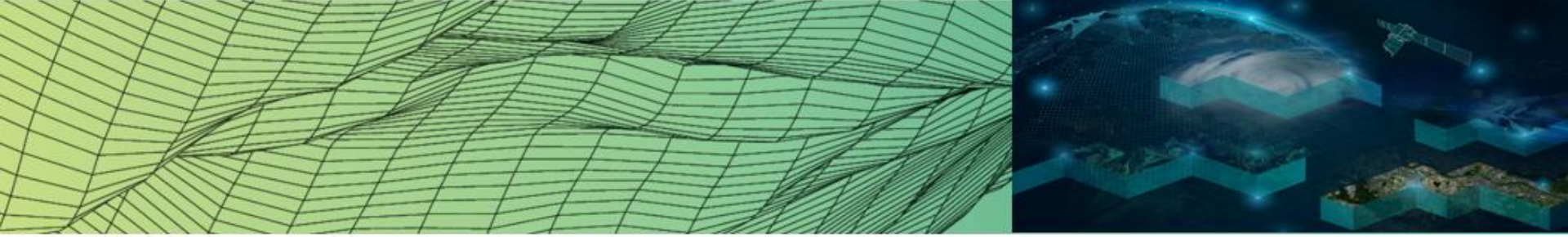


Journée Intéropérabilité et Innovation :
Le FAIR data en appui à la compréhension de la Terre et de son environnement
Atelier thématique : DWG métiers OGC

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.AI.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

"Polycloud" – 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

Jean François Rainaud Jean-francois.Rainaud@geosiris.com

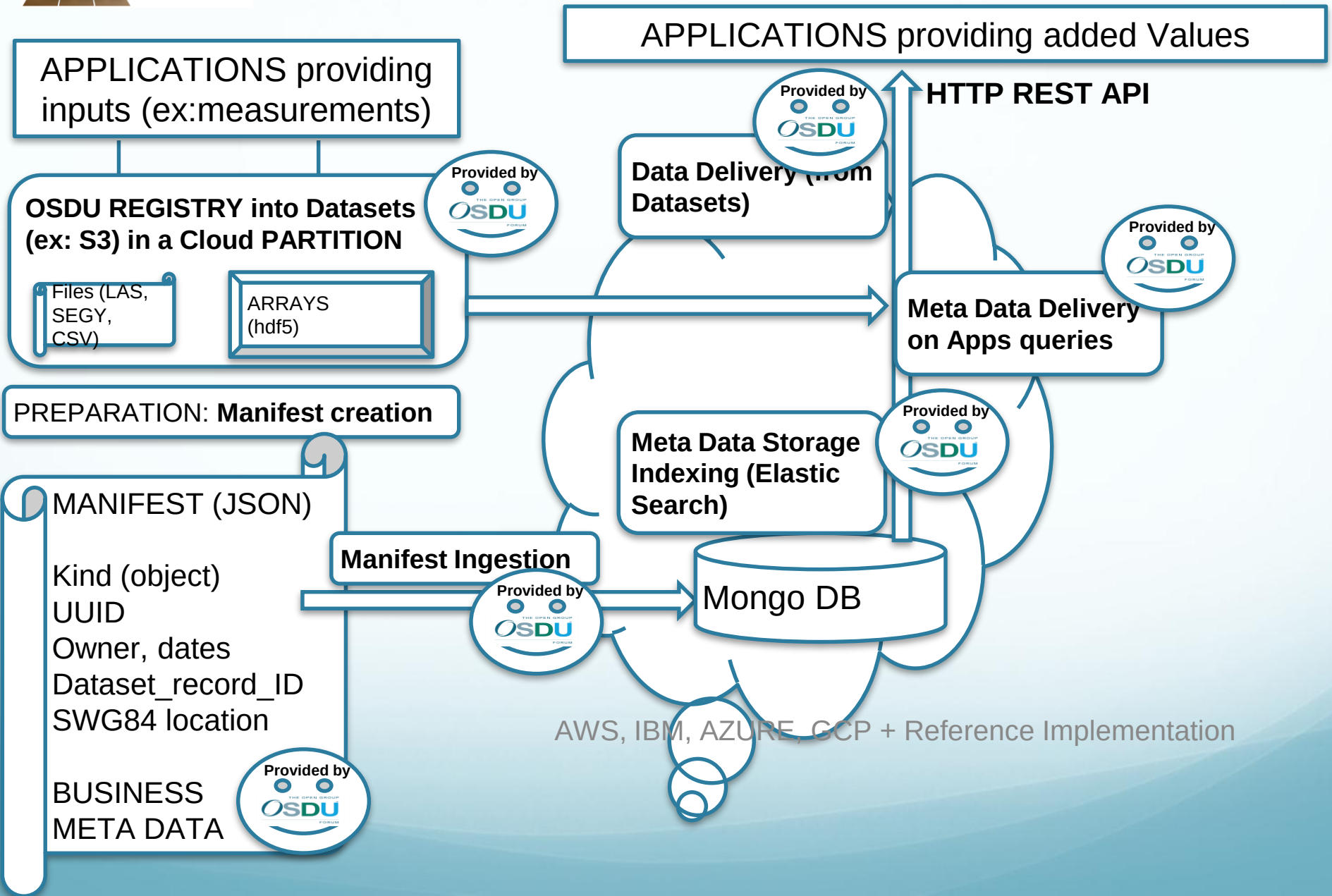
OSDU Members

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

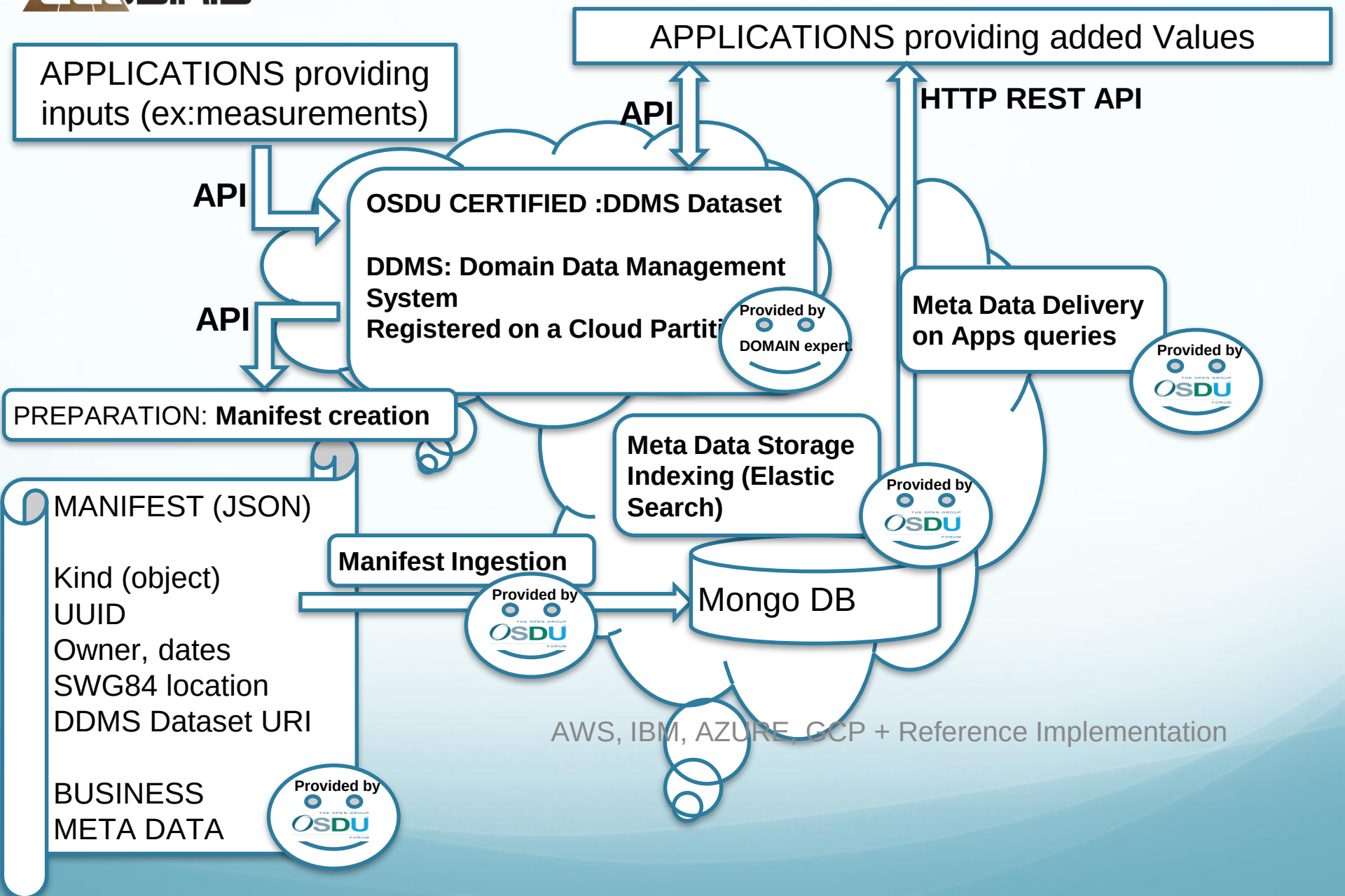
- Energy operators
- Software & technology companies
- Service companies
- Research organizations
- Academia
- And more…

[Current Members | The Open Group Website]

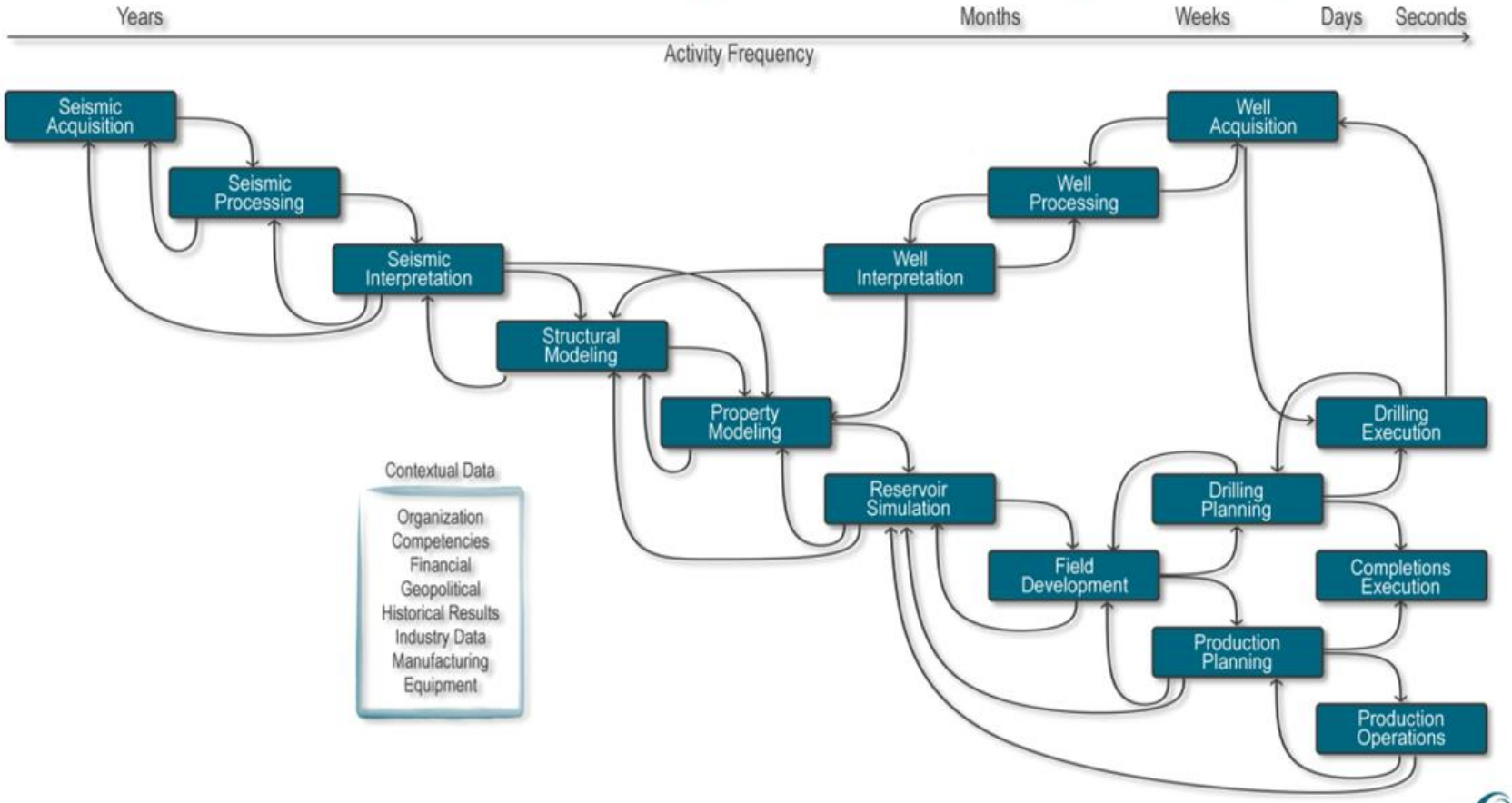
How OSDU Platform works (from files)



How OSDU Platform works (from DDMS API)

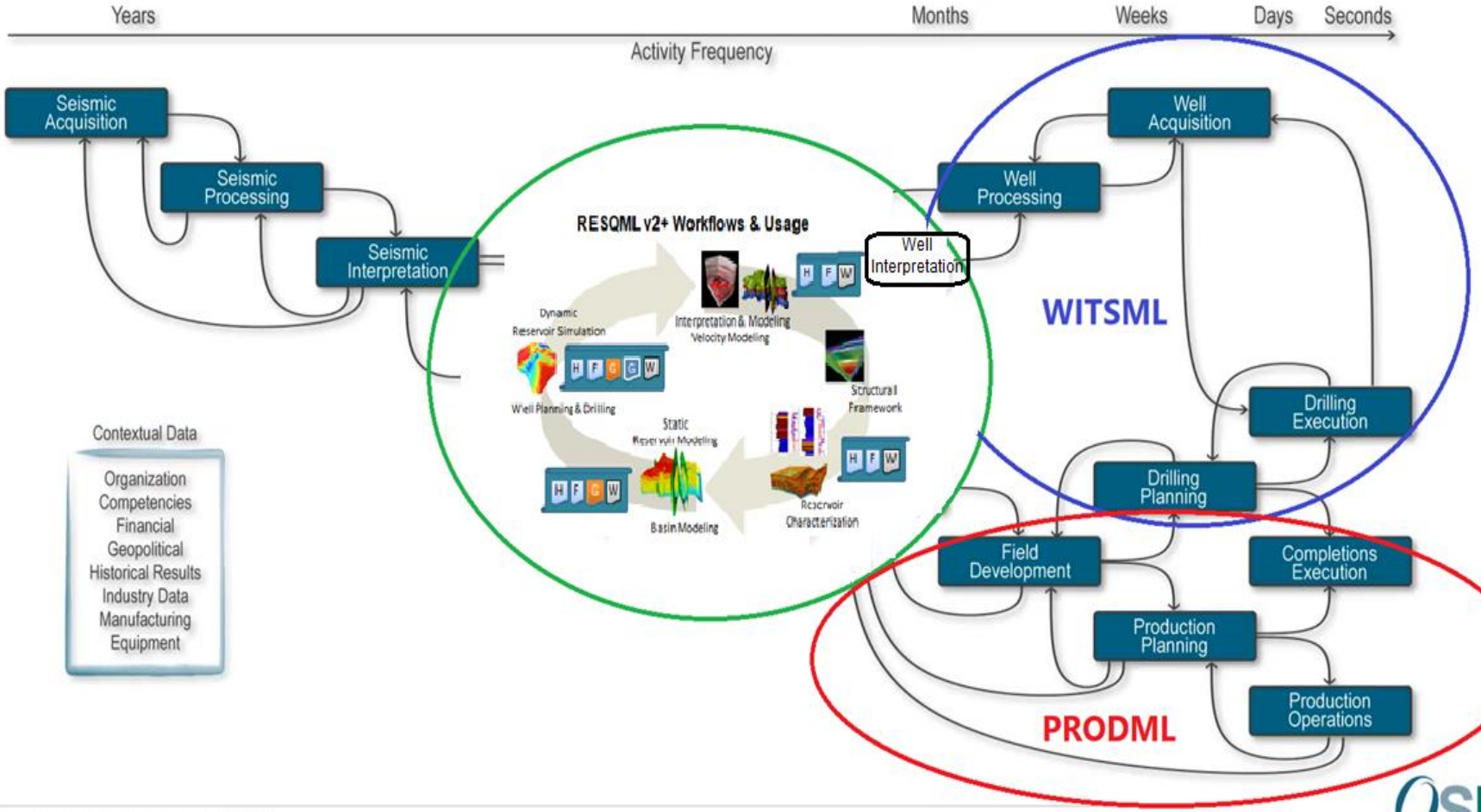


OSDU Initial data footprint driving design



Copyright © The Open Group 2021

OSDU Initial data footprint driving design



Copyright © The Open Group 2021

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)

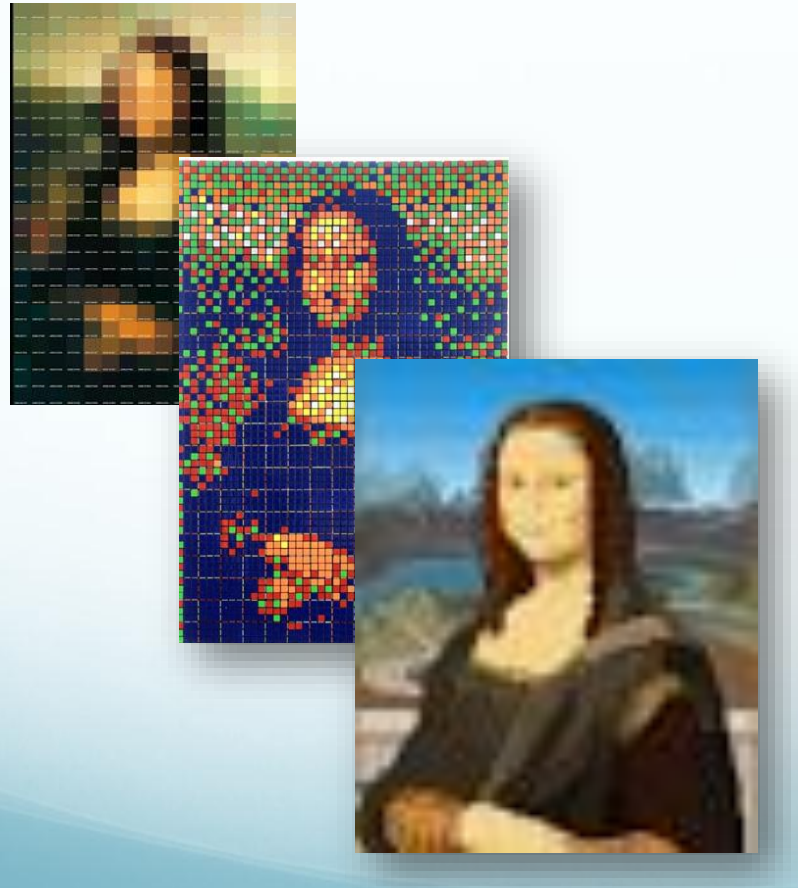
Representations: The Indexing scheme (Topology) to handle the Geometry of the object :
points, polylines, Triangulated or 2D grid surfaces, discretized or implicit 3D Volumes

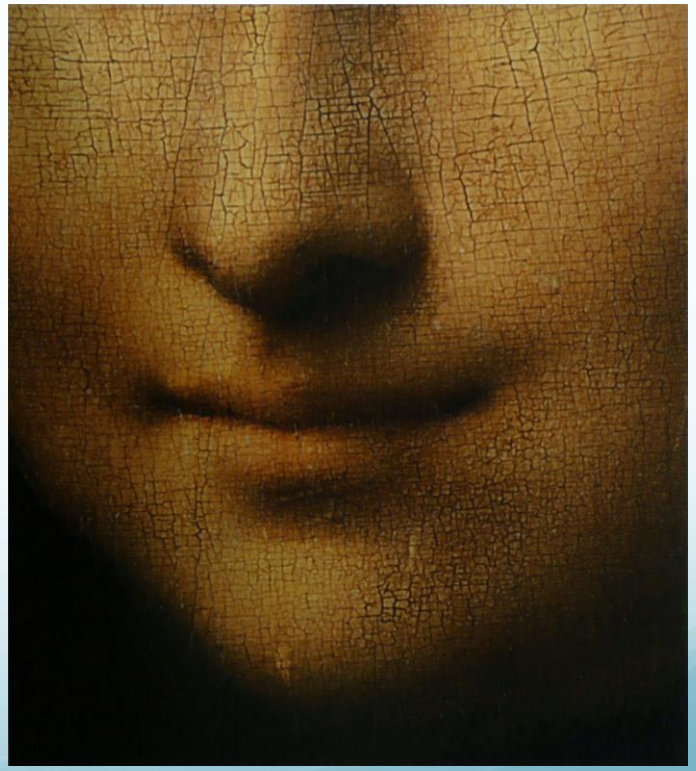
Properties: A range of many property Kind with their unit Of Measure, of great number of domains using the representation indexing scheme to associate Values to Interpretations .

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 : Energistics Transfer Protocol

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)

Copyright © The Open Group 2021

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

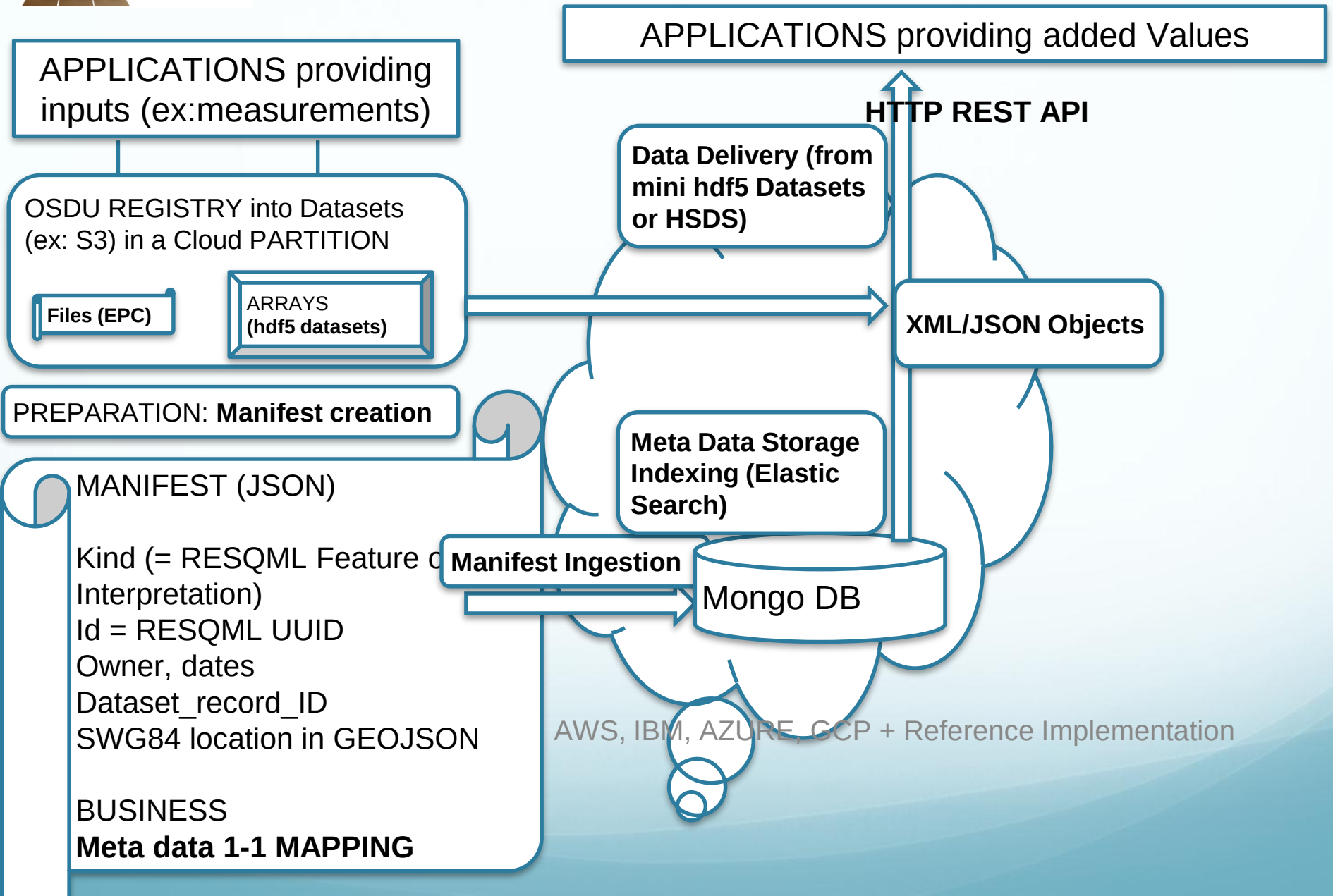
RESQML **Representations :**

- 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.

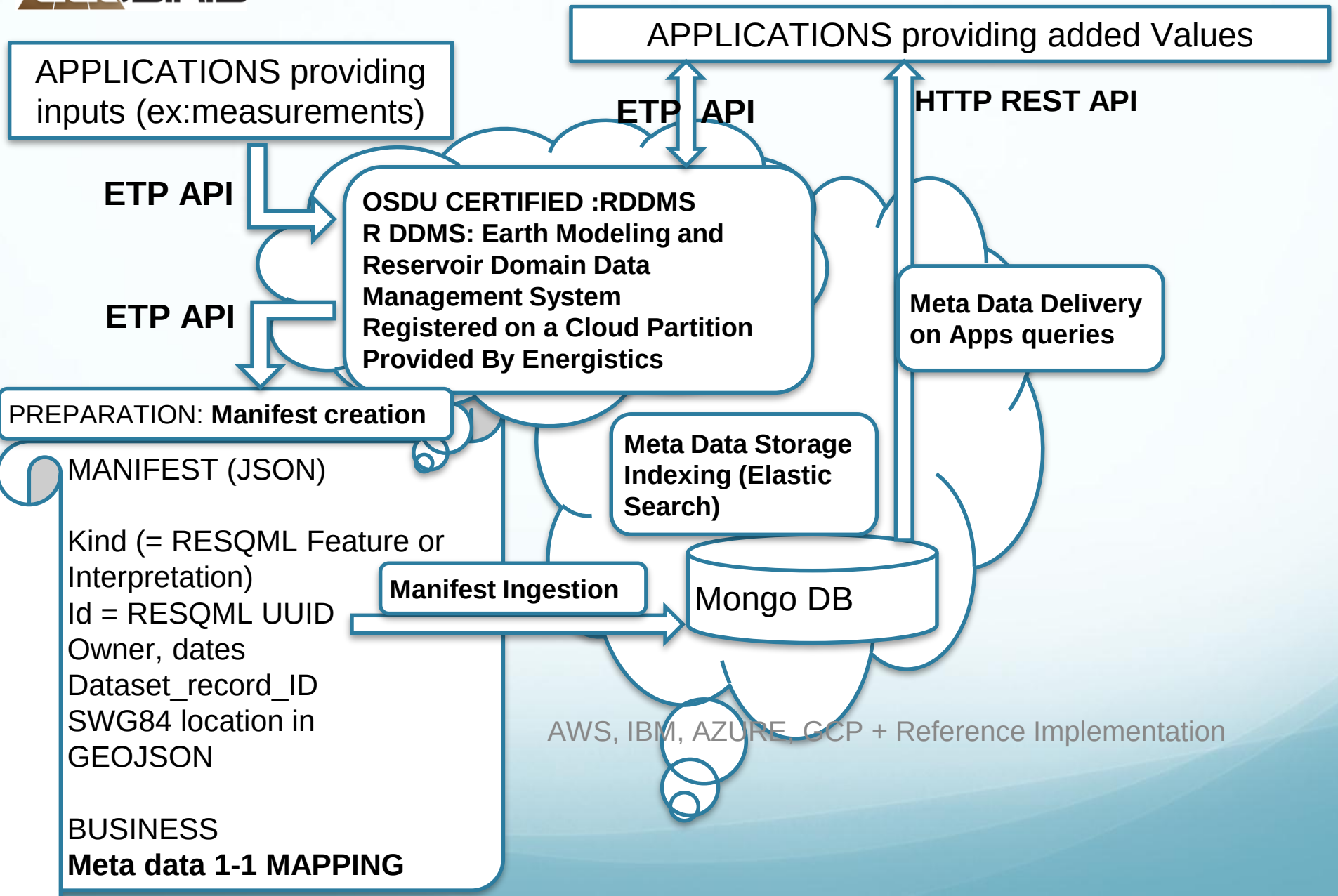
RESQML **Properties :**

- 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.

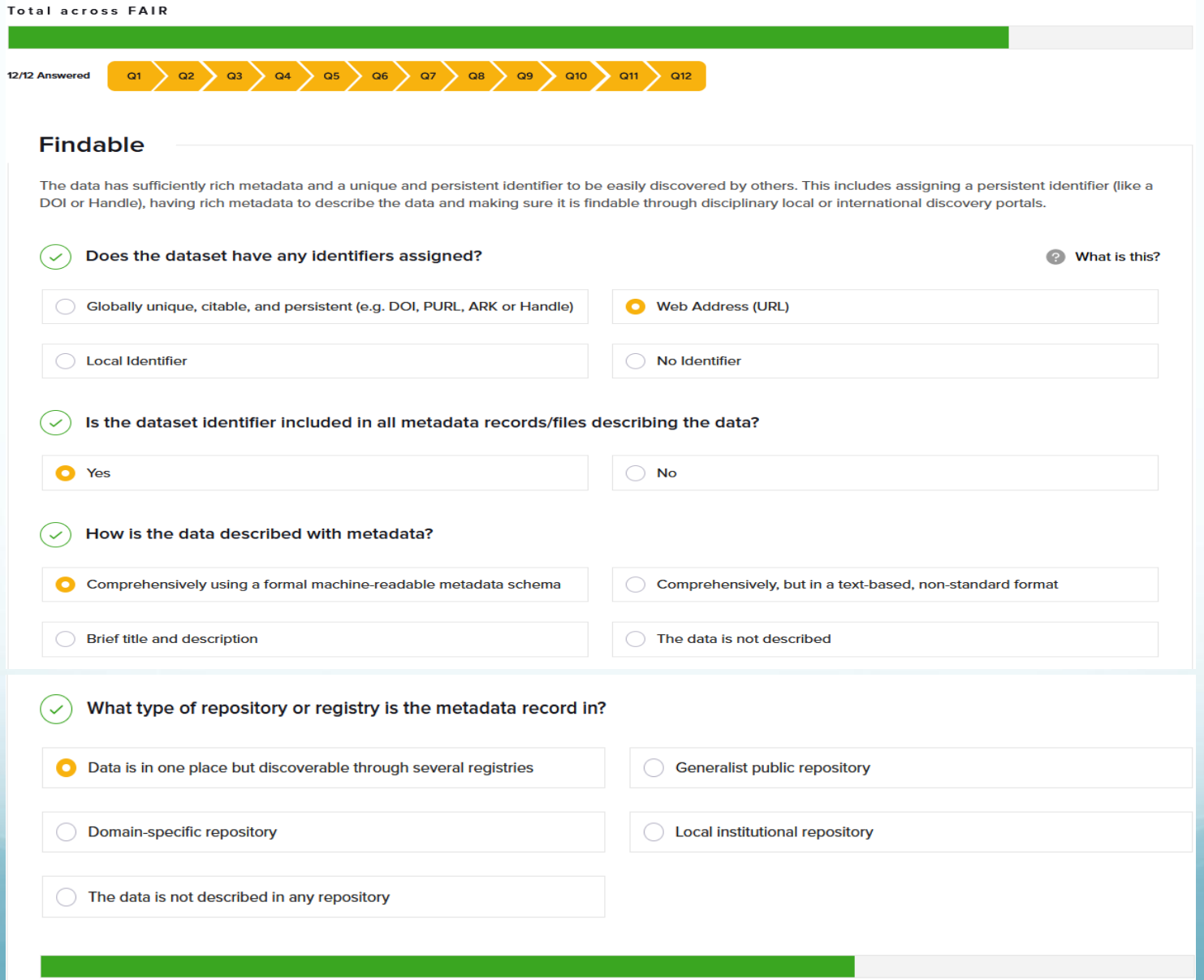
How OSDU and RESQML are linked (files)



How OSDU Platform and ETP Server are linked



OSDU/RESQML F.A.I.R Analysis



OSDU/RESQML F.A.I.R Analysis

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

OSDU/RESQML F.A.I.R Analysis

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 »

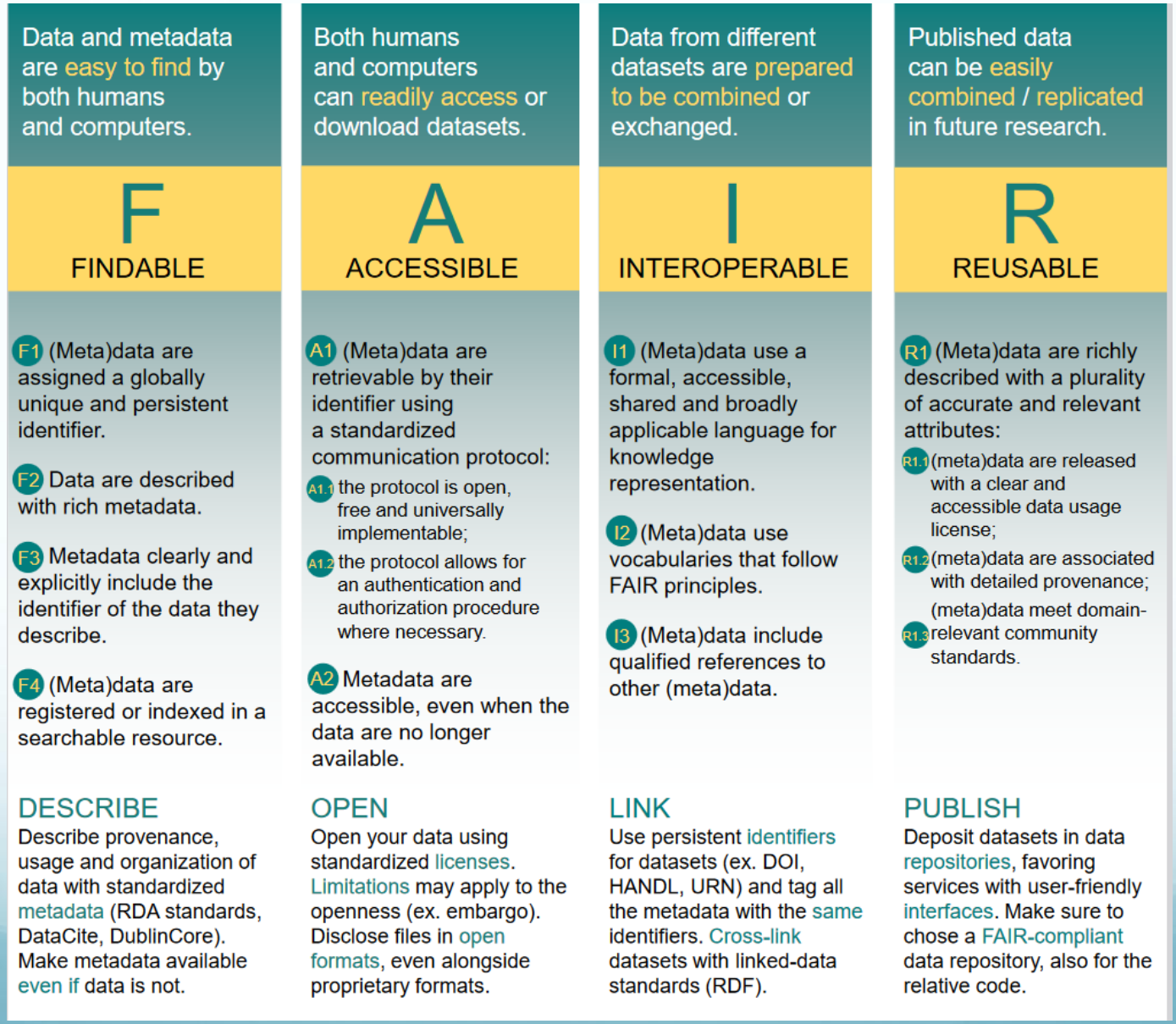
Merci de votre attention !

annexe

Les 14 principes, en résumé:

• identifiant pérenne • métadonnées enrichies • données cherchables et trouvables en ligne	• données récupérables en ligne par des protocoles standardisés • restrictions si nécessaire	• formats et standards communs • vocabulaires contrôlés	• données documentées (pex. avec un fichier README) • licence et information de provenance claires
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.	Dans un repository qui stocke les données de manière sécurisée	Pour être compatibles et combinables avec d'autres jeux de données	Afin que les données soient interprétées correctement

FAIR DATA PRINCIPLES [1]



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 [3] – 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 [4] – 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 [5]	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		