WP3 OSSR / WP4 CEVO

ExecutionPlanner
data model

D.Morris
Institute for Astronomy,
Edinburgh University

OSSR Technical development
April 2022

OSSR Python library

How to install & use the library

instructions in project README

```
pip install eossr

from eossr.api import get_ossr_records
```

D.Morris
Institute for Astronomy,
Edinburgh University

OSSR Python library

GitLab repository also includes executable components

- Docker container with Python library installed
- Jupyter notebooks with example code

How to enable ESAP science platform to discover and launch these

D.Morris
Institute for Astronomy,
Edinburgh University

OSSR/ESAP metadata

How to discover executable components – include a list at a known location.

```
{
"executables": [
    {
    "uuid": "5cec9f3d-79c0-49ad-9a5a-455596c55341",
    "type": "urn:types/jupyter-notebook",
    "title": "Find ESCAPE OSSR records",
    "description": "How to find and access Zenodo records from the OSSR dataset.",
    ....
    },
    {
    "uuid": "1d5e06cc-e1a9-4eb9-8d99-d
    "type": "urn:types/jupyter-notebook",
    "title": "Upload records on the OSSR",
    "description": "How to create Zenodo records in the OSSR dataset.",
    ....
    }
  ]
}
```

Unique identifier

Machine readable type

Human readable title and description

D.Morris
Institute for Astronomy,
Edinburgh University

OSSR/ESAP metadata

How to launch executable components – include details of the resources required.

```
{
"executables": [
    {
    "uuid": "5cec9f3d-79c0-49    a5a-455596c55341",
    "type": "urn:types/jupyte    notebook",
    ....
    "compute-resources": [ .. ],
    "storage-resources": [ .. ],
    "data-resources": [ .. ]
    }
  ]
}
```

Number and size of compute resources

Storage resources linked to task execution

Data resource inputs and outputs

D.Morris
Institute for Astronomy,
Edinburgh University

OSSR/ESAP metadata

How to launch executable components – include details of the resources required.

```
{
"executables": [
    {
    "uuid": "5cec9f3d-79c0-49ad-9a5a-455596c55341",
    "type": "urn:types/jupyter-notebook",
    ....
    "compute-resources": [
        {
        "mincores": 1,
        "minmemory": 1
        }
      ]
    }
  ]
}
```

Small footprint enables the task to run anywhere

D.Morris
Institute for Astronomy,
Edinburgh University

OSSR/ESAP metadata

How to launch executable components – include details of the resources required.

```
{
"executables": [
    {
    "uuid": "5cec9f3d-79c0-49ad-9a5a-455596c55341",
    "type": "urn:types/jupyter-notebook",
    ....
    "compute-resources": [
        {
        "mincores": 54,
        "minmemory": 86
        }
      ]
    }
  ]
}
```

Large footprint helps ESAP find a suitable platform

D.Morris
Institute for Astronomy,
Edinburgh University

OSSR/ESAP metadata

How to launch executable components – include details of the resources required.

```
{
"executables": [
    {
    "uuid": "5cec9f3d-79c0-49ad-9a5a-455596c55341",
    "type": "urn:types/jupyter-notebook",
    ....
    "compute-resources": [ .. ],
    "data-resources": [
        {
        "uuid": "401f6251-be15-4f8a-bad7-38ef4d6cb7c1",
        "data-content": "urn:astronomy-observation",
        "data-format":  "urn:parquet-binary"
        ...
        }
      ]
    }
  ]
}
```

Data resources gives ESAP a shopping list of data to find

D.Morris
Institute for Astronomy,
Edinburgh University

OSSR/ESAP metadata

Small footprint enables the task to run anywhere

Range of different use cases

OSSR Python library

Single compute node
1 cpu core, 1G of memory

Gaia data mining platform
HDBSCAN clustering algorithm

8 compute nodes
210 cpu cores, 344G memory
190 hours to execute
5Tbyte dataset

Large footprint helps ESAP find a suitable platform

D.Morris
Institute for Astronomy,
Edinburgh University

OSSR/ESAP metadata

User might increase requirements to make space for experimenting

Range of different use cases

OSSR Python library

Single compute node
~~1~~ 4 cpu core, ~~1G~~ 16G of memory

Gaia data mining platform

HDBSCAN clustering algorithm

~~8~~ 6 compute nodes
~~210~~ 50 cpu cores, ~~344G~~ 100G memory
~~5T~~ 500Mbyte dataset

User might decrease requirements to process a smaller dataset on more platforms

D.Morris
Institute for Astronomy,
Edinburgh University