# PATOF
# D3 Metadata Factory

Ding-Ze Hu  Martin Köhler

# D3 Outlook

- **D3 [29 Feb 2024]** - Alpha version of the description of the “FAIR Metadata Factory”
    - Drafted document is expected around the due date.
- **FAIR Metadata Factory** - A framework that allows us to create related or dependent metadata records following a general pattern.
- Define **XML** and **XSD**
    - **XML** file with defined format will be the output of the factory
    - Define **XSD** file to check the validity of the output XML file
- A **Python program** to implement the **Metadata Factory**:
    - accesses and fetches data from experiments.
    - manipulate and engineer the data
    - output to XML metadata record of the data with the defined format

# Metadata Schema – XML

- Use of DataCite: widely used DOI registry
- Design a metadata schema for PATOF – schema based on rules defined by DataCite
- Two XML files. One is for DOI minting which contains basic metadata information. The other one is for actual usage in which apart from the basic metadata information, experiment specific metadata is also included.
    - Metadata included in the one for DOI minting:

      ex. identifier, title, author, subject, …, etc…
    - Metadata included in the one for actual usage:

      ex. identifier, title, author, subject, …, etc… **plus** ex. firmware, DOOCS_property, sampling_rate, etc…

# Metadata Schema – XML

## For actual usage

```
<resource xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://datacite.org/schema/kernel-4" xsi:schemaLocation="http://datacite.org/schema/kernel-4 https://schema.datacite.org/meta/kernel-4.4/metadata.xsd">
        <identifier identifierType="DOI"></identifier>
        <alternateIdentifiers>
                <alternateIdentifier alternateIdentifierType="recid"></alternateIdentifier>
        </alternateIdentifiers>
        <creators>
                <creator>
                        <creatorName nameType="Organizational"></creatorName>
                </creator>
        </creators>
        <titles>
                <title></title>
        </titles>
        <publisher>(to be determined when register the object)</publisher>
        <publicationYear></publicationYear>
        <resourceType resourceTypeGeneral="Dataset"></resourceType>
        <subjects>
                <subject subjectScheme="experiment"></subject>
                <subject subjectScheme="experiment"></subject>
                <subject subjectScheme="collaboration"></subject>
                <subject subjectScheme="collaboration"></subject>
                <subject subjectScheme="keyword"></subject>
        </subjects>
        <contributors>
                <contributor contributorType="Distributor">
                        <contributorName></contributorName>
                </contributor>
                <contributor contributorType="HostingInstitution">
                        <contributorName></contributorName>
                </contributor>
                <contributor contributorType="Distributor">
                        <contributorName></contributorName>
                </contributor>
                <contributor contributorType="ContactPerson">
                        <contributorName></contributorName>
                        <givenName></givenName>
                        <familyName></familyName>
                        <nameIdentifier nameIdentifierScheme="ORCID" schemeURI="http://orcid.org"></nameIdentifier>
                </contributor>
        </contributors>
```

```
−<column12>
    <name rel="rdfs:seeAlso" resource="https://doi.org/10.1351/goldbook.P04712">y_0 for polarization 0</name>
    <units rel="owl:sameAs" property="qudt:Number" resource="https://doi.org/10.1351/goldbook.C01370">count</units>
    <meanValue>78005.29086282304</meanValue>
    <rangeBottom>216.166</rangeBottom>
    <rangeTop>146147.084</rangeTop>
  −<dataset type="xsd:float">
      <value>34889.059</value>
      <value>57867.741</value>
      <value>48403.713</value>
      <value>76495.574</value>
      <value>97206.345</value>
      <value>120959.259</value>
      <value>104223.584</value>
      <value>42705.002</value>
      <value>62608.645</value>
      <value>57638.797</value>
      <value>75986.432</value>
```

**Deep and useful information of dataset**

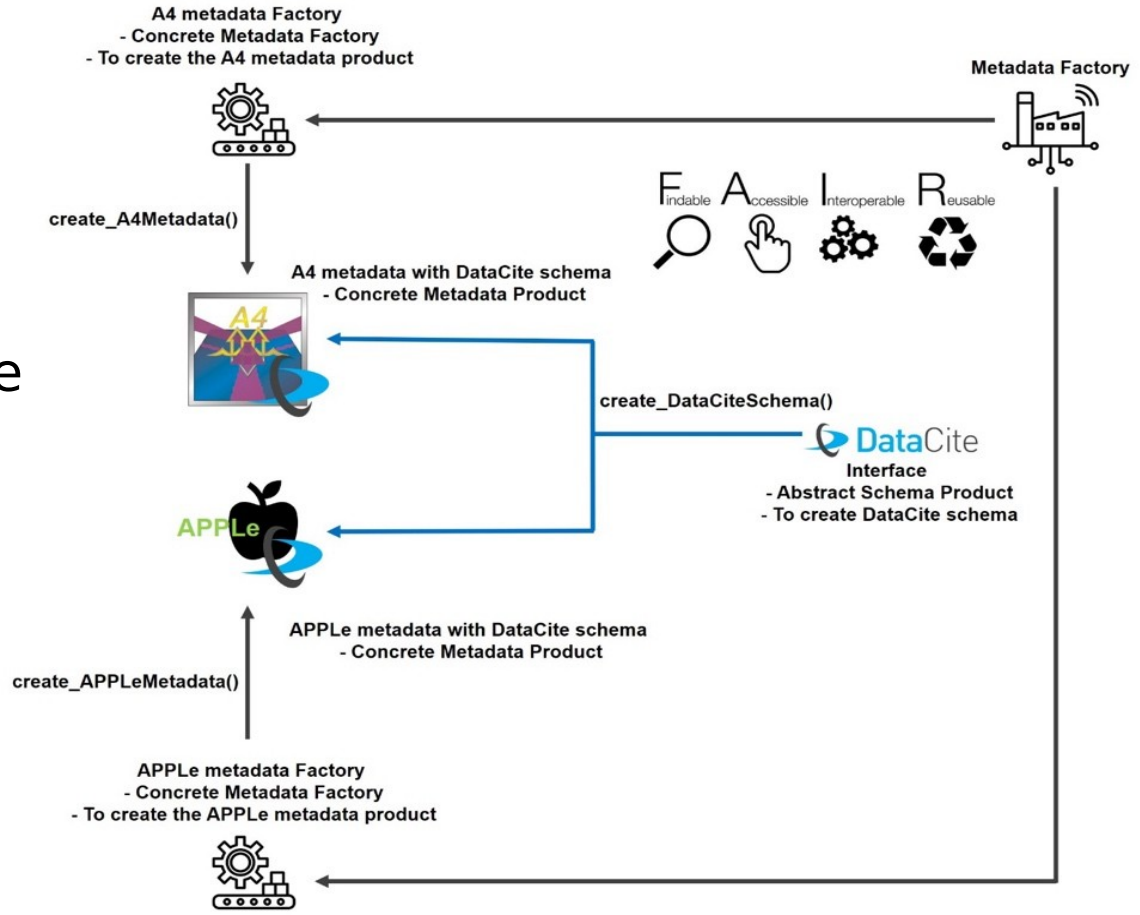
## For DOI minting

```
<resource xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://datacite.org/schema/kernel-4" xsi:schemaLocation="http://datacite.org/schema/kernel-4 https://schema.datacite.org/meta/kernel-4.4/metadata.xsd">
        <identifier identifierType="DOI"></identifier>
        <alternateIdentifiers>
                <alternateIdentifier alternateIdentifierType="recid"></alternateIdentifier>
        </alternateIdentifiers>
        <creators>
                <creator>
                        <creatorName nameType="Organizational"></creatorName>
                </creator>
        </creators>
        <titles>
                <title></title>
        </titles>
        <publisher>(to be determined when register the object)</publisher>
        <publicationYear></publicationYear>
        <resourceType resourceTypeGeneral="Dataset"></resourceType>
        <subjects>
                <subject subjectScheme="experiment"></subject>
                <subject subjectScheme="experiment"></subject>
                <subject subjectScheme="collaboration"></subject>
                <subject subjectScheme="collaboration"></subject>
                <subject subjectScheme="keyword"></subject>
        </subjects>
        <contributors>
                <contributor contributorType="Distributor">
                        <contributorName></contributorName>
                </contributor>
                <contributor contributorType="HostingInstitution">
                        <contributorName></contributorName>
                </contributor>
                <contributor contributorType="Distributor">
                        <contributorName></contributorName>
                </contributor>
                <contributor contributorType="ContactPerson">
                        <contributorName></contributorName>
                        <givenName></givenName>
                        <familyName></familyName>
                        <nameIdentifier nameIdentifierScheme="ORCID" schemeURI="http://orcid.org"></nameIdentifier>
                </contributor>
        </contributors>
```

Validate the XML metadata records against the defined XSD, to ensure that XML is valid and follows the schema we want

# Metadata Factory

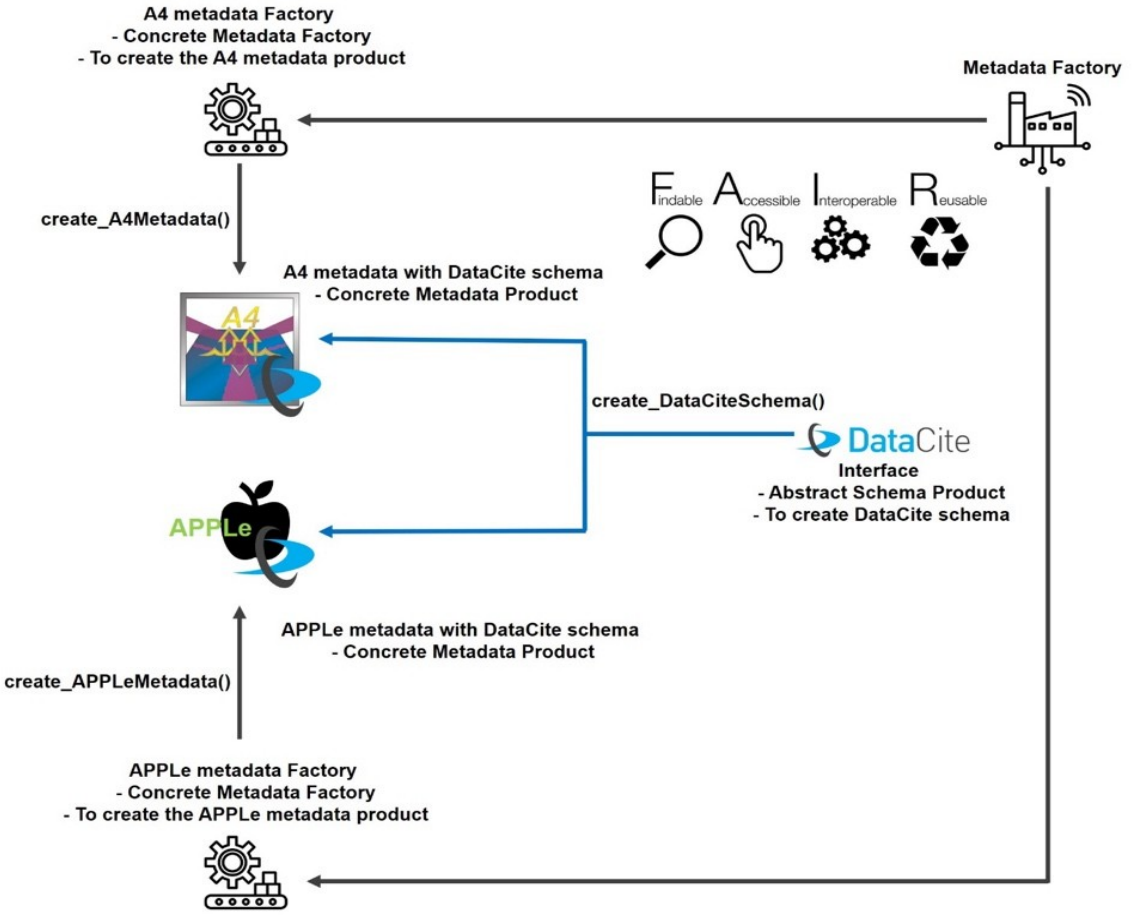
- **Metadata Factory** Pattern, a framework that allows us to create related or dependent products that follows a general pattern.
- DataCite as the **core** in metadata schema
- **Experiment-specific extension** for the varied  metadata fields needed in individual experiment.
- Metadata in general follows the schema from DataCite. Make changes and add additional stuff to the metadata according to the dataset from individual experiments.



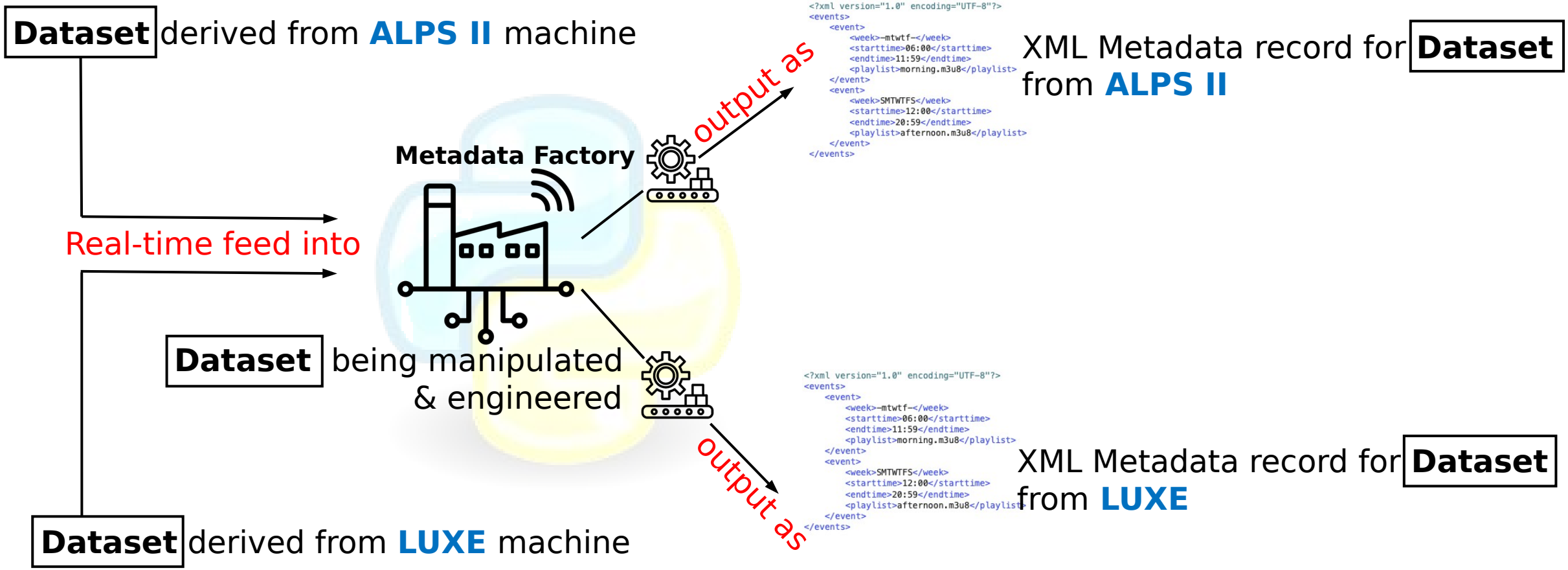
# Metadata Factory

Assume that we are design a family of two products.

- **Abstract Product**: One abstract class is created, for creating DataCite schema. This class contains abstract methods that are mandatory for the construction of the products. This abstract class is referred to as interfaces.
- **Concrete Products**: Concrete products inherit the abstract methods from the abstract class i.e. abstract product. Using the interface, different families of products can be created.
- **Concrete Factories**: Concrete Factories create Concrete Products as directed by the Abstract Factory. The concrete factories are only capable of creating those products that are specified in them – **A4 metadata factory** creates A4 metadata, while **APPLe metadata factory** creates APPLe metadata.
- **Abstract Factory:** The Abstract factory possesses interface to create the abstract product i.e. they contain several methods that return abstract product.



# Workflow for the Python program (Metadata Factory)



# !!Example only!! Python syntax for the factory

```python
class patofSchema(ABC):
    """
    The Abstract Factory
    """

    @abstractmethod
    def create_patof_schema(self):
        pass

class alps2ProductsFactory(patofSchema):
    """
    Type: Concrete Factory
    Implement the operations to create concrete product objects.
    """

    def create_patof_schema(self):
        return alps2_data_with_patof_schema()

class luxeProductsFactory(AbstractFactory):
    """
    Type: Concrete Factory
    Implement the operations to create concrete product objects.
    """

    def create_patof_schema(self):
        return luxe_data_with_patof_schema()
```

# D5 Outlook

- **D5 [31 Dec 2024]** - “Cookbook” describing a general “FAIR Metadata Factory” based on D3
- A living cookbook can be used for future experiments and even different discipline
- Will start in the future when the Metadata Factory is developed to a certain level.

# Thank you for your attention

Welcome any feedback!

Ding-Ze Hu
DESY-Library
ding-ze.hu@desy.de

Martin Köhler
DESY-Library
martin.koehler@desy.de