

E2.1 SERVICIOS Y ESQUEMA DEL CATÁLOGO

ESPACIO DE DATOS LINGÜÍSTICO (SER 15/23 OTT)

Resumen

Este documento describe los principales servicios del conector Inesdata que se van a usar en el despliegue del espacio de datos lingüístico EDL. En esta primera versión del entregable se incluyen los servicios que hasta la fecha están disponibles en el repositorio github de Inesdata. Además, el esquema del catálogo no se incluye en esta versión debido a que en la versión actual de los componentes de los espacios de datos de Inesdata el catálogo no persiste en una base de datos.

Andrés García-Silva

07/05/2024

Expert.ai Expert.ai Language Technology Research Lab

Historia

revisión	Fecha	Descripción	Autor
0.1	07/05/2024	Primera versión completa	Andrés García-Silva
0.2	09/05/2024	Revisión del documento	Andrés García-Silva
0.3	13/05/2024	Revisión del documento	José Manuel Gómez Pérez
1.0	16/05/2024	Versión final del documento	Andrés García-Silva

Contents

1	Introducción.....	4
2	Servicio de inicio de sesión	4
3	Registro del plano de datos.....	5
4	Creación de activos.....	6
4.1	Llamada del servicio para un recurso disponible vía http:.....	6
4.2	Llamada del servicio para un recurso disponible vía s3:	6
5	Creación de una política.....	7
6	Creación de un contrato.....	8
7	Visualizar los activos disponibles en el espacio de datos	9
8	Negociación de contratos.....	13
9	Visualización del estado de la negociación de un contrato	14
10	Inicio de transferencia de activos	14
10.1	Transferencia de activos a un recurso http	14
10.2	Transferencia de activos a almacenamiento s3	15
11	Visualización del estado de la transferencia de ficheros	15
12	Conclusiones.....	16

1 Introducción

El espacio de datos lingüísticos EDL se apoya en los componentes INESData para el despliegue de espacios de datos. Entre los componentes uno de los principales es el conector que permite que un usuario del espacio de datos pueda proveer activos al espacio de datos y consumir activos disponibles en el espacio de datos. El conector garantiza los principios de los espacios de datos incluyendo la confianza y la propiedad de los datos.

El flujo de trabajo para un usuario que busca compartir activos en el espacio de datos incluye los siguientes servicios del conector:

- Iniciar sesión en el espacio de datos
- Registrar el plano de datos donde se almacenan indicando el tipo de transferencia soportada
- Crear activos a compartir indicando su ubicación
- Crear políticas de acceso y contratación a los activos usando odrl como vocabulario principal.
- Creación de un contrato que vincule políticas de acceso y contratación con los activos

Para un usuario que actúa como consumidor de activos en el espacio de datos el flujo de trabajo incluye los siguientes servicios del conector:

- Iniciar sesión en el espacio de datos
- Registrar el plano de datos donde se almacenan los activos recibidos por el usuario indicando los métodos de transferencia soportados
- Visualización de los activos disponibles y accesibles en el catálogo del espacio de datos
- Negociación del contrato que el proveedor del activo ha definido.
- Iniciar la transferencia del activo desde el proveedor al consumidor una vez se ha negociado el contrato.

En este documento se describen los servicios del conector que ofrecen soporte a los flujos de trabajo de los usuarios consumidores y proveedores de activos en el espacio de datos.

2 Servicio de inicio de sesión

Este servicio se utiliza para iniciar sesión como usuario existente en keycloak y obtener un token de acceso. Al utilizar cualquier otro servicio de conector, el usuario debe enviar el token de acceso como prueba de ser un usuario autenticado en el espacio de datos. El servicio está expuesto en el puerto 8080.

```
POST /realms/dataspace/protocol/openid-connect/token HTTP/1.1
Host: localhost:8080
Content-Type: application/x-www-form-urlencoded
Content-Length: 79
client_id=dataspace-
users&grant_type=password&username=%username%&password=%password%
```

Salida del servicio:

```
{
  "access_token": "access_token_value",
  "expires_in": number,
  "refresh_expires_in": number,
  "refresh_token": "refresh_token_value",
  "token_type": "Bearer",
```

```
"not-before-policy": 0,  
"session_state": "session_state_value",  
"scope": "email profile"  
}
```

3 Registro del plano de datos

En la configuración del conector se define en que puerto se exponen los servicios de gestión que soporta. En este caso el puerto definido para estos servicios, incluyendo el registro del plano de datos, se exponen en el puerto 19193. El servicio de registro del plano de datos es accesible en la ruta /management/v2/dataplanes.

Este servicio permite registrar el plano de datos del conector. El plano de datos es donde los datos residen y se transfieren. El conector de Inesdata soporta transferencia de datos usando el protocolo Http y servicios de almacenamiento en la nube como AmazonS3 que ofrecen una interfaz de servicio web.

En la configuración del conector se define que el plano de datos se expone en el puerto 19192 en la ruta /control/transfer internamente y en el puerto 19292 en la ruta /public para acceso público.

```
POST /management/v2/dataplanes HTTP/1.1  
Host: localhost:19193  
Content-Type: application/json  
Authorization: Bearer %Access_token%  
Content-Length: 436  
  
{  
  "@context": {  
    "@vocab": "https://w3id.org/edc/v0.0.1/ns/"  
  },  
  "@id": "http-pull-provider-dataplane",  
  "url": "http://connector-c1:19192/control/transfer",  
  "allowedSourceTypes": [  
    "HttpData",  
    "AmazonS3"  
  ],  
  "allowedDestTypes": [  
    "HttpProxy",  
    "HttpData",  
    "AmazonS3"  
  ],  
  "properties": {  
    "https://w3id.org/edc/v0.0.1/ns/publicApiUrl": "http://connector-c1:19291/public/"  
  }  
}
```

Salida del servicio

```
{
  "@type": "IdResponse",
  "@id": "http-pull-provider-dataplane",
  "createdAt": 1715098050963,
  "@context": {
    "@vocab": "https://w3id.org/edc/v0.0.1/ns/",
    "edc": "https://w3id.org/edc/v0.0.1/ns/",
    "odrl": "http://www.w3.org/ns/odrl/2/"
  }
}
```

4 Creación de activos

El conector expone un servicio para la creación de activos en la ruta `/management/v3/assets`. Este servicio recibe un documento json con el id del activo, el nombre, y el tipo del activo. Además, se incluye información de la dirección donde se encuentra el activo. Esta dirección incluye el tipo de dirección que define si es un recurso disponible vía http o si está en un repositorio s3. Para los recursos web se especifica la url donde se encuentra y para los recursos en S3 se especifica el bucket y el nombre del fichero entre otros.

4.1 Llamada del servicio para un recurso disponible vía http:

Para crear un activo disponible como un recurso http se usa la siguiente llamada.

```
POST /management/v3/assets HTTP/1.1
Host: localhost:19193
Content-Type: application/json
Authorization: Bearer %Access_token%
Content-Length: 357

{
  "@context": {
    "@vocab": "https://w3id.org/edc/v0.0.1/ns/"
  },
  "@id": "c1-http-asset1",
  "properties": {
    "name": "c1-http-asset1",
    "contenttype": "application/json"
  },
  "dataAddress": {
    "type": "HttpData",
    "name": "c1-http-asset1",
    "baseUrl": "http://http-proxy:4000/test/call",
    "proxyPath": "true"
  }
}
```

4.2 Llamada del servicio para un recurso disponible vía s3:

Para crear un activo disponible en un repositorio s3 se usa la siguiente llamada:

```
POST /management/v3/assets HTTP/1.1
Host: localhost:19193
Content-Type: application/json
Authorization: Bearer %Access_token%
Content-Length: 341

{
  "@context": {
    "@vocab": "https://w3id.org/edc/v0.0.1/ns/"
  },
  "@id": "c1-s3-asset2",
  "properties": {
    "name": "c1-s3-asset2",
    "contenttype": "text/plain"
  },
  "dataAddress": {
    "type": "AmazonS3",
    "region": "us-east-1",
    "keyName": "s3-c1-file.txt",
    "bucketName": "bucket-connector-1"
  }
}
```

Salida del servicio

```
{
  "@type": "IdResponse",
  "@id": "c1-s3-asset2",
  "createdAt": 1715172777013,
  "@context": {
    "@vocab": "https://w3id.org/edc/v0.0.1/ns/",
    "edc": "https://w3id.org/edc/v0.0.1/ns/",
    "odrl": "http://www.w3.org/ns/odrl/2/"
  }
}
```

5 Creación de una política

La versión actual del conector de INESData no aplica las políticas de acceso ni ejecución. De todos modos, es necesario crear una política para poder crear un contrato asociado a un activo. El servicio de creación de políticas está disponible en la ruta /management/v2/policydefinitions

La siguiente llamada crea una política vacía.

```
POST /management/v2/policydefinitions HTTP/1.1
Host: localhost:19193
Content-Type: application/json
Authorization: Bearer %Access_token%
Content-Length: 271
```

```
{
  "@context": {
    "@vocab": "https://w3id.org/edc/v0.0.1/ns/",
    "odrl": "http://www.w3.org/ns/odrl/2/"
  },
  "@id": "emptyPolicy",
  "policy": {
    "@type": "http://www.w3.org/ns/odrl/2/Set",
    "odrl:permission": [],
    "odrl:prohibition": [],
    "odrl:obligation": []
  }
}
```

Salida del servicio

```
{
  "@type": "IdResponse",
  "@id": "emptyPolicy",
  "createdAt": 1715175303371,
  "@context": {
    "@vocab": "https://w3id.org/edc/v0.0.1/ns/",
    "edc": "https://w3id.org/edc/v0.0.1/ns/",
    "odrl": "http://www.w3.org/ns/odrl/2/"
  }
}
```

6 Creación de un contrato

Un contrato aplica una política de acceso y una política de contrato a un conjunto de activos. El servicio de creación de un contrato está disponible en la ruta: /management/v2/contractdefinitions

POST /management/v2/contractdefinitions HTTP/1.1

Host: localhost:19193

Content-Type: application/json

Authorization: Bearer %Access_token%

Content-Length: 204

```
{
  "@context": {
    "@vocab": "https://w3id.org/edc/v0.0.1/ns/"
  },
  "@id": "allowAllContract",
  "accessPolicyId": "emptyPolicy",
  "contractPolicyId": "emptyPolicy",
  "assetsSelector": []
}
```



```
}
```

Salida del servicio

```
{
  "@type": "IdResponse",
  "@id": "allowAllContract",
  "createdAt": 1715175770239,
  "@context": {
    "@vocab": "https://w3id.org/edc/v0.0.1/ns/",
    "edc": "https://w3id.org/edc/v0.0.1/ns/",
    "odrl": "http://www.w3.org/ns/odrl/2/"
  }
}
```

7 Visualizar los activos disponibles en el espacio de datos

Cada conector tiene una copia del catálogo de activos disponibles en el espacio de datos. El servicio para consultar el catálogo está en la ruta `/management/federatedcatalog/` y permite definir criterios para llevar a cabo la búsqueda de activos. Cuando no se define ningún criterio de búsqueda se retornan todos los activos a los que el usuario tiene acceso.

POST `/management/federatedcatalog` HTTP/1.1

Host: `localhost:29193`

Content-Type: `application/json`

Authorization: `Bearer`

Content-Length: `102`

```
{
  "edc:operandLeft": "",
  "edc:operandRight": "",
  "edc:operator": "",
  "edc:Criterion": ""
}
```

Salida del servicio

El servicio retorna cada uno de los datasets distribuidos por cada conector. Para cada conector, se presentan los datasets que expone en el espacio de datos, y la política de contratación también llamada de oferta, y los medios de distribución soportados entre los que se incluyen: usar un proxy http en la modalidad push, o un recurso web disponible vía http (HttpData) en las modalidades push y pull, o como un recurso AmazonS3 en la modalidad push.

Además, se expone la información del servicio del catálogo de cada conector en el elemento `http://connector-c1:19194/protocol`.

```
[
  {
    "@id": "b5cfaa16-b350-41b4-82f2-f4a68ca5f654",
    "@type": "http://www.w3.org/ns/dcat#Catalog",

```

```

    "https://w3id.org/dspace/v0.8/participantId": "connector-c1",
    "http://www.w3.org/ns/dcat#dataset": [
      {
        "@id": "c1-http-asset1",
        "@type": "http://www.w3.org/ns/dcat#Dataset",
        "odrl:hasPolicy": {
          "@id":
"YWxsb3dBbGxDb250cmFjdA==:YzEtaHR0cC1hc3NldDE=:Y2E5MjYjYjAtNmQ1Mi00ZjA1LTNmNTk
tYzQ5NzZmY2EwZmYy",
          "@type": "odrl:Offer",
          "odrl:permission": [],
          "odrl:prohibition": [],
          "odrl:obligation": []
        },
        "http://www.w3.org/ns/dcat#distribution": [
          {
            "@type": "http://www.w3.org/ns/dcat#Distribution",
            "http://purl.org/dc/terms/format": {
              "@id": "HttpProxy-PUSH"
            },
            "http://www.w3.org/ns/dcat#accessService": {
              "@id": "2427c339-437f-4d67-8000-e4c057299865",
              "@type": "http://www.w3.org/ns/dcat#DataService"
            }
          },
          {
            "@type": "http://www.w3.org/ns/dcat#Distribution",
            "http://purl.org/dc/terms/format": {
              "@id": "HttpData-PULL"
            },
            "http://www.w3.org/ns/dcat#accessService": {
              "@id": "2427c339-437f-4d67-8000-e4c057299865",
              "@type": "http://www.w3.org/ns/dcat#DataService"
            }
          },
          {
            "@type": "http://www.w3.org/ns/dcat#Distribution",
            "http://purl.org/dc/terms/format": {
              "@id": "HttpData-PUSH"
            },
            "http://www.w3.org/ns/dcat#accessService": {
              "@id": "2427c339-437f-4d67-8000-e4c057299865",
              "@type": "http://www.w3.org/ns/dcat#DataService"
            }
          },
          {
            "@type": "http://www.w3.org/ns/dcat#Distribution",
            "http://purl.org/dc/terms/format": {
              "@id": "AmazonS3-PUSH"
            },
            "http://www.w3.org/ns/dcat#accessService": {
              "@id": "2427c339-437f-4d67-8000-e4c057299865",

```

```

        "@type": "http://www.w3.org/ns/dcat#DataService"
      }
    },
    {
      "name": "c1-http-asset1",
      "id": "c1-http-asset1",
      "contenttype": "application/json"
    },
    {
      "@id": "c1-s3-asset2",
      "@type": "http://www.w3.org/ns/dcat#Dataset",
      "odrl:hasPolicy": {
        "@id":
"YWxsb3dBbGxDb250cmFjdA==YzEtczMtYXNzZXQy:MzYwYTg5Y2ItMzhjNy00M2U0LTlmODctYzUwOTAyM2I2OTk2",
        "@type": "odrl:Offer",
        "odrl:permission": [],
        "odrl:prohibition": [],
        "odrl:obligation": []
      },
      "http://www.w3.org/ns/dcat#distribution": [
        {
          "@type": "http://www.w3.org/ns/dcat#Distribution",
          "http://purl.org/dc/terms/format": {
            "@id": "HttpProxy-PUSH"
          },
          "http://www.w3.org/ns/dcat#accessService": {
            "@id": "2427c339-437f-4d67-8000-e4c057299865",
            "@type": "http://www.w3.org/ns/dcat#DataService"
          }
        },
        {
          "@type": "http://www.w3.org/ns/dcat#Distribution",
          "http://purl.org/dc/terms/format": {
            "@id": "HttpData-PULL"
          },
          "http://www.w3.org/ns/dcat#accessService": {
            "@id": "2427c339-437f-4d67-8000-e4c057299865",
            "@type": "http://www.w3.org/ns/dcat#DataService"
          }
        },
        {
          "@type": "http://www.w3.org/ns/dcat#Distribution",
          "http://purl.org/dc/terms/format": {
            "@id": "HttpData-PUSH"
          },
          "http://www.w3.org/ns/dcat#accessService": {
            "@id": "2427c339-437f-4d67-8000-e4c057299865",
            "@type": "http://www.w3.org/ns/dcat#DataService"
          }
        }
      ]
    }
  ]
}

```

```

        "@type": "http://www.w3.org/ns/dcat#Distribution",
        "http://purl.org/dc/terms/format": {
            "@id": "AmazonS3-PUSH"
        },
        "http://www.w3.org/ns/dcat#accessService": {
            "@id": "2427c339-437f-4d67-8000-e4c057299865",
            "@type": "http://www.w3.org/ns/dcat#DataService"
        }
    },
    ],
    "name": "c1-s3-asset2",
    "id": "c1-s3-asset2",
    "contenttype": "text/plain"
}

],
"http://www.w3.org/ns/dcat#service": {
    "@id": "2427c339-437f-4d67-8000-e4c057299865",
    "@type": "http://www.w3.org/ns/dcat#DataService",
    "http://purl.org/dc/terms/terms": "connector",
    "http://purl.org/dc/terms/endpointUrl": "http://connector-
c1:19194/protocol"
},
"originator": "http://connector-c1:19194/protocol",
"participantId": "connector-c1",
"@context": {
    "@vocab": "https://w3id.org/edc/v0.0.1/ns/",
    "edc": "https://w3id.org/edc/v0.0.1/ns/",
    "odrl": "http://www.w3.org/ns/odrl/2/"
}
},
{
    "@id": "242e0299-0d16-4fb3-94e1-d1fd39756938",
    "@type": "http://www.w3.org/ns/dcat#Catalog",
    "https://w3id.org/dspace/v0.8/participantId": "connector-c2",
    "http://www.w3.org/ns/dcat#dataset": [],
    "http://www.w3.org/ns/dcat#service": {
        "@id": "49c90e68-9a81-43db-b557-75b2cc85f928",
        "@type": "http://www.w3.org/ns/dcat#DataService",
        "http://purl.org/dc/terms/terms": "connector",
        "http://purl.org/dc/terms/endpointUrl": "http://connector-
c2:29194/protocol"
    },
    "originator": "http://connector-c2:29194/protocol",
    "participantId": "connector-c2",
    "@context": {
        "@vocab": "https://w3id.org/edc/v0.0.1/ns/",
        "edc": "https://w3id.org/edc/v0.0.1/ns/",
        "odrl": "http://www.w3.org/ns/odrl/2/"
    }
}
}
]

```

8 Negociación de contratos

Para poder acceder a un activo es necesario negociar un contrato con el proveedor del activo. El servicio de negociación de contratos está en la ruta `/management/v2/contractnegotiations` del conector. Se debe especificar la url de la parte con la que se va a negociar en el elemento `counterPartyAddress`. La url de la parte se devuelve como resultado de la consulta al catálogo en el elemento `originator`.

Además, es necesario definir el activo que se quiere negociar en el atributo del objeto `policy target`.

```
POST /management/v2/contractnegotiations HTTP/1.1
Host: localhost:29193
Content-Type: application/json
Authorization: Bearer %Access_token%
Content-Length: 587

{
  "@context": {
    "@vocab": "https://w3id.org/edc/v0.0.1/ns/"
  },
  "@type": "NegotiationInitiateRequestDto",
  "connectorId": "provider",
  "counterPartyAddress": "http://connector-c1:19194/protocol",
  "consumerId": "connector-c2",
  "providerId": "connector-c1",
  "protocol": "dataspace-protocol-http",
  "policy": {
    "@context": "http://www.w3.org/ns/odrl.jsonld",
    "@id": "[[[ID_CONTRACT]]]",
    "@type": "Offer",
    "permission": [],
    "prohibition": [],
    "obligation": [],
    "assigner": "connector-c2",
    "target": "c1-http-asset1"
  }
}
```

Salida del servicio

```
{
  "@type": "IdResponse",
  "@id": "68d4911a-a256-40f4-9fd2-8c0c8db9db94",
  "createdAt": 1715182409194,
  "@context": {
    "@vocab": "https://w3id.org/edc/v0.0.1/ns/",
    "edc": "https://w3id.org/edc/v0.0.1/ns/",
  }
}
```

```
    "odrl": "http://www.w3.org/ns/odrl/2/"
  }
}
```

9 Visualización del estado de la negociación de un contrato

Para visualizar el estado de una negociación iniciada se pasa el id del contrato devuelto por el servicio de creación de contratos al final de la ruta `/management/v2/contractnegotiations/`.

```
GET /management/v2/contractnegotiations/%CONTRACT_ID% HTTP/1.1
Host: localhost:29193
Authorization: Bearer
```

Cuando el contrato se haya llevado a cabo el sistema retorna un id de la negociación: `contractagreementid`.

10 Inicio de transferencia de activos

Una vez el contrato ha sido aprobado por el proveedor, el consumidor puede solicitar su transferencia usando el servicio desplegado en la ruta `/management/v2/transferprocesses/`. La transferencia puede realizarse usando `s3` o `http`.

10.1 Transferencia de activos a un recurso http

En el cuerpo del mensaje se especifica el id del contrato, el id del activo que se solicita y la url del servicio del protocolo de datos del proveedor. El servicio retorna un id de la transferencia.

```
POST /management/v2/transferprocesses HTTP/1.1
Host: localhost:29193
Content-Type: application/json
Authorization: Bearer %Access_token%
Content-Length: 348

{
  "@context": {
    "@vocab": "https://w3id.org/edc/v0.0.1/ns/"
  },
  "@type": "TransferRequestDto",
  "connectorId": "connector-c2",
  "counterPartyAddress": "http://connector-c1:19194/protocol",
  "contractId": "%contractagreementid%",
  "assetId": "c1-http-asset1",
  "protocol": "dataspace-protocol-http",
  "dataDestination": {
    "type": "HttpProxy"
  }
}
```

10.2 Transferencia de activos a almacenamiento s3

La transferencia de activos a un almacenamiento s3 solo modifica el destino de los datos donde se debe indicar los datos de acceso y el endpoint s3 donde se quiere almacenar los datos transferidos.

```
POST /management/v2/transferprocesses HTTP/1.1
Host: localhost:29193
Content-Type: application/json
Authorization: Bearer %Access_token%
Content-Length: 737

{
  "@context": {
    "@vocab": "https://w3id.org/edc/v0.0.1/ns/"
  },
  "@type": "TransferRequestDto",
  "connectorId": "connector-c2",
  "counterPartyAddress": "http://connector-c1:19194/protocol",
  "contractId": "%contractagreementid%",
  "assetId": "c1-s3-asset2",
  "protocol": "dataspace-protocol-http",
  "dataDestination": {
    "properties": {
      "type": "AmazonS3",
      "region": "us-east-1",
      "bucketName": "bucket-connector-2",
      "accessKeyId": "wEfyrv2vH7Wj3ON03g8V",
      "secretAccessKey": "JlHVGgSgeRkbs6WtKTugR4I5uU7oA1VRf8kBtrGd"
    },
    "endpointOverride": "http://minio-c2:9000",
    "type": "AmazonS3"
  },
  "managedResources": true,
  "transferType": {
    "contentType": "text/plain",
    "isFinite": true
  }
}
```

11 Visualización del estado de la transferencia de ficheros

Para visualizar el estado de la solicitud de transferencia de un fichero se pasa el id de la transferencia como el parámetro final de la ruta /management/v2/transferprocesses/. Cuando la transferencia se ha llevado a cabo el servicio retorna el id del contrato.

```
GET /management/v2/transferprocesses/%transfer_id% HTTP/1.1
Host: localhost:29193
Authorization: Bearer %Access_token%
```

12 Conclusiones

Se ha descrito los principales servicios del conector de Inesdata que soportan los flujos de trabajo de los usuarios proveedores y consumidores de activos en el espacio de datos. Sin embargo, para poder desplegar el EDL funcionalmente aún faltan otros servicios que permitan gestionar los planos de datos, activos, políticas y contratos incluyendo operaciones básicas de borrado y actualización. Se espera que en la próxima versión del entregable los componentes Inesdata soporten estas operaciones y puedan integrarse en el documento.