGH

Working with large data

Data and compute within ESCAPE



How do you know where the data is ?

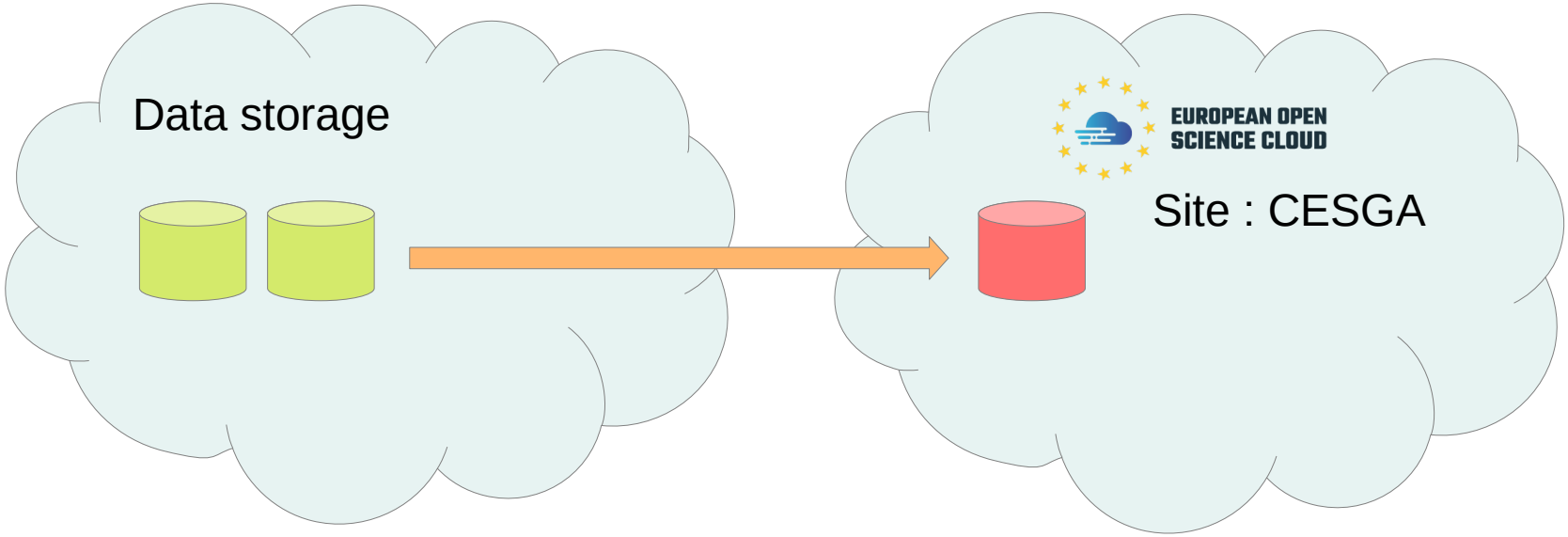
How do you know where the compute is ?

How do you know how long staging will take ?

D.Morris
Institute for Astronomy,
Edinburgh University

Working with large data

Data and/or compute outside ESCAPE



How do you know where the data is ?

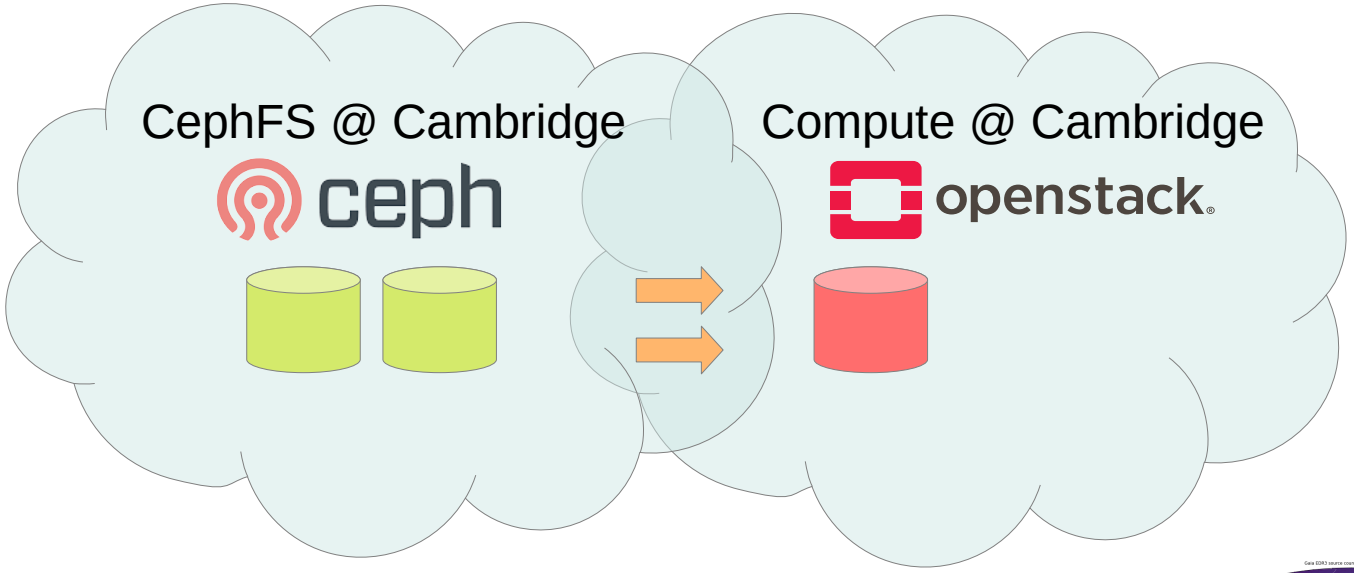
How do you know where the compute is ?

How do you know how long staging will take ?

D.Morris
Institute for Astronomy,
Edinburgh University

Working with large data

Data and compute within the same facility



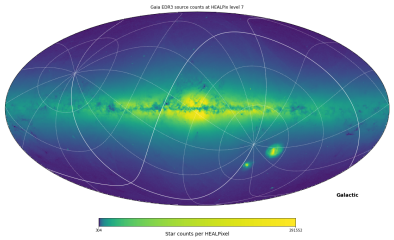
gaia

Gaia Data Mining Platform (Gaia DMp)

3.7 Tbytes of numerical data

several **hours** data transfer

within the same system



D.Morris
Institute for Astronomy,
Edinburgh University

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

D.Morris
Institute for Astronomy,
Edinburgh University

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/mycontainer"
   }
}
```

complex

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

D.Morris
Institute for Astronomy,
Edinburgh University

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 is not a problem

Customer : Can I buy an <apple>

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

Customer : Can I buy an <apple>

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

Bakery does not need to understand anything about <fruit>.

D.Morris
Institute for Astronomy,
Edinburgh University

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 is not a problem

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

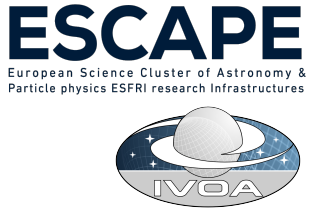
Bakery : No, we are a <bakery>, we only sell <bread>.

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

GameStop : Yes, we can offer 2 different configurations.

Bakery does not need to understand what <Gbytes> are.

D.Morris
Institute for Astronomy,
Edinburgh University



User interface
display and edit

ESAP

OSSR metadata

Execution platform

D.Morris
Institute for Astronomy,
Edinburgh University

Resource list

Cores

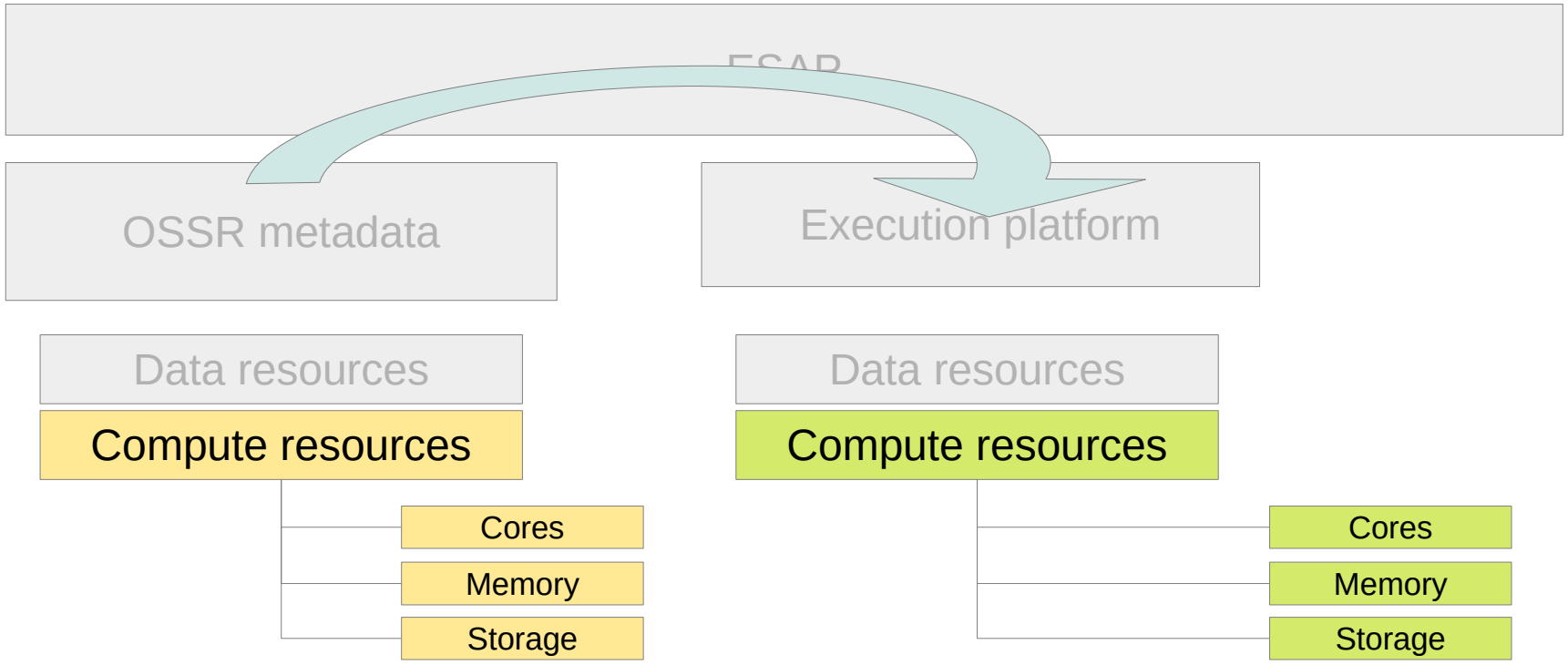
Memory

Storage

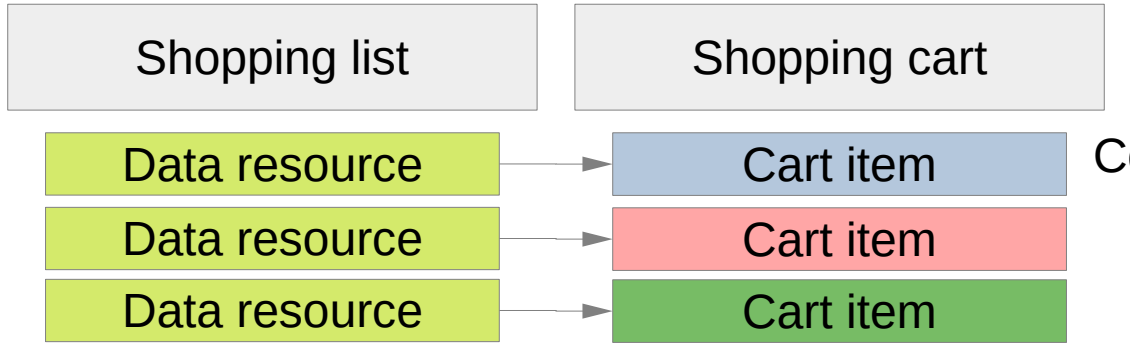
User can modify requirements to match their intended use case

Increase requirements to give themselves more space

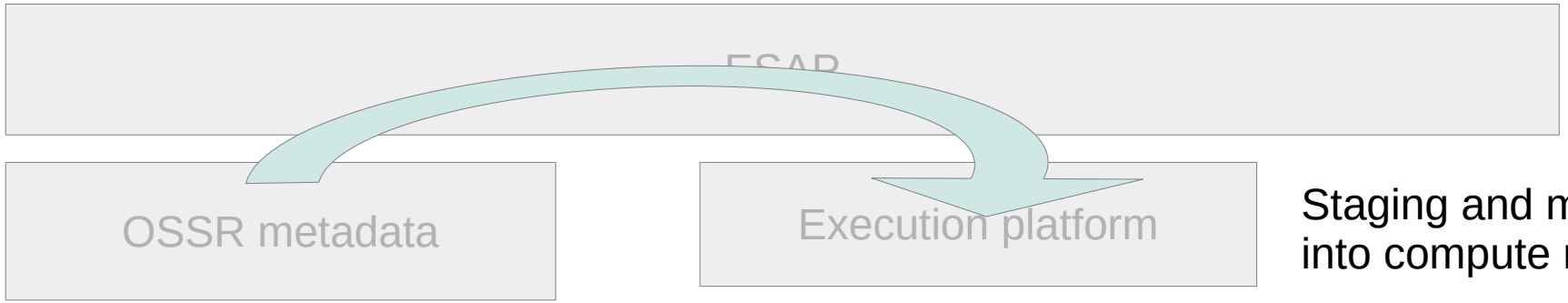
Decrease requirements to be able to run on more platforms



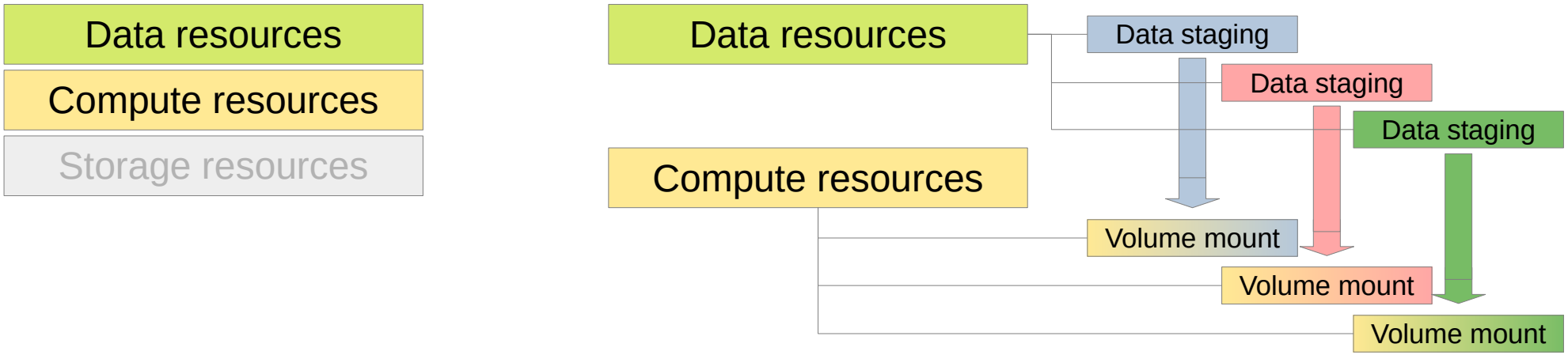
D.Morris
Institute for Astronomy,
Edinburgh University



Connecting data in shopping list to data in shopping cart



Staging and mounting data into compute resources



D.Morris
Institute for Astronomy,
Edinburgh University

Example image processing notebook.

Requires images and calibration data

Data located in shopping cart

Basic task description

```
{
  "type": "https://purl.org/esap.escape.eu/task-type/binder-notebook",
  "description": "A notebook that applies a set of calibration data to a set of images.",
  "task": {
    "repository": {
      "type": "gitlab",
      "url": "https://gitlab.in2p3.fr//..."
    },
    "path": "calibrator.ipynb"
  }
}
```

How do we do authentication for protected repositories ?

D.Morris
Institute for Astronomy,
Edinburgh University

Example image processing notebook.

Requires images and calibration data

Data located in shopping cart

Image data

```
{
 ....
 "data-resources": [
  {
   "uuid": "b77cc829-ac12-43b5-9dbe-48cfa727d14c",
   "type": "urn:data-reference",
   "description": "The image data to be processed.",
   "data-type": "https://purl.org/esap.escape.eu/data-type/optical-images",
   "mime-type": "image/fits"
  },
  ....
  ....
 ]
 ....
}
```

D.Morris
Institute for Astronomy,
Edinburgh University

Example image processing notebook.

Requires images and calibration data

Data located in shopping cart

Image data


```
{
 ....
 "data-resources": [
  {
   "uuid": "b77cc829-ac12-43b5-9dbe-48cfa727d14c",
   "type": "urn:data-reference",
   "description": "The image data to be processed.",
   "data-type": "https://purl.org/esap.escape.eu/data-type/optical-images",
   "mime-type": "image/fits",
   "location":{
    "type": "https://purl.org/esap.escape.eu/data-location/shopping-cart",
    "uuid": "9620d56c-eea1-4b98-aa85-ca5c7e439178",
    "href": "https://sdc-dev.astron.nl/shopping-cart/9620d56c-eea1-4b98-aa85-ca5c7e439178"
    }
  },
 ]
}
```


D.Morris,
Institute for Astronomy,
Edinburgh University