

**Journée Intéropérabilité et Innovation :**  
**Le FAIR data en appui à la compréhension**  
**de la Terre et de son environnement**  
**Atelier thématique : DWG métiers OGC**

**Université Gustave Eiffel - Bâtiment Bienvenue**  
**16 Av. Newton, 77420 Champs-sur-Marne, France**  
**23-01-2024**

OSDU Forum ( <https://osduforum.org> ) + RESQML (Energistics)

Is it A F.A.I.R. Solution ?

## OSDU Forum Mission

Working with The Open Group to create  
certifiable standards to support interoperability

Using the Open Source model  
to promote collaboration

The Open Group OSDU Forum delivers an **Open Source**,  
**standards-based**, **technology-agnostic** data platform for the  
**energy industry** that stimulates innovation, industrializes  
data management, and reduces time to market for new  
solutions.

Starting in Subsurface Oil and Gas,  
but extending to support the wider  
needs of Energy Companies

"Polycoud" – runs on Public Clouds  
and on premises, taking advantage  
of the technology available on each

Not a data storage solution or a data  
management application - but a  
modern, cloud-native data platform

Copyright © The Open Group 2021

12 **OSDU**

# OSDU Members

The OSDU Forum currently has 220+ Member Companies with a diverse set of goals and objectives:

- Energy operators
- Software & technology companies
- Service companies
- Research organizations
- Academia
- And more...

[Current Members | The Open Group Website](#)

## OPERATOR MEMBERS

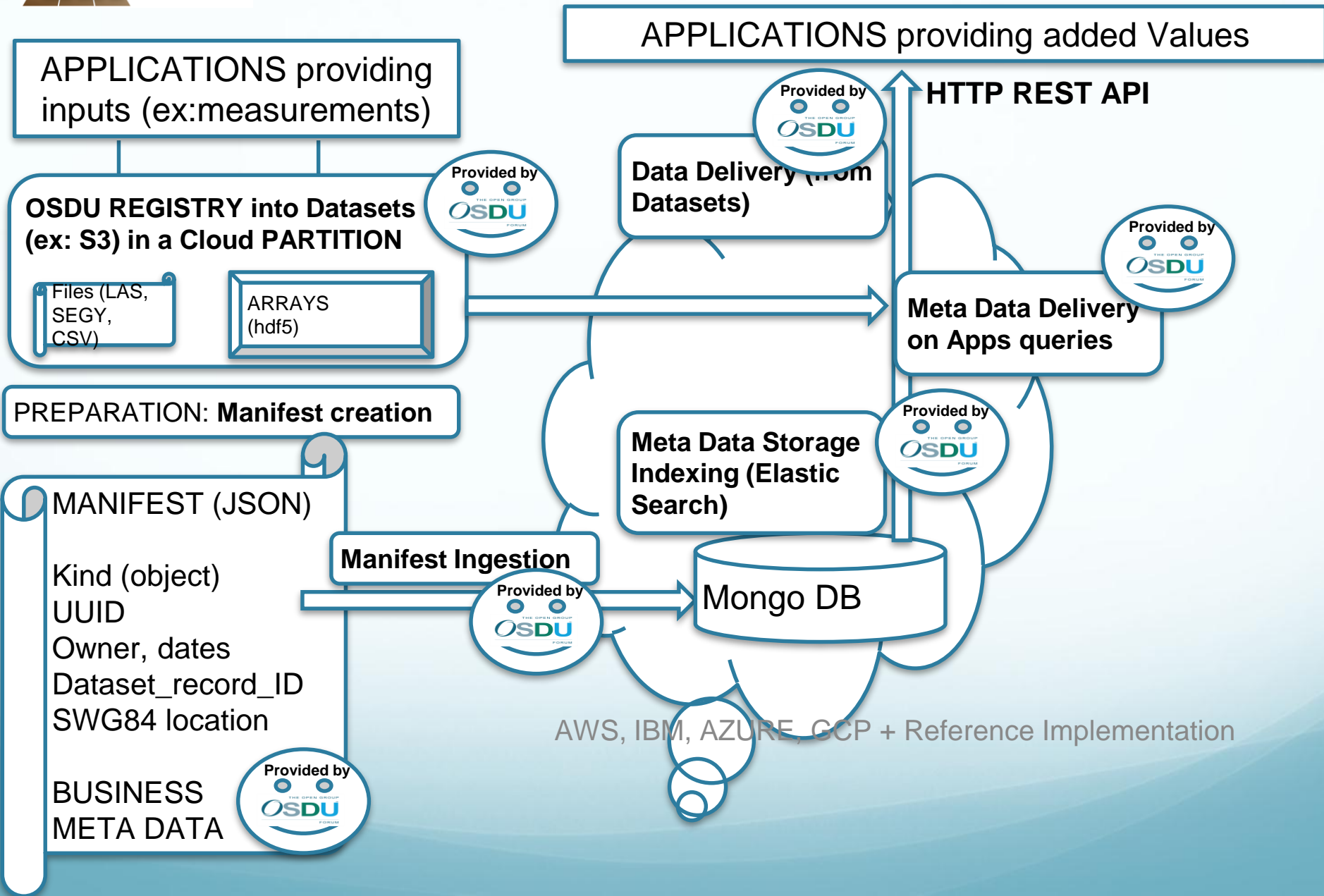


## SUPPLIER MEMBERS

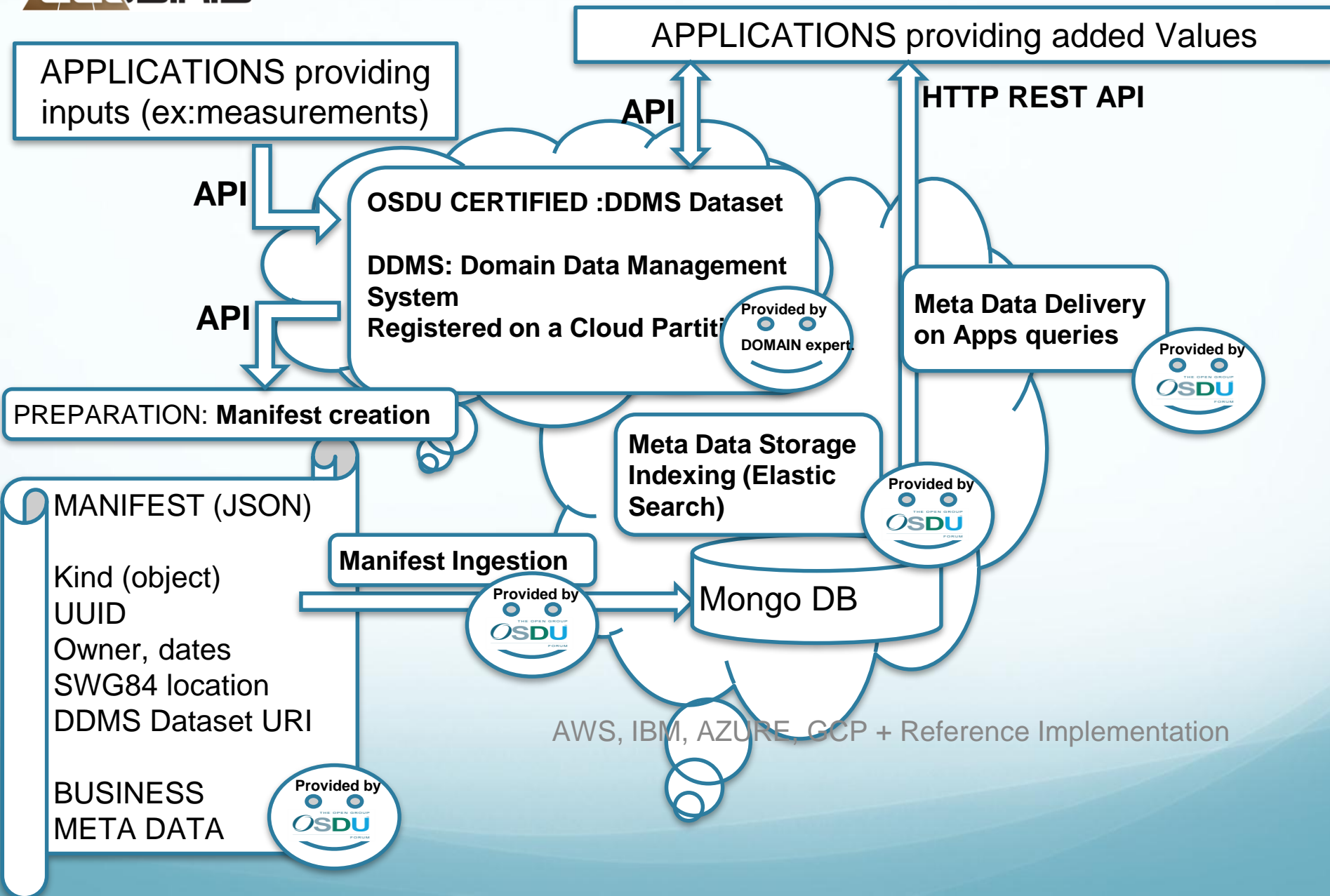




# How OSDU Platform works (from files)



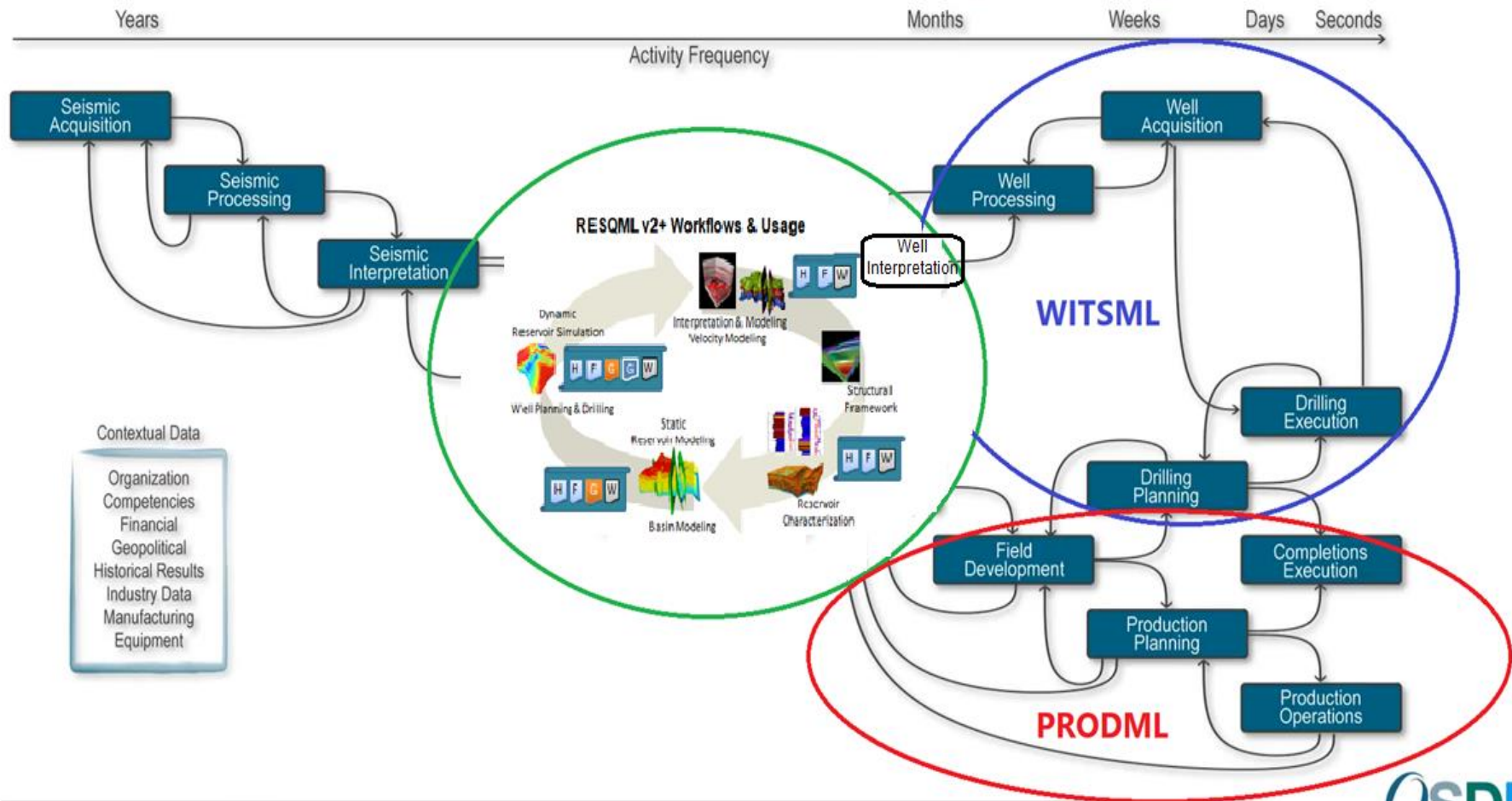
# How OSDU Platform works (from DDMS API)







# OSDU Initial data footprint driving design





# Reminder :The Energistics Data Concepts

A **Generic** Data Model To Cover 3D Geoscience Earth Modeling (including Geotechnics)

Based on 4 Main components (Energistics's Objects):

**Features:** The Object's Reference : The key to access the meta data and the data.

- 2 categories of Geologic Features : Boundary, RockVolume,

- 3 Categories of Technical Features : WellBore, Cultural, SeismicSurvey

- 3 Categories of Organizations (geol features assembled in Earth-models)

  - Structural, Stratigraphic, RockCompartments

**Interpretations:** The knowledge on the Object ( Meta Data )

**Representations:** The Indexing scheme (Topology) to handle the Geometry of the object :

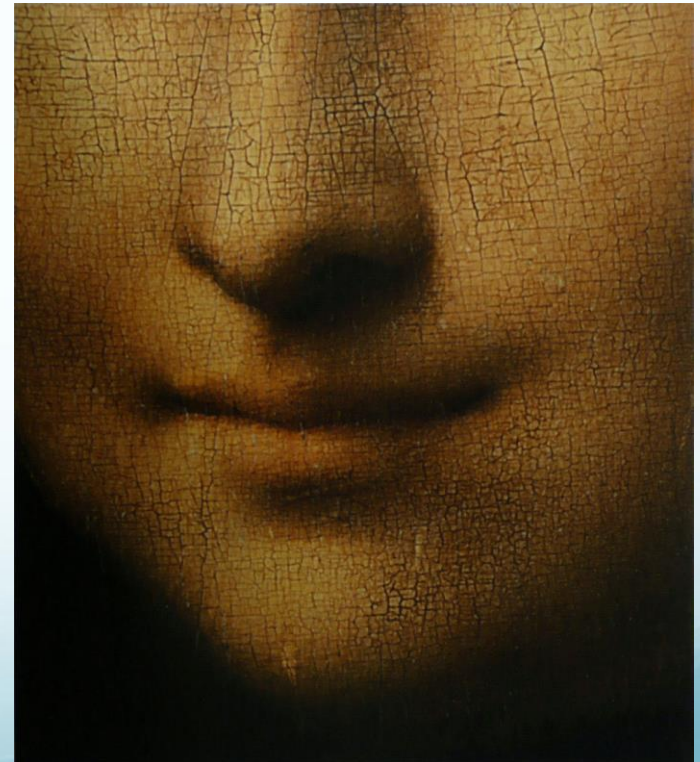
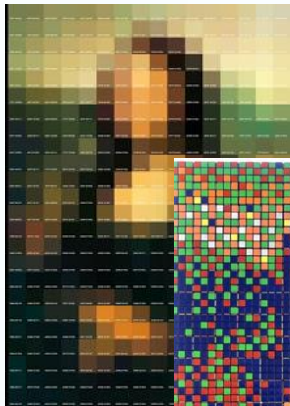
- points, polylines, Triangulated or 2D grid surfaces, discretized or implicit 3D Volumes

**Properties:** A range of many property Kind with their unit Of Measure, of great number of domains using the representation indexing scheme to associate Values to Interpretations .

## ***Why this concept of Interpretation ?***

### ***Which does not exists elsewhere***

- Improve consistency of the Earth & Reservoir models (meta-data / data)
- Clarify and reduce geoscience study cycles
- Allow to follow up the Audit trail of Earth Models & Reservoir development decisions







# RESQML : How exchange with the Applications

## 1/ By Using Files :

An Energistic object is an XML File today (or a JSON File tomorrow morning)

- An EPC (Energistics Package Container) is a zip file containing these XML files
- An hdf5 File (OGC, FAIR compatible) is delivered with this EPC.
- Apps can read and Write EPC+Hdf5 (C++, C#, Java, Python APIs).

## 2/ By using Websocket based ETP Protocol (Energistic Transfer Protocol)

Allows the Apps to connect to an **ETP server** on an URL:

- Provides a dataspace concept ( equivalent to EPC)
- Provides GET/PUT/RETRIEVE/DELETE RESQML Object exchange (XML or JSON)
- Provides GET/PUT/RETRIEVE/DELETE Binary Arrays exchange ( hdf5 Datasets)

## ETP : What is it?

- ETP is a communication protocol i.e. precise set of rules for the exchange of data between endpoints, in the form of messages organized into subprotocols.
- ETP messages:
  - Are defined in Avro schemas and serialized using Avro
  - Are transported in accordance with the message framing of the WebSocket (WS) protocol

## ETP domain resources

- ETP messages transport domain resources such as a well, horizon, fault , stratigraphic grid, perforation, PVT, etc..
- Resources are simply transported in an AVRO message byte array field. It contains RESQML or WITSML or PRODML format encoded resource (today in XML, tomorrow possibly in JSON)



# RESQML/OSDU Data Definition mapping

**The OSDU Earth and Reservoir Domain is totally based Energistics's Data Model**

**RESQML Features** are **Master Data** in OSDU Data Definition

**RESQML Interpretations** are **Work Product Components** in OSDU Data Definition

**RESQML Representations :**

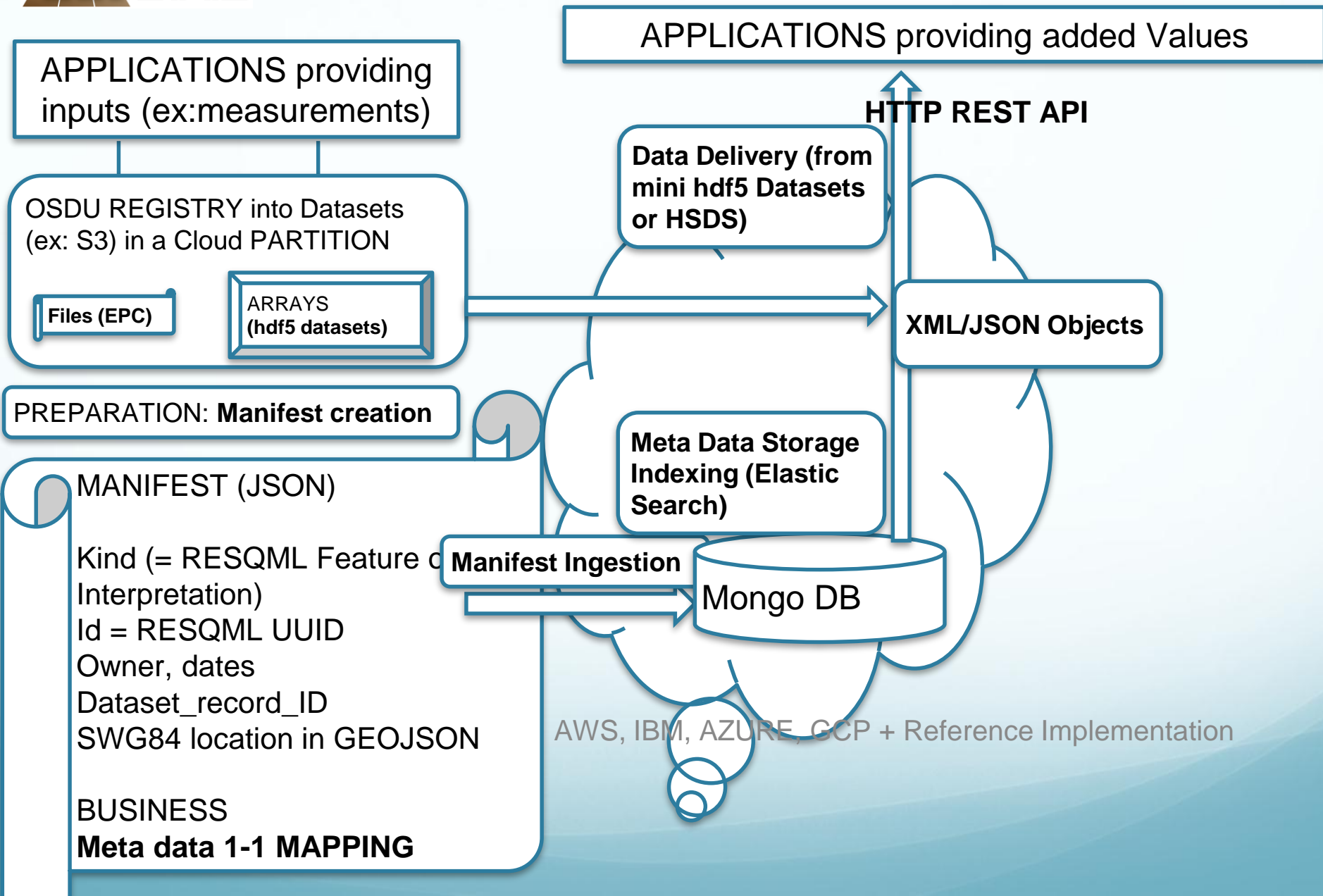
- Only the characterization of the Representation : points, polylines, Triangulated or 2D grid surfaces, discretized or implicit 3D Volumes are indicated in the OSDU data definition
- The Values are send from an hdf5 file or an HSDS http REST API for files or from ETP Data Arrays.

**RESQML Properties :**

- only the property Kinds are indicated in the OSDU data definition, The Values are send from an hdf5 file or an HSDS http REST API for files or from ETP Data Arrays.



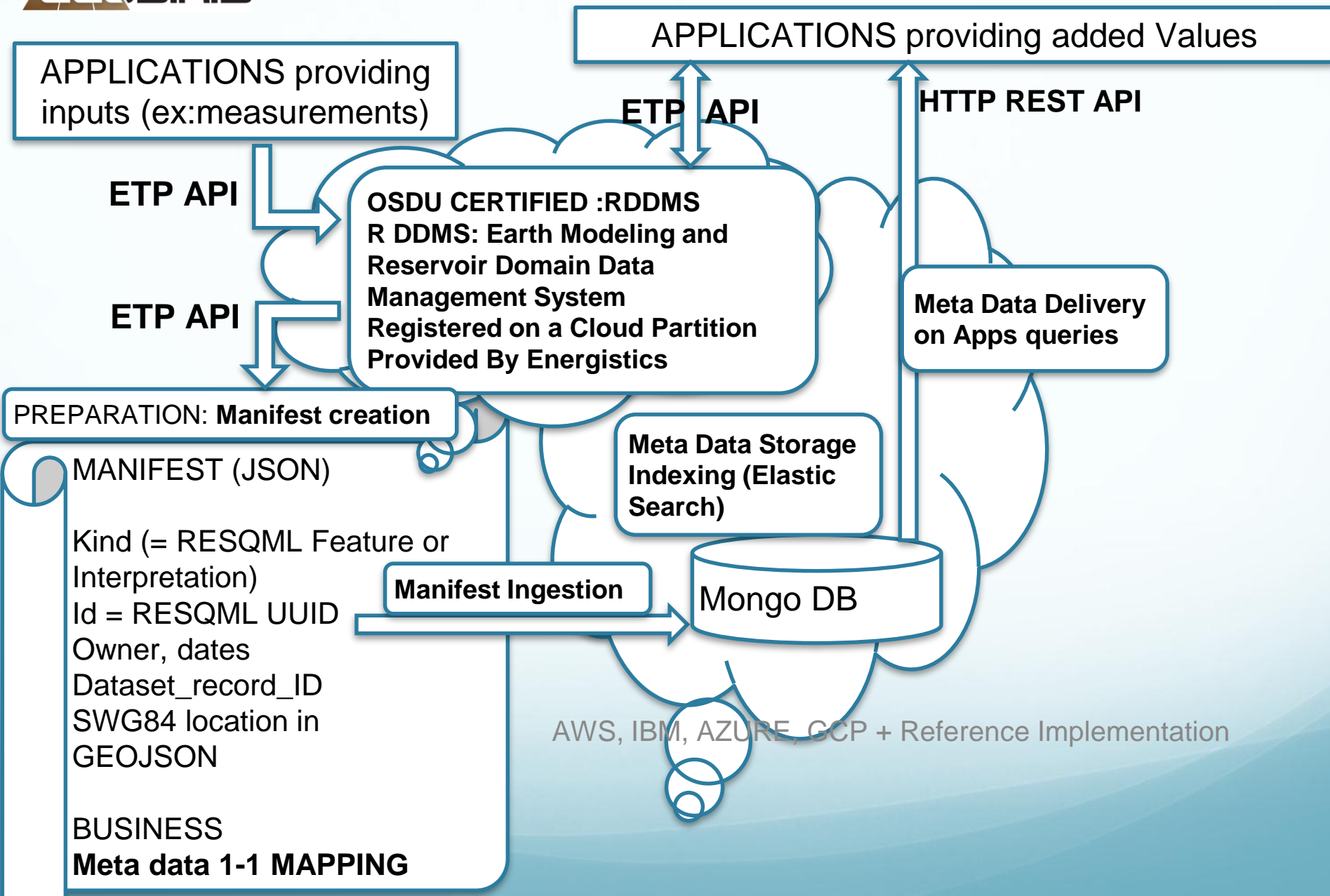
# How OSDU and RESQML are linked (files)







# How OSDU Platform and ETP Server are linked





# OSDU/RESQML F.A.I.R Analysis

Total across FAIR

12/12 Answered

Q1

Q2

Q3

Q4

Q5

Q6

Q7

Q8

Q9

Q10

Q11

Q12

## Findable

The data has sufficiently rich metadata and a unique and persistent identifier to be easily discovered by others. This includes assigning a persistent identifier (like a DOI or Handle), having rich metadata to describe the data and making sure it is findable through disciplinary local or international discovery portals.



Does the dataset have any identifiers assigned?



What is this?



Globally unique, citable, and persistent (e.g. DOI, PURL, ARK or Handle)



Web Address (URL)



Local Identifier



No Identifier



Is the dataset identifier included in all metadata records/files describing the data?



Yes



No



How is the data described with metadata?



Comprehensively using a formal machine-readable metadata schema



Comprehensively, but in a text-based, non-standard format



Brief title and description



The data is not described



What type of repository or registry is the metadata record in?



Data is in one place but discoverable through several registries



Generalist public repository



Domain-specific repository



Local institutional repository



The data is not described in any repository

## Accessible

The data is retrievable by humans and machines through a standardised communication protocol, with authentication and authorisation where necessary. The data does not necessarily have to be open. Data can be sensitive due to privacy concerns, national security or commercial interests. When it's not able to be open, there should be clarity and transparency around the conditions governing access and reuse.



How accessible is the data?



What is this?

☐ Publicly accessible

☐ Fully accessible to persons who meet explicitly stated conditions, e.g. ethics approval for sensitive data

☒ A de-identified / modified subset of the data is publicly accessible

☐ Embargoed access after a specified date

☐ Unspecified conditional access e.g. contact the data custodian for access

☐ Access to metadata only

☐ No access to data or metadata



Is the data available online without requiring specialised protocols or tools once access has been approved?

☐ Standard web service API (e.g. OGC)

☒ Non-standard web service (e.g. OpenAPI/Swagger/informal API)

☐ File download from online location

☐ By individual arrangement

☐ No access to data



Will the metadata record be available even if the data is no longer available?

☒ Yes

☐ No

☐ Unsure

# OSDU/RESQML F.A.I.R Analysis

## Interoperable

The associated data and metadata uses a 'formal, accessible, shared, and broadly applicable language for knowledge representation'. This involves using community accepted languages, formats and vocabularies in the data and metadata. Metadata should reference and describe relationships to other data, metadata and information through identifiers.



What (file) format(s) is the data available in?



What is this?



In a structured, open standard, machine-readable format



In a structured, open standard, non-machine-readable format



Mostly in a proprietary format



What best describes the types of vocabularies/ontologies/tagging schemas used to define the data elements?



Standardised open and universal using resolvable global identifiers linking to explanations



Standardised vocabularies/ontologies/schema without global identifiers



No standards have been applied in the description of data elements



Data elements not described



How is the metadata linked to other data and metadata (to enhance context and clearly indicate relationships)?



Metadata is represented in a machine readable format, e.g. in a linked data format such as Resource Description Framework (RDF).



The metadata record includes URI links to related metadata, data and definitions



There are no links to other metadata




## Reusable

The associated metadata provides rich and accurate information, and the data comes with a clear usage licence and detailed provenance information. Reusable data should maintain its initial richness. For example, it should not be diminished for the purpose of explaining the findings in one particular publication. It needs a clear machine readable licence and provenance information on how the data was formed. It should also use discipline-specific data and metadata standards to give it rich contextual information that will allow reuse.



Which of the following best describes the license/usage rights attached to the data?

 What is this?



Standard machine-readable license (e.g. Creative Commons)



Standard text based license



Non-standard machine-readable license (clearly indicating under what conditions the data may be reused)



Non-standard text-based license



No license



How much provenance information has been captured to facilitate data reuse?



Fully recorded in a machine readable format



Fully recorded in a text format



Partially recorded



No provenance information is recorded





L'association OSDU /RESQML  
semble avoir toutes les qualités  
d'une « FAIR Proposal »

Merci de votre attention !



# annexe

## Les 14 principes, en résumé:

			
<ul style="list-style-type: none"> <li>identifiant pérenne</li> <li>métadonnées enrichies</li> <li>données cherchables et trouvables en ligne</li> </ul>	<ul style="list-style-type: none"> <li>données récupérables en ligne par des protocoles standardisés</li> <li>restrictions si nécessaire</li> </ul>	<ul style="list-style-type: none"> <li>formats et standards communs</li> <li>vocabulaires contrôlés</li> </ul>	<ul style="list-style-type: none"> <li>données documentées (pex. avec un fichier README)</li> <li>licence et information de provenance claires</li> </ul>
<p>Si le jeu de données ne peut pas être mis à disposition librement, il est recommandé qu'au minimum sa description soit librement accessible en ligne.</p>	<p>Dans un repository qui stocke les données de manière sécurisée</p>	<p>Pour être compatibles et combinables avec d'autres jeux de données</p>	<p>Afin que les données soient interprétées correctement</p>



# FAIR DATA PRINCIPLES [1]

Data and metadata are **easy to find** by both humans and computers.

## F

### FINDABLE

- F1** (Meta)data are assigned a globally unique and persistent identifier.
- F2** Data are described with rich metadata.
- F3** Metadata clearly and explicitly include the identifier of the data they describe.
- F4** (Meta)data are registered or indexed in a searchable resource.

### DESCRIBE

Describe provenance, usage and organization of data with standardized **metadata** (RDA standards, DataCite, DublinCore). Make metadata available **even if** data is not.

Both humans and computers can **readily access** or download datasets.

## A

### ACCESSIBLE

- A1** (Meta)data are retrievable by their identifier using a standardized communication protocol:
  - A1.1** the protocol is open, free and universally implementable;
  - A1.2** the protocol allows for an authentication and authorization procedure where necessary.
- A2** Metadata are accessible, even when the data are no longer available.

### OPEN

Open your data using standardized **licenses**. **Limitations** may apply to the openness (ex. embargo). Disclose files in **open formats**, even alongside proprietary formats.

Data from different datasets are **prepared to be combined** or exchanged.

## I

### INTEROPERABLE

- I1** (Meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation.
- I2** (Meta)data use vocabularies that follow FAIR principles.
- I3** (Meta)data include qualified references to other (meta)data.

### LINK

Use persistent **identifiers** for datasets (ex. DOI, HANDL, URN) and tag all the metadata with the **same** identifiers. **Cross-link** datasets with linked-data standards (RDF).

Published data can be **easily combined / replicated** in future research.

## R

### REUSABLE

- R1** (Meta)data are richly described with a plurality of accurate and relevant attributes:
  - R1.1** (meta)data are released with a clear and accessible data usage license;
  - R1.2** (meta)data are associated with detailed provenance;
  - R1.3** (meta)data meet domain-relevant community standards.

### PUBLISH

Deposit datasets in data **repositories**, favoring services with user-friendly **interfaces**. Make sure to chose a **FAIR-compliant** data repository, also for the relative code.

# Acceptable Archiving Formats

When selecting file formats for archiving, the formats should ideally be: Non-proprietary, unencrypted, uncompressed, commonly used by the research community

- Compliant to an open, documented standard: interoperable among diverse platforms and applications, fully published and available royalty-free, fully and independently implementable by multiple software providers on multiple platforms without any intellectual property [2]

TYPE OF DATA	APPROPRIATE	ACCEPTABLE	DEPRECATED
Tabular (extensive metadata)	CSV – HDF5	TXT – HTML – TEX – FASTQ <sup>[3]</sup> – POR	
Tabular (minimal metadata)	CSV – TAB – ODS – SQL – TSV	XML (if appropriate DTD) – XLSX	XLS – XLSB
Textual / Presentation	TXT – PDF – ODT – ODM – TEX – MD – HTM – XML – EXTXYZ <sup>[4]</sup> – ODF	PPTX – RTF – DOCX – PDF (with embedded forms) – EPS – IPF	DOC – PPT – DVI – PS
Code / Computation	M – R – PY – IYPNB – RSTUDIO – RMD – NETCDF – AIML	SDD	MAT – RDATA
Image & Spectroscopy	TIF – PNG – SVG – JPEG – FITS	JCAMP – JPG – JP2 – TIF – TIFF – PDF – GIF – BMP – DM3 – OIR – LSM <sup>[5]</sup>	INDD – AIT – PSD – SPC
Audio	FLAC – WAV – OGG – MXL – MIDI – MEI – HUMDRUM	MP3 – AIF	
Video	MP4 – MJ2 – AVI – MKV	OGM – MP4 – WEBM	WMV – MOV – QT
Geospatial	NETCDF – tabular GIS attribute data – SHP – SHX – DBF – PRJ – SBX – SBN – POSTGIS – TIF – TFW – GEOJSON	MDB – MIF	
3D structures & images	X3D – X3DV – X3DB – PDF3D – POV – PDBML	DWG – DXF – PDB	PXP
Generic	XML – JSON – RDF		