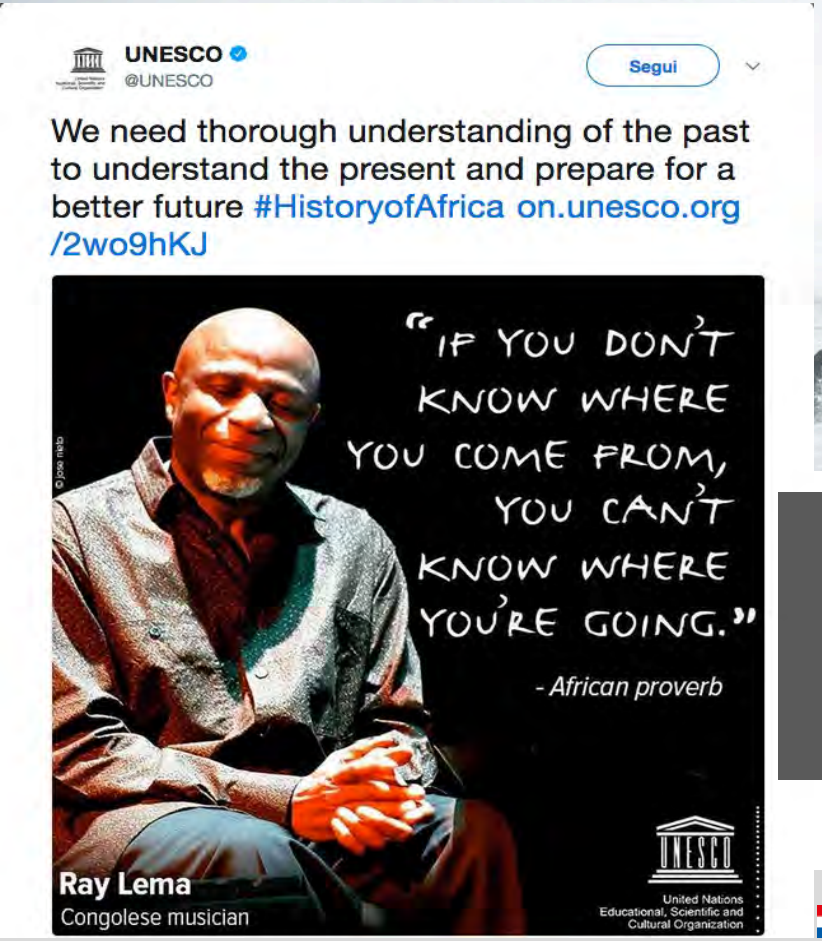
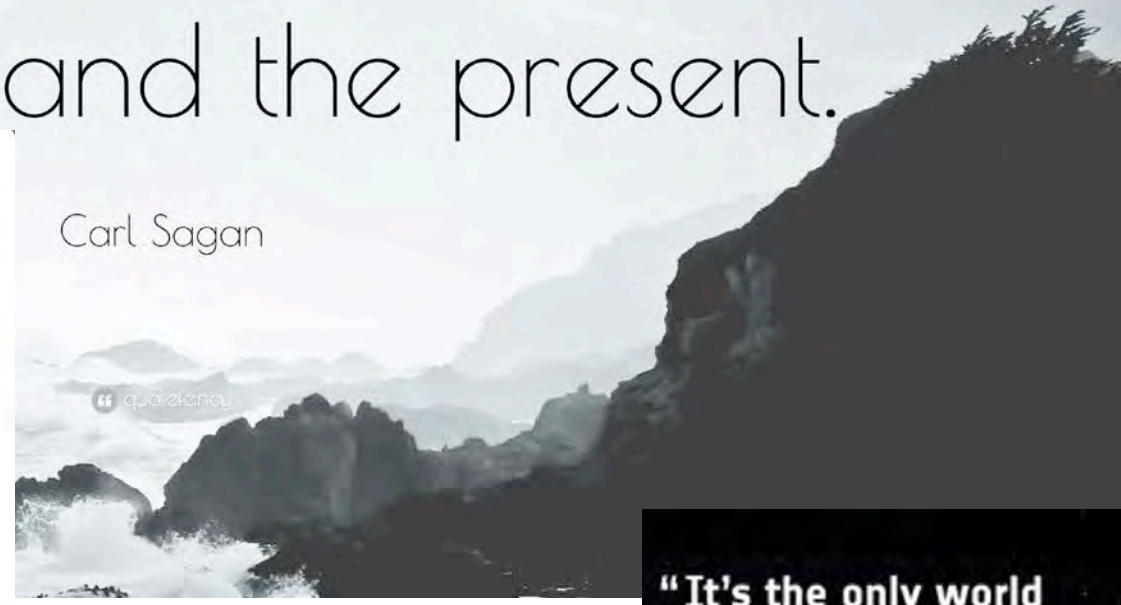
ESA SPACE DATA AND INFORMATION LONG TERM PRESERVATION, DISCOVERY AND ACCESS

BiDS 2019 – 19 February

Long Term Data Preservation TEAM (ESA)

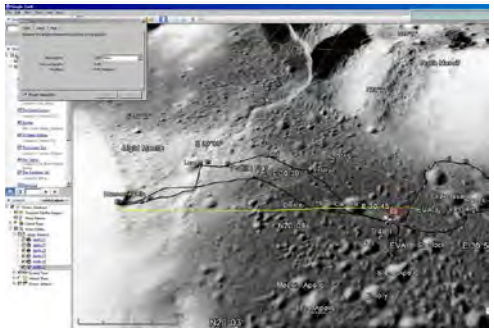
You have to know the past to understand the present.





Preserve and know the past to understand the present, and shape and enable the future

THE HERITAGE SPACE DATA AND INFORMATION PROGRAMME - MOTTO



PRESERVE AND KNOW THE PAST

UNDERSTAND THE PRESENT

SHAPE AND ENABLE THE FUTURE

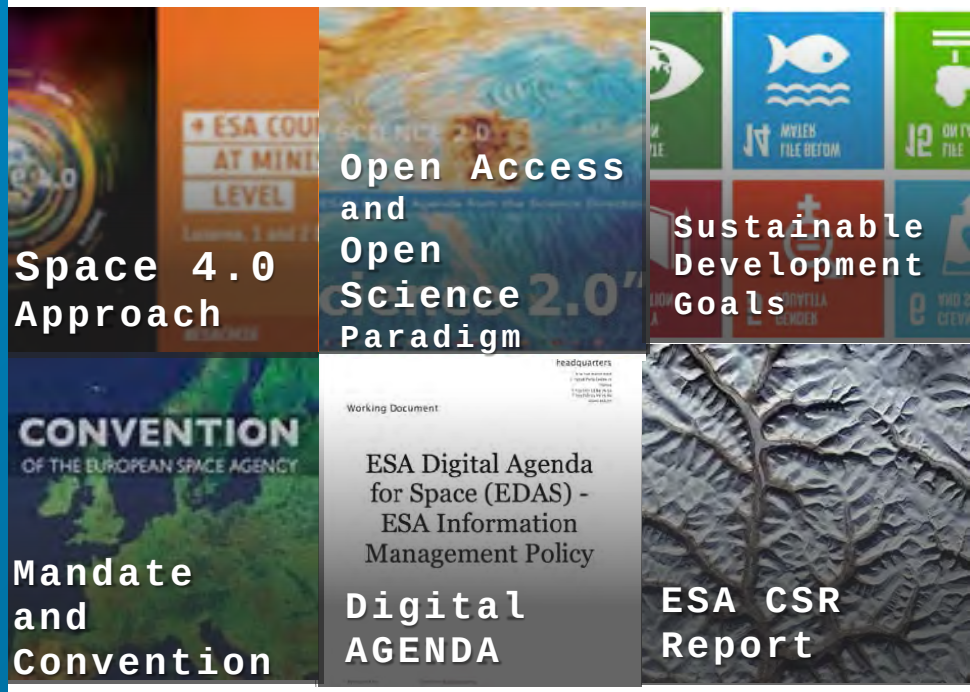
ESA VISION – REFERENCE FRAMEWORK

The European Space Agency, has the mandate to assure the long term preservation, sharing and exploitation of space data and its associated Information.

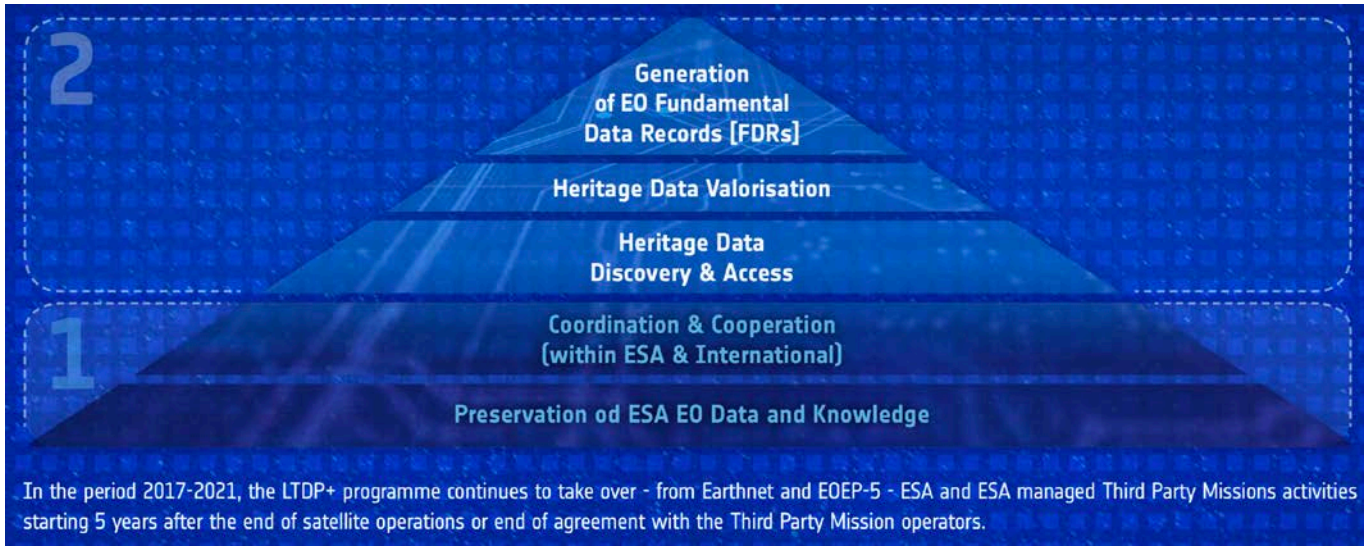
ESA's aim is to turn space exploration and space-related activities into an overall societal project involving a wide variety of stakeholders. To this end, it brings together and coordinates as many countries as possible under the banner of space missions. It is a basic principle that ESA deals with its stakeholders openly and with real transparency, an approach that has contributed to its long-term success.

Working towards better corporate responsibility and sustainability practices and ethical behavior requires that ESA focus on various aspects externally but also internally with the workforce. Each ESA office has a role to play in leading the Agency to increase its commitment to environmentally and socially responsible practices.

Space 4.0 in Europe will strongly depend on the appropriate management of the huge value in space data & associated knowledge held and shared by ESA and its partners across MSs, with industry and with the science communities.
ESA shall "Hold ALL of its information and data digitally and online accessible"



HERITAGE SPACE DATA AND INFORMATION PROGRAMME (LTDP+) - Activities



BIG DATA PARADIGM SHIFT

In the context of Big Data, it is vital to manage the vast amount and variety of the information around the data, as this information facilitates our capacity to extract information and meaning from the data.

Strategy and Pillars

Preservation

Discovery

Access

Exploitation

Valorization

Data Manager

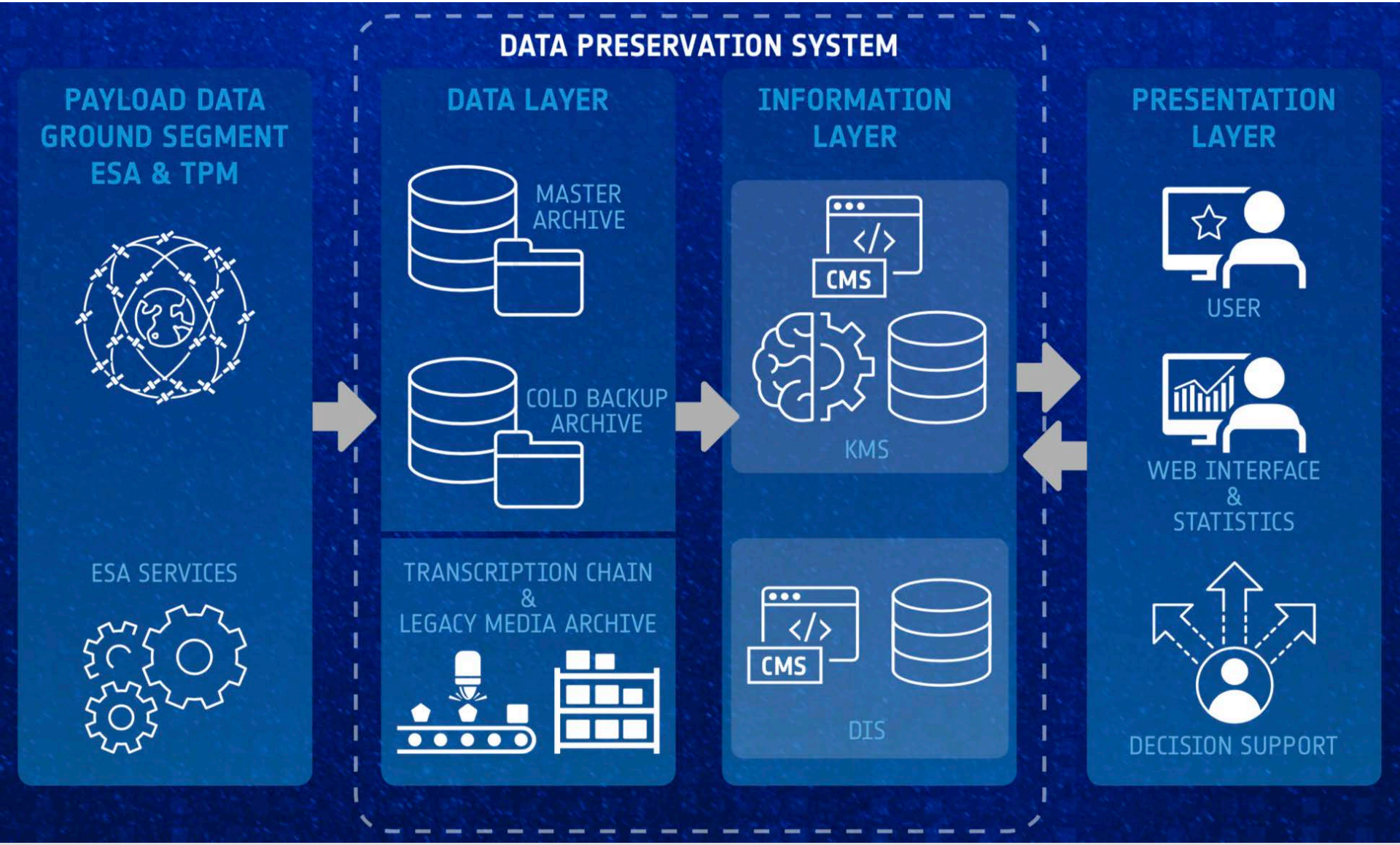
Data Curator

Data Scientist

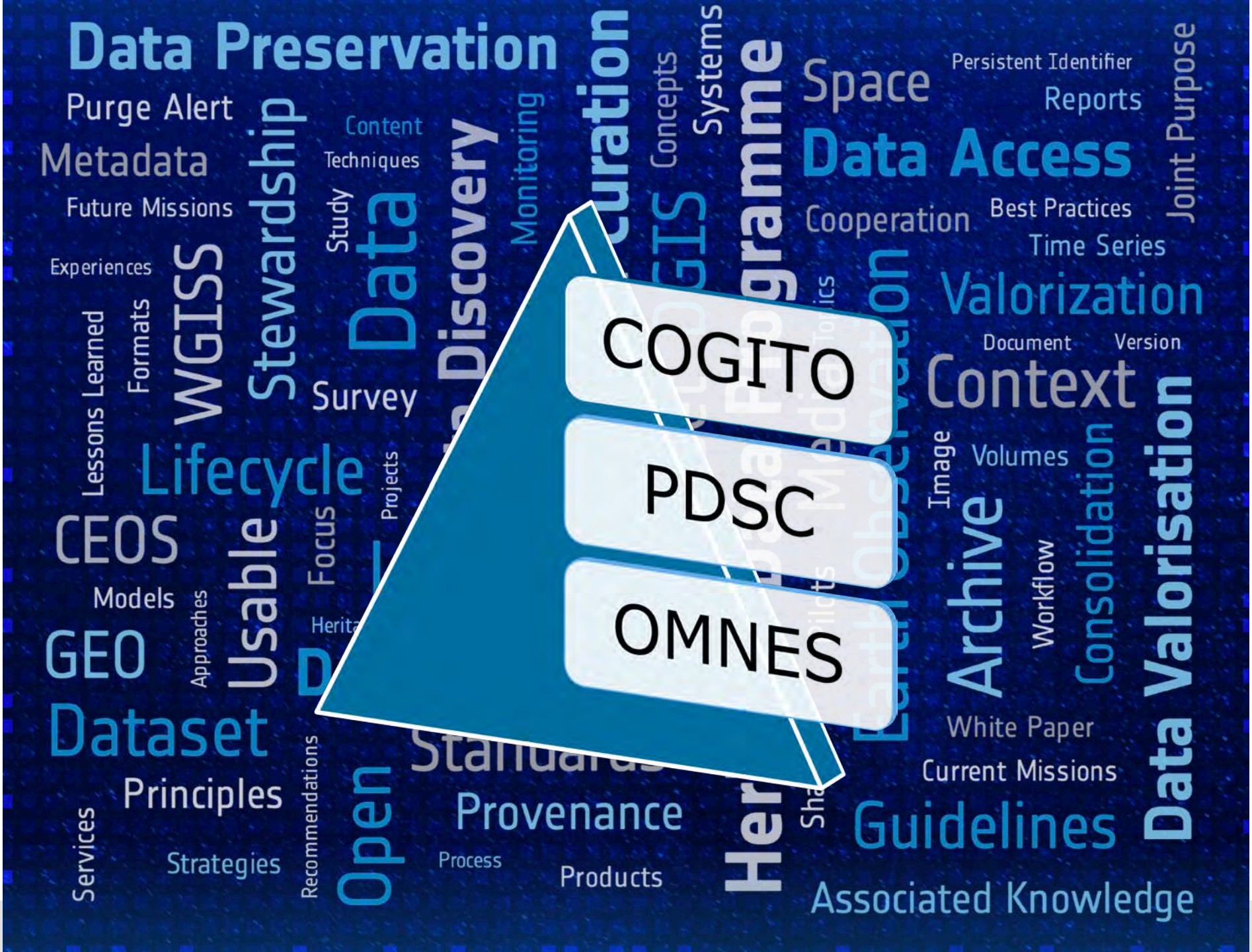
STANDARDS AND BEST PRACTICES

- Define Policy, Guidelines, Best Practices and Standards applicable to the Long Term Data Preservation, Discovery and Access Strategy Goals
  - CCSDS Data Archive and Interoperability WG
  - ISO Standards
  - CEOS Best Practices for Long Term Data Preservation
  - OGC Interoperability Standards
  - Data Management and Stewardship Maturity Matrix
  - Research Data Life Management
  - Data Management Plan

EO SPACE DATA PRESERVATION SYSTEM



DATA ASSOCIATED INFORMATION MANAGEMENT SYSTEM - CONCEPT


COGITO
PDSC
OMNES
Data Preservation
Purge Alert
Metadata
Future Missions
Experiences
Lessons Learned
Formats
WGISS
Stewardship
Content
Techniques
Study
Data
Survey
Discovery
Monitoring
Curation
Concepts
Systems
GIS
Programme
Space
Persistent Identifier
Reports
Data Access
Cooperation
Best Practices
Time Series
Joint Purpose
Valorization
Document
Version
Context
Image
Volumes
Archive
Consolidation
Workflow
Data Valorisation
Lifecycle
Projects
Focus
CEOS
Models
Usable
Approaches
GEO
Dataset
Principles
Services
Strategies
Recommendations
Open
Standards
Provenance
Process
Products
White Paper
Current Missions
Guidelines
Associated Knowledge

OMNES

Knowledge for Tomorrow Mankind

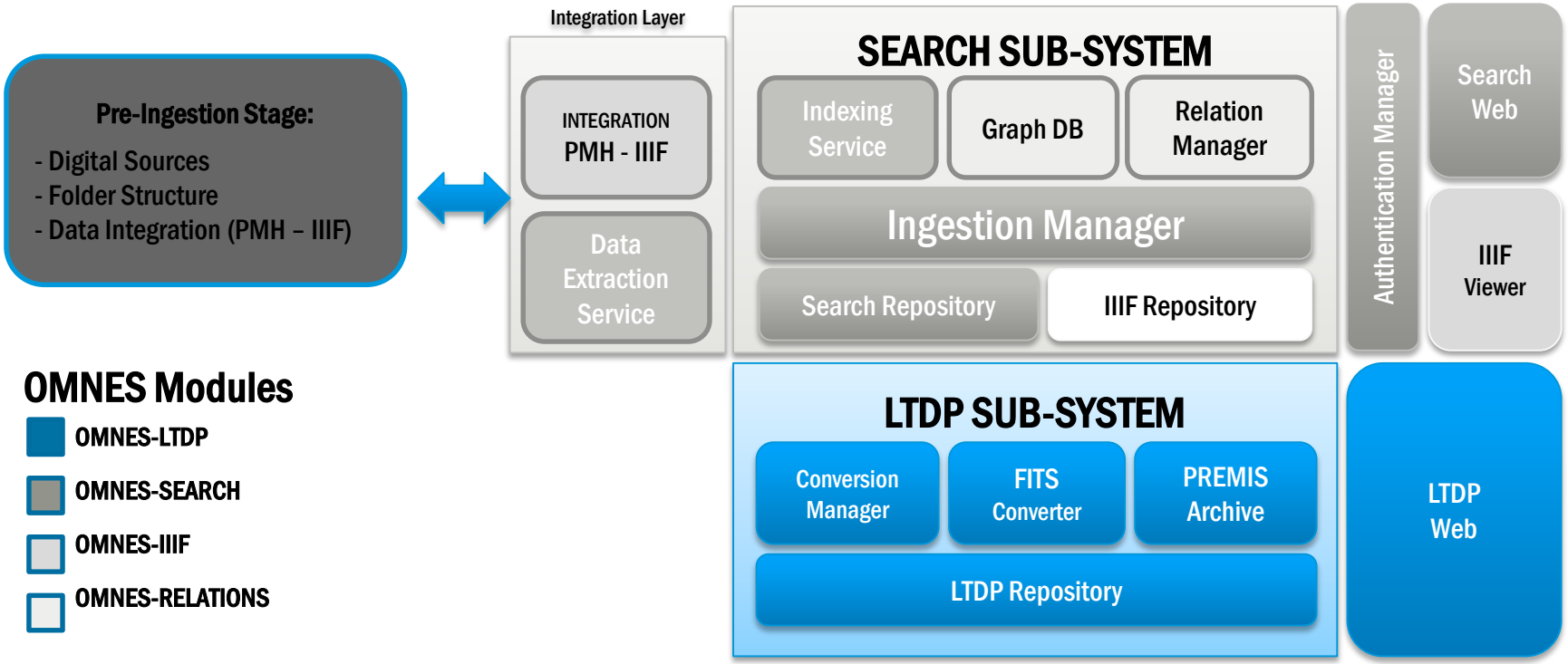
Innovative Heritage

Management Solution

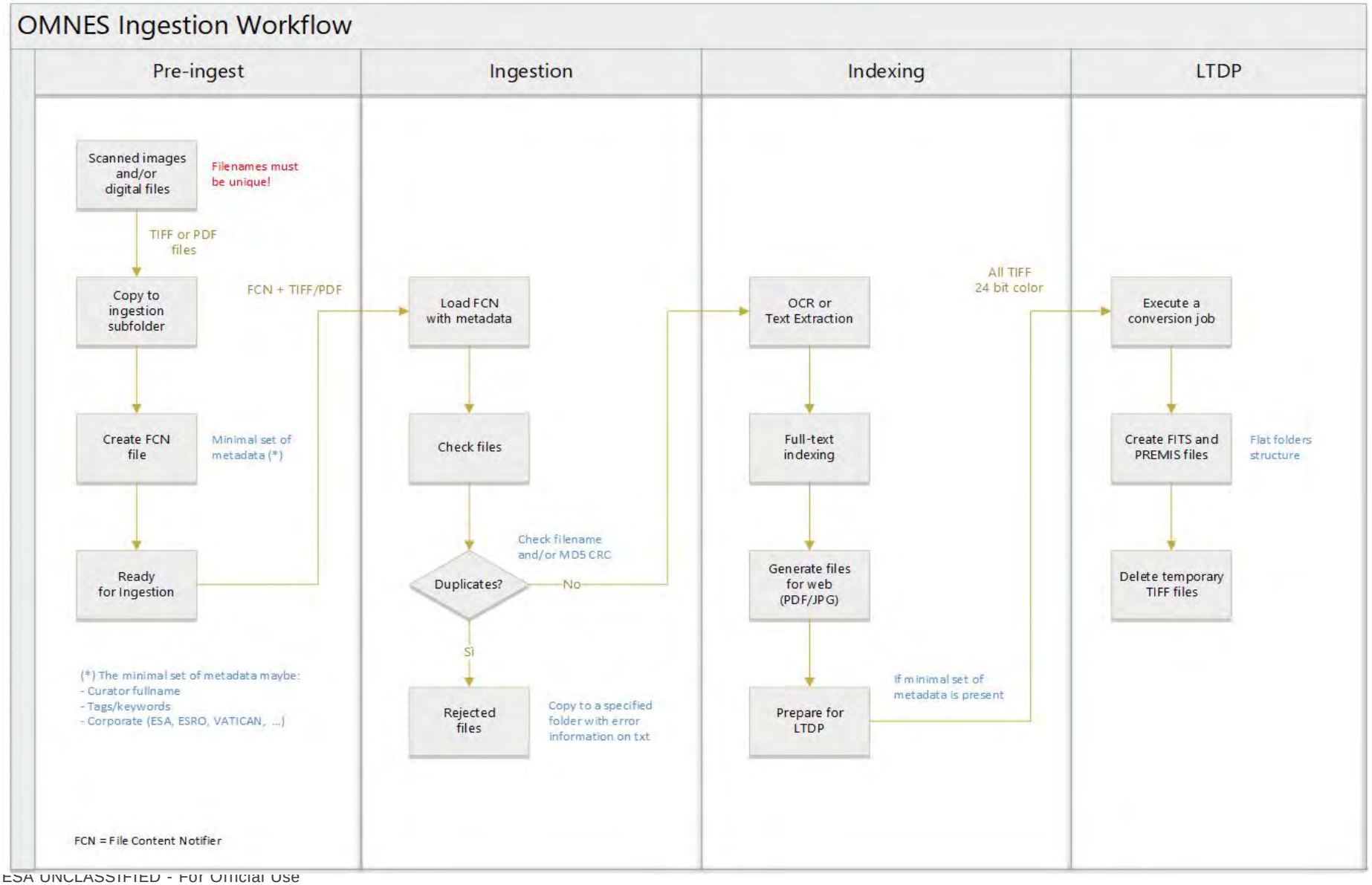
DB SERET srl

OMNES ARCHITECTURE

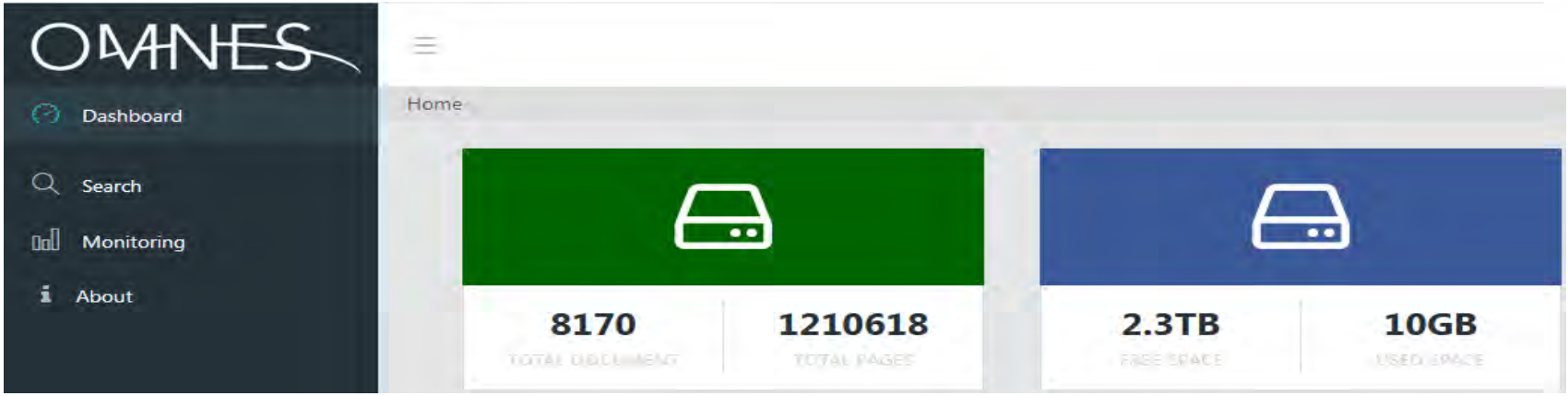
The objective of the OMNES platform is to ingest digital information, such as documentation or images, and to preserve it in a digital repository, with an appropriate long term archive format.



OMNES INGESTION WORKFLOW



OMNES- MENU & DASHBOARD



The new ISE UI Menu include 3 Macro Area

- Search Area
- Archive Area
- Administration Area

The new ISE UI MENU

"Upload Resources" feature allow the users to upload the Resources and start the OMNES ingestion flow using a web feature.

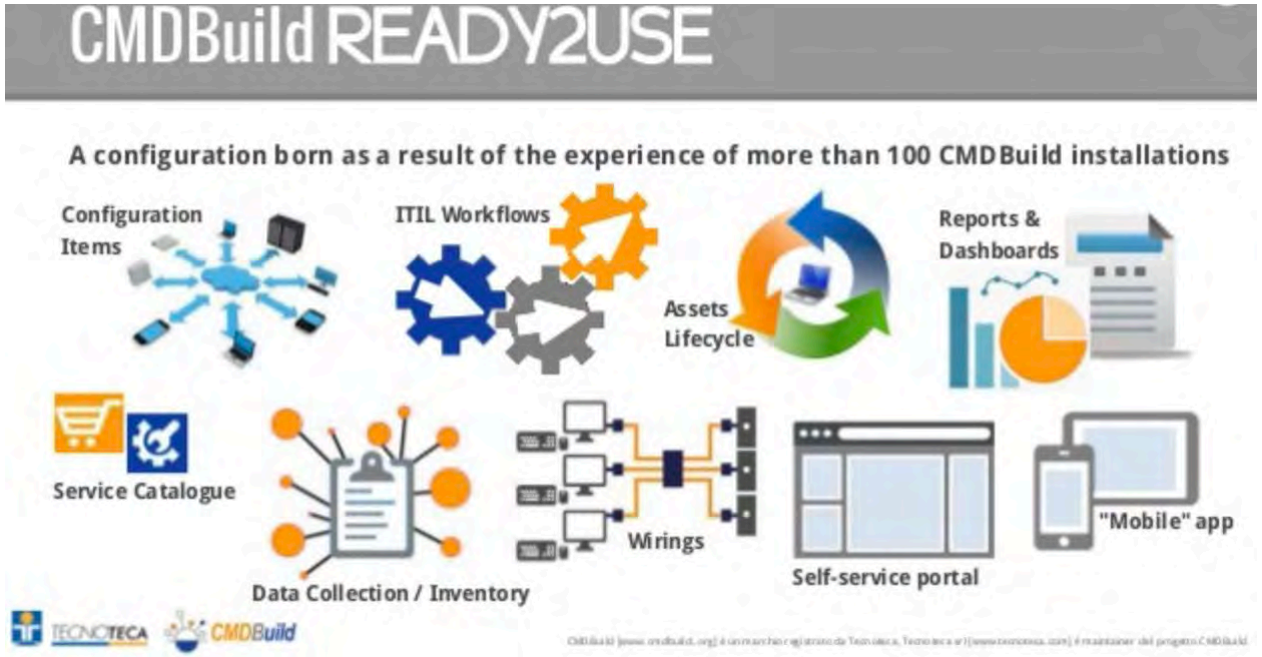
This feature not need manual FCN file generation because manage it automatically from data entry of Dublin Core metadata

It manage both, single and multiple resources

It manage the Automatic or Manual DC_Identifier Metadata

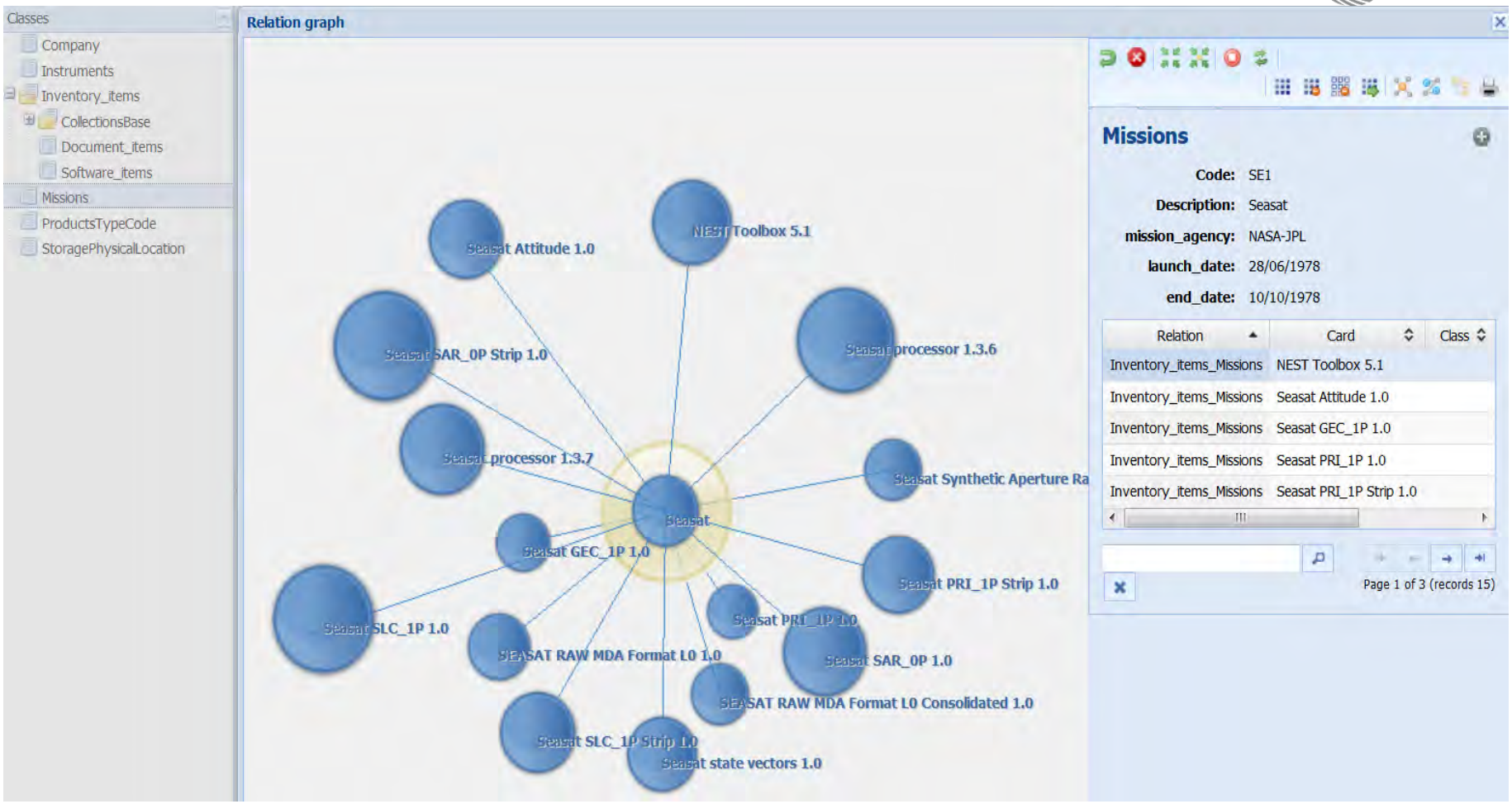
KMS COMPONENT: PDSC

Preserved DataSet Content is the brain of the KMS



The PDSC Management System is composed of an open source web enterprise environment, CMDBuild, and an Associated Knowledge repository, Alfresco. CMDBuild allows the management of IT assets' configuration so that the organisation may be able to keep under control the situation of data and associated digital resources, knowing at all times the composition, dislocation, functional relations, rules for updating over time.

PDSC - NETWORK OF RESOURCES



Once the associated knowledge and tools are linked to the corresponding data records (including their physical location), the data set is considered consolidated from an information point of view.

PDSC - MISSIONS TAILORING FOR LTDP

MISSION

Mission:	JERS 1	**Orbit period [min]:**	96.00	**Agency:**	JAXA-NASDA
Launch Date:	1992-02-11	**Orbit altitude:**	570		
Mission end:	1998-10-12	**Orbit type:**	Sun Synchronous		
		Cycle [days]:	44		
		LST:	10:45		

Mission url: https://directory.eoportal.org/web/eoportal/satellite-missions/j/jers-1

Description	Resolution	Swath [Km]	Waveband	Geometry
JERS-1 L-band Synthetic Apature Radar	18m (range) x 18m (azimuth at 3 looks)		Microwave: L-Band 1275 MHzMW (~1.0 cm - ~100 cm) L-Band (2 - 1 GHz)	Side-looking
JERS-1 Optical Sensor	18.3m (range) x 24.2m (azimuth)		VIS: 0.42 - 0.50 μm, 0.52 - 0.60 μm, 0.61 - 0.69 μm, NIR: 0.76 - 0.89 μm VIS (~0.40 μm - ~0.75 μm) SWIR (~1.3 μm - ~3.0 μm)	Cross-track scanning
Sensors number:2				

DATA RECORDS: AUXILIARIES

Auxiliary	Product type	Instrument	Date start Date stop	Ingestion format	Size [Mb]	Auxiliary number
Jers Definitive Ephemeris Jaxa 1.0	AUX_SV_RES		1991-04-03 1998-10-12	no sip	333.047	2436
Jers Orbit Scenario File 1.0	AUX_OR_REF		1991-04-03 1998-10-01	no sip	0.000	1
Jers Restitude state vectors 1.0	AUX_SV_RES		1991-04-03 1998-10-01	no sip	0.336	98
Auxiliaries records: 3					**333.383**	**2535**

DATA RECORDS: COLLECTIONS

Description	Product format	Ingestion format	Date start Date stop	Products number Products size [GB]	Notes/Software
Jers Level0 OPS	WILMA	no sip	1992-10-28 1998-10-11	2089 3212.000	
Jers Level0 RAW	CEOS	sip	1992-07-13 1998-10-08	81522 7065.600	
					Jers processor 2.05 2.05
Jers Level0 SAR	WILMA	no sip	1992-07-13 1998-10-08	3001 11048.000	
Jers Level1 OPS	geoTIFF	sip	1992-10-28 1998-10-11	55068 1740.800	
Jers Level1 PRI	CEOS	sip	1992-07-13 1998-10-08	82156 3891.200	
					Jers processor 2.05 2.05
Jers Level1 SLC	CEOS	sip	1992-07-13 1998-10-08	81975 21504.000	
					Jers processor 2.05 2.05
Collections number: 6				**Products: 305811 Size: 48461.602**	

PDSC - MISSIONS TAILORING FOR LTDP

DATA COLLECTIONS TAILORING

Collection	Software	InAux
Jers Level0 OPS 1.0	0	0
Jers Level0 RAW 1.0	1	0
Jers Level0 SAR 1.0	0	0
Jers Level1 OPS 1.0	0	0
Jers Level1 PRI 1.0	1	0
Jers Level1 SLC 1.0	1	0

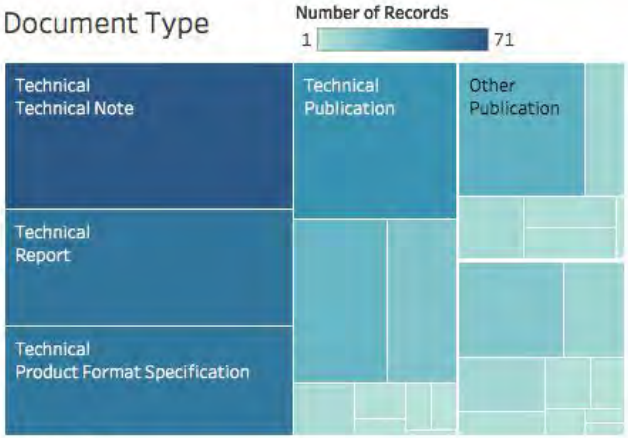
MISSION TAILORING

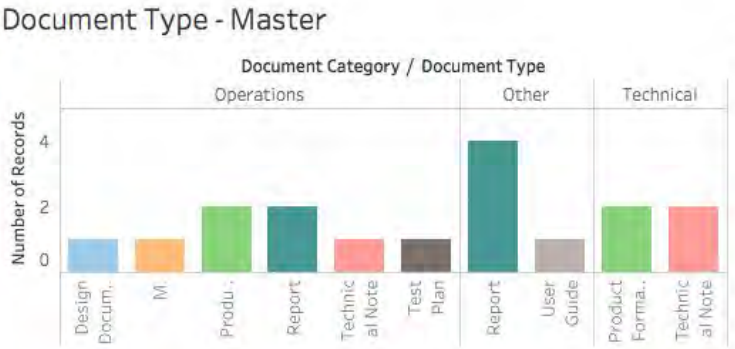
Id	Description	No. Items
MC 1.1	Scientific Scenario and User Communities	0
MC 1.2	Mission Requirement Document	0
MC 1.3	Mission Operation Plan	0
MD 1.1	Mission Requirements Specifications	0
MD 1.2	Space to Ground segment ICDs	0
MD 1.3A	Sensor / Instrument requirements	0

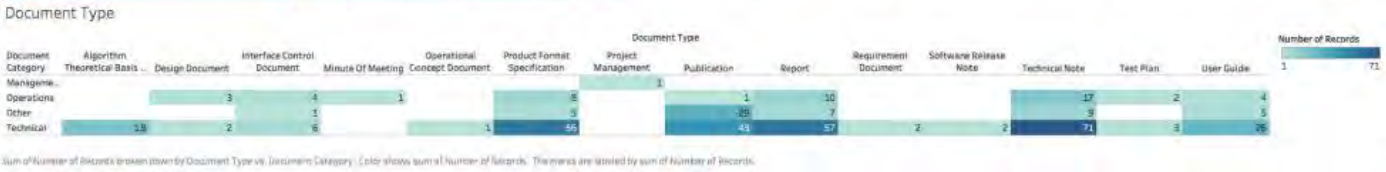
...

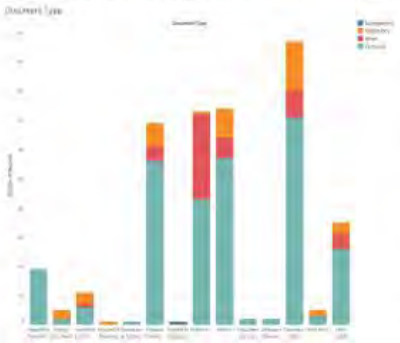
Id	Description	No. Items
MO 1.18	Product qualification and quality assurance monitoring reports	0
MO 1.19	Sensor/Instrument evolution and history records	0
MO 1.20	Referred publications and papers	0
MO 1.21	Tandem and/or combined campaigns; comparisons	0
MO 1.22	Cross- campaign; cross- comparisons and cross- calibration activities documentation and Data	0
MO 1.23	Data Access Policy	0
PM 1.1A	Data consolidation & reprocessing strategy; implementation plans; and consolidated/ reprocessed data. Processing.	0
PM 1.1B	Data consolidation & reprocessing strategy; implementation plans; and consolidated/ reprocessed data.Ancillary; Auxiliary	3
PM 1.1C	Data consolidation & reprocessing strategy; implementation plans; and consolidated/ reprocessed data. PA/QA	0
PM 1.2	Data consolidation & reprocessing strategy; implementation plans; and consolidated/ reprocessed data.L0; L1; L2	4
PM 1.3	Data consolidation & reprocessing strategy; implementation plans and consolidated/ reprocessed data.Data/Image processing	0
PM 1.4	Data consolidation & reprocessing strategy; implementation plans and consolidated/ reprocessed data.Browse Metadata	0

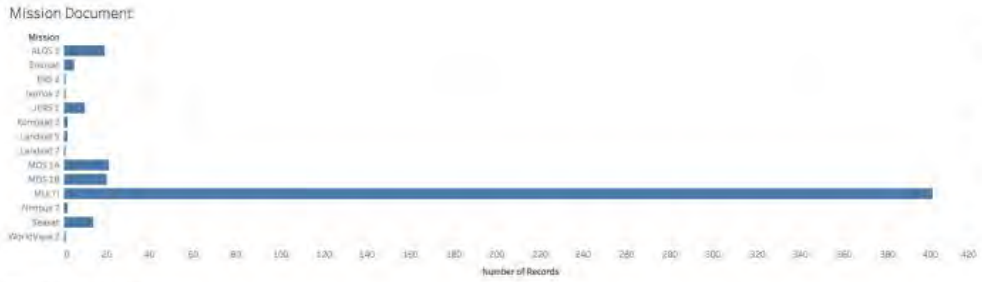
PDSC – INFORMATION VISUALIZATION TOOL







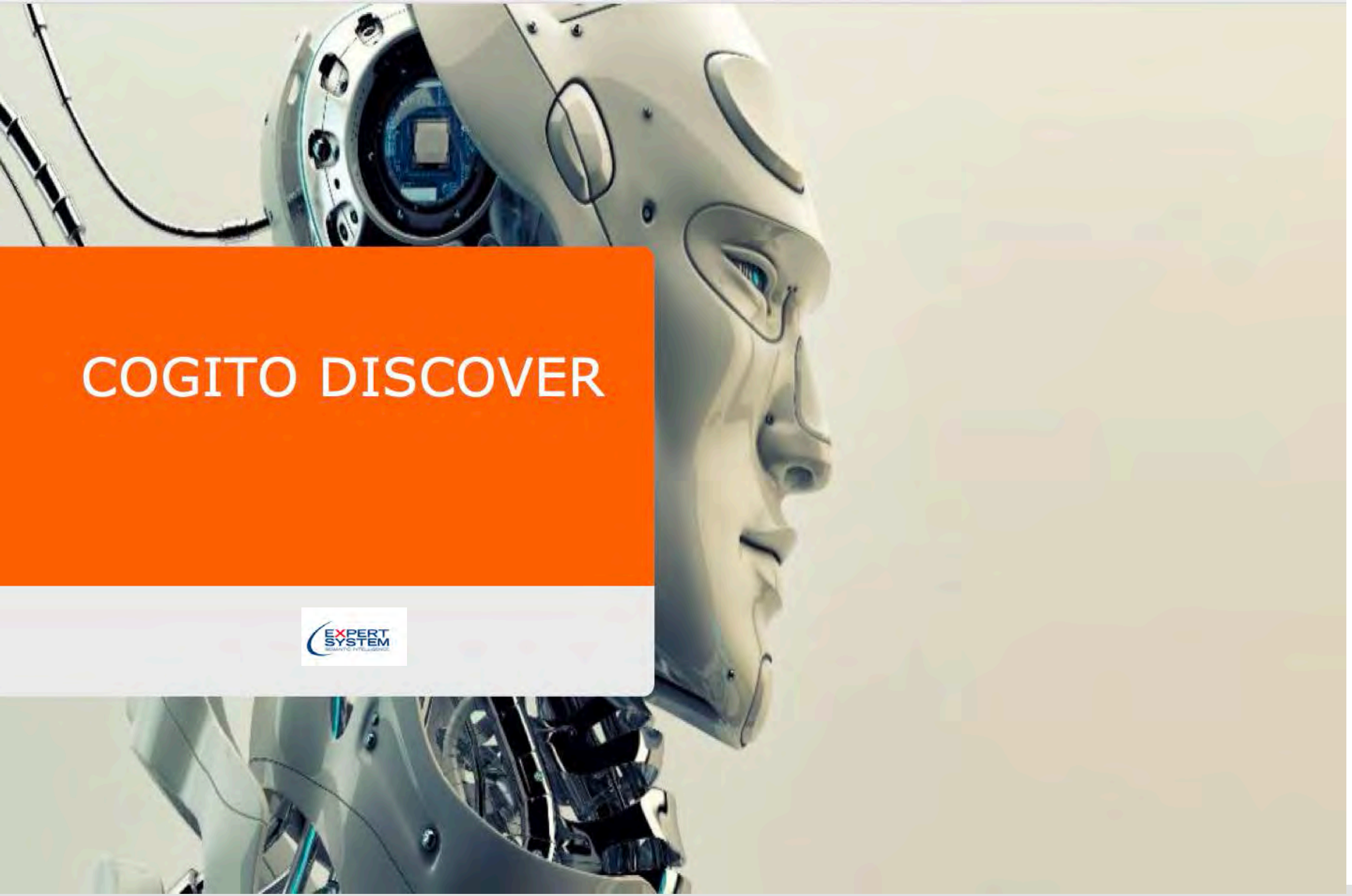




Mission Document

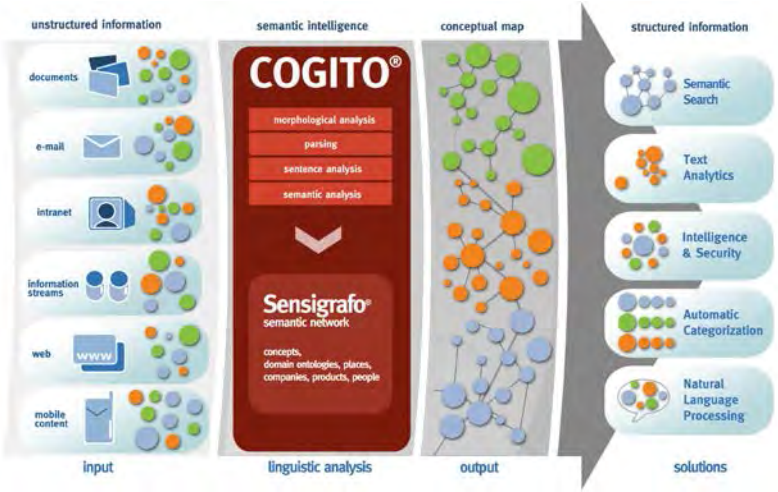
Mission	Instrument	Number of Records (1 – 400)
ALOS 1	Null	9
	Advanced Visible and Nea..	1
	Panchromatic Remote-se..	5
	Phased Array type L-band ..	4
Envisat	Advanced Synthetic Apert..	3
	Advanced Synthetic Apert..	1
	Advanced Synthetic Apert..	1
ERS 2	Null	1
Ikonos 2	Ikonos 2 Optical Sensor A..	1
JERS 1	Null	3
	JERS L-band Synthetic Ap..	5
	JERS Optical Sensor	2
Kompsat 2	Multi Spectral Camera	2
Landsat 5	Thematic Mapper	2
Landsat 7	Enhanced Thematic Mapp..	1
MOS 1A	Null	5
	Multispectral Electronic S..	8
	Visible and Thermal Infrar..	8
MOS 1B	Null	4
	Multispectral Electronic S..	8
	Visible and Thermal Infrar..	8
MULTI	Null	400
	Thematic Mapper	1
Nimbus 7	Coastal Zone Color Scanner	2
Seasat	Null	2
	Seasat Synthetic Apertur..	12
WorldView 2	Multi Spectral Camera	1

KMS COMPONENT: COGITO DISCOVER



COGITO DISCOVER

HOW COGITO WORKS

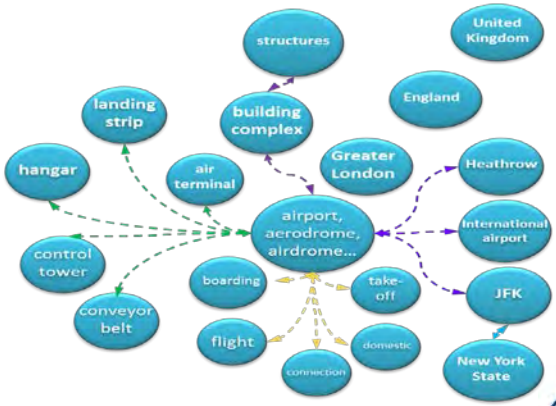


COGITO solution reads a text like a human does, deriving meaning from the text as a whole using powerful semantics analysis; cognitive and contextual reasoning and machine learning; and not just applying text analytics and statistical measures.

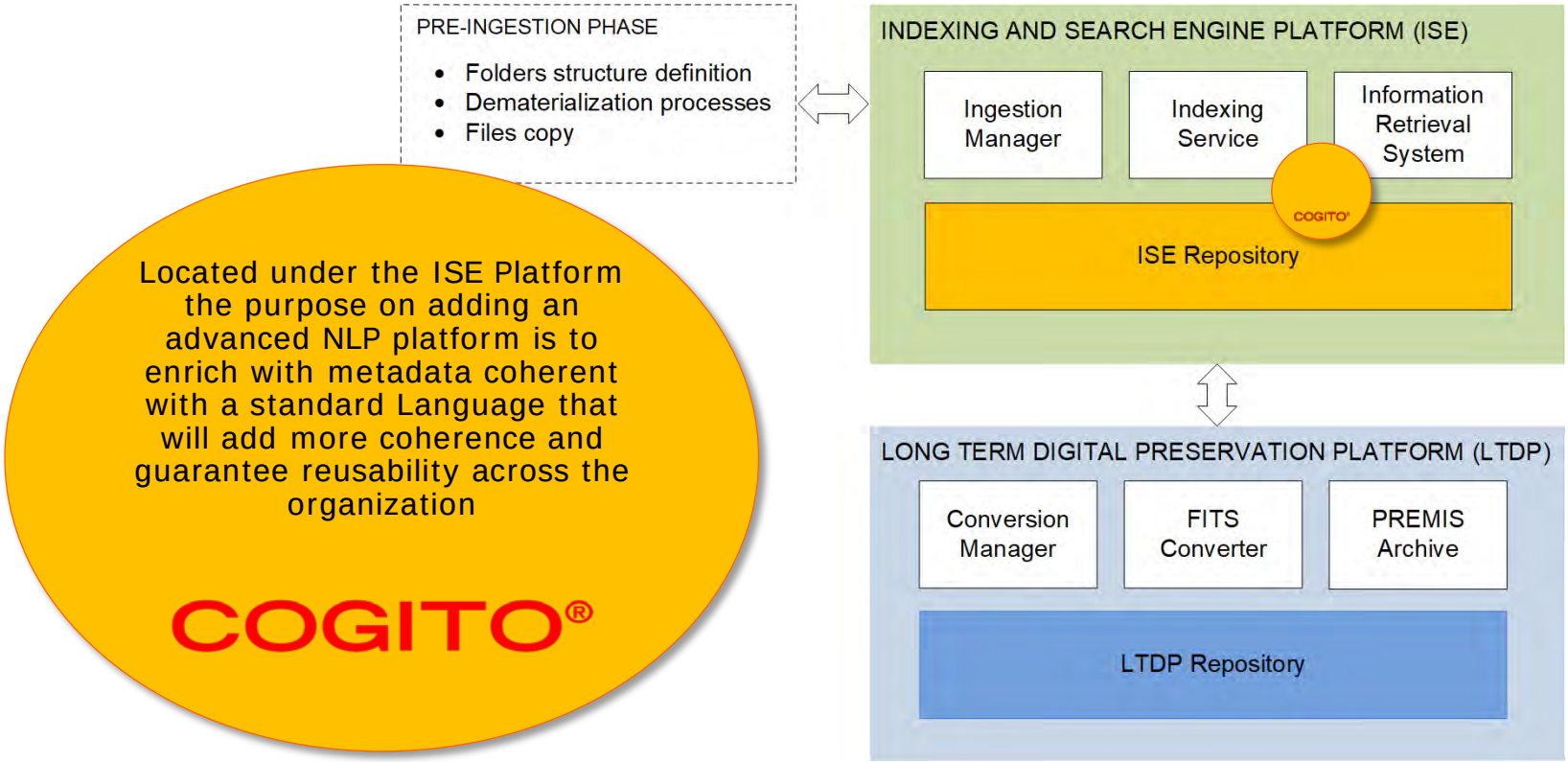
Natural language text is read and analyzed

1. **Sentence splitting / parsing**
   - Divide text into words
2. **Morphological analysis**
   - Understand word forms (e.g. Verb conjugations)
3. **Sentence / logical / grammatical analysis**
   - Understand how words relate to other words, and what their function is in the sentence
4. **Semantics analysis / disambiguation**
   - Understanding sentences and texts as a whole, taking into account words synonyms, context, plausibility

And then translated into a cognitive or conceptual map

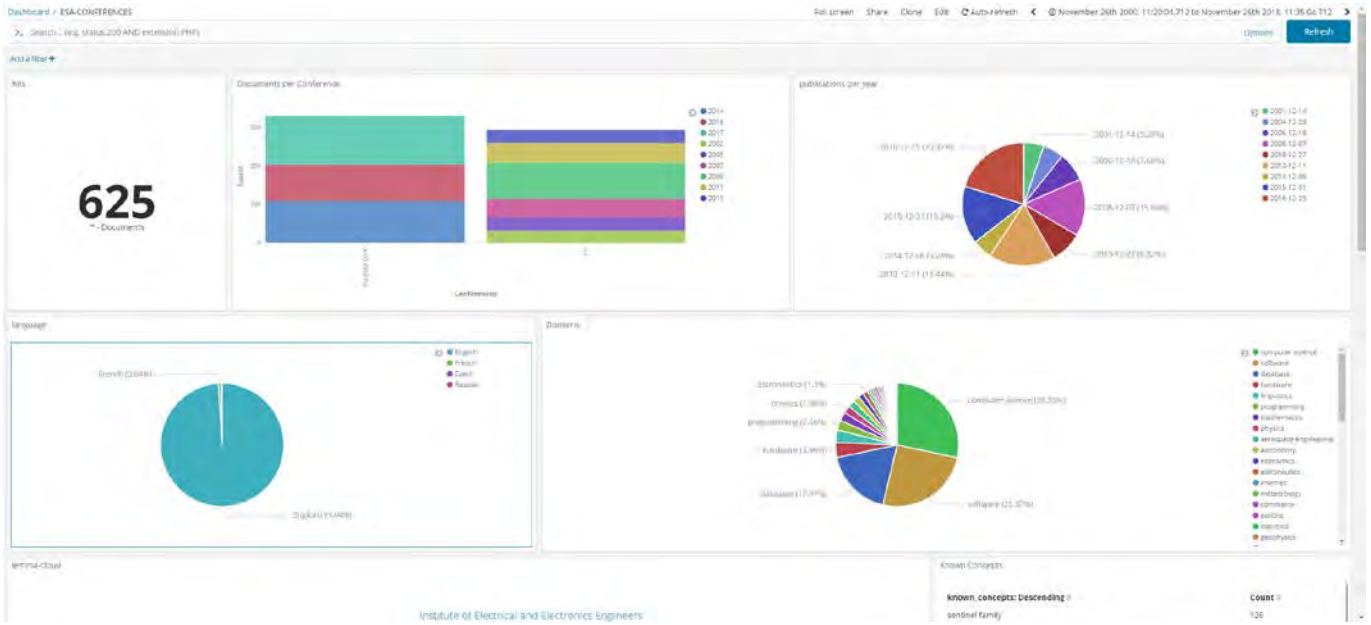


COGITO HIGH -LEVEL SYSTEM ARCHITECTURE

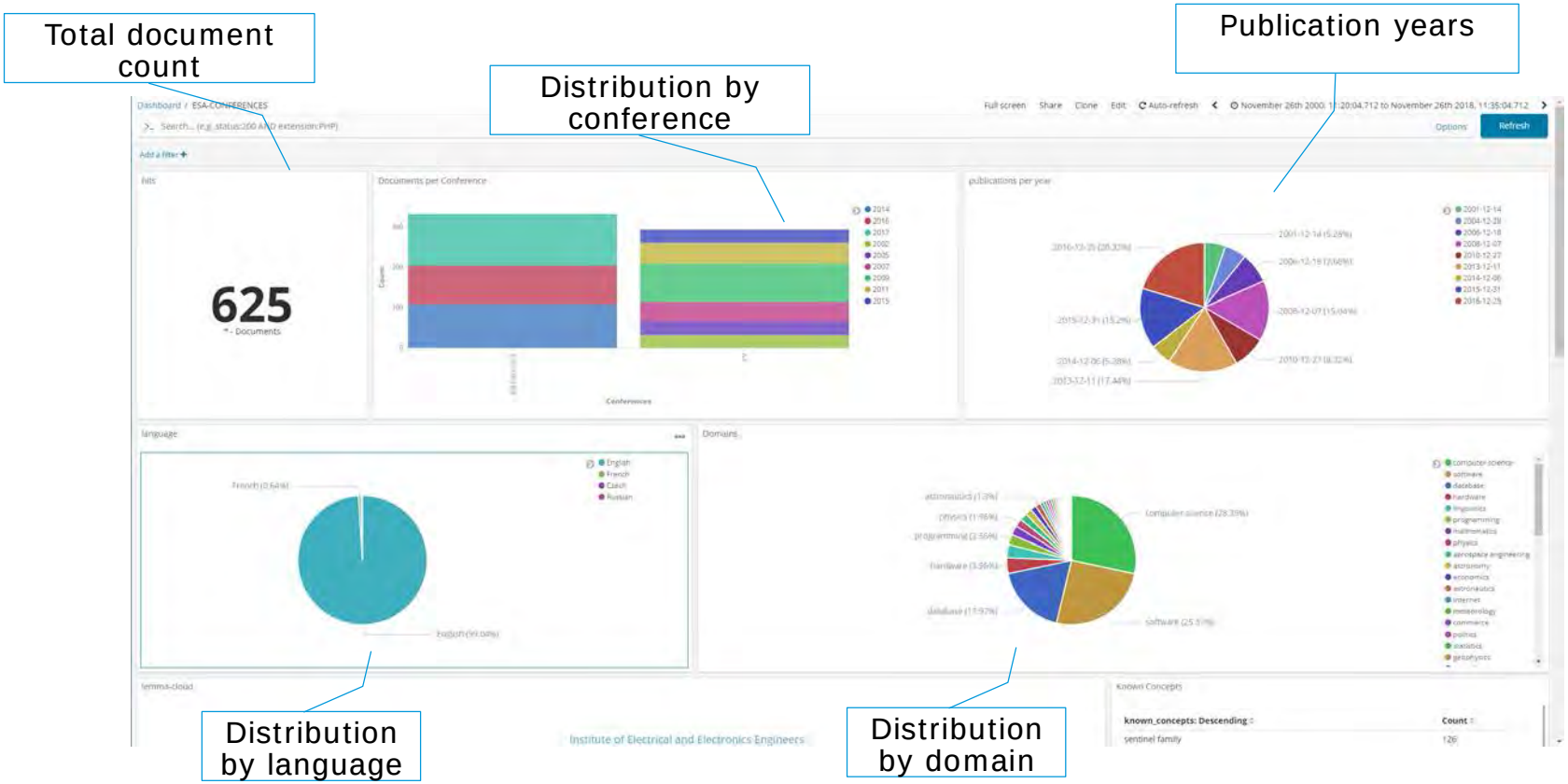


ESA USE CASES

The purpose of this dashboard is to illustrate the terminology update implemented so far during the project. Around 1.400 new terms and concepts have been distilled from the documentation provided by ESA, including the ESA vocabulary and glossary of terms, the ESA taxonomy, and a corpus of 625 scientific papers from the proceedings of ESA conferences, including PV and BiDS conferences.



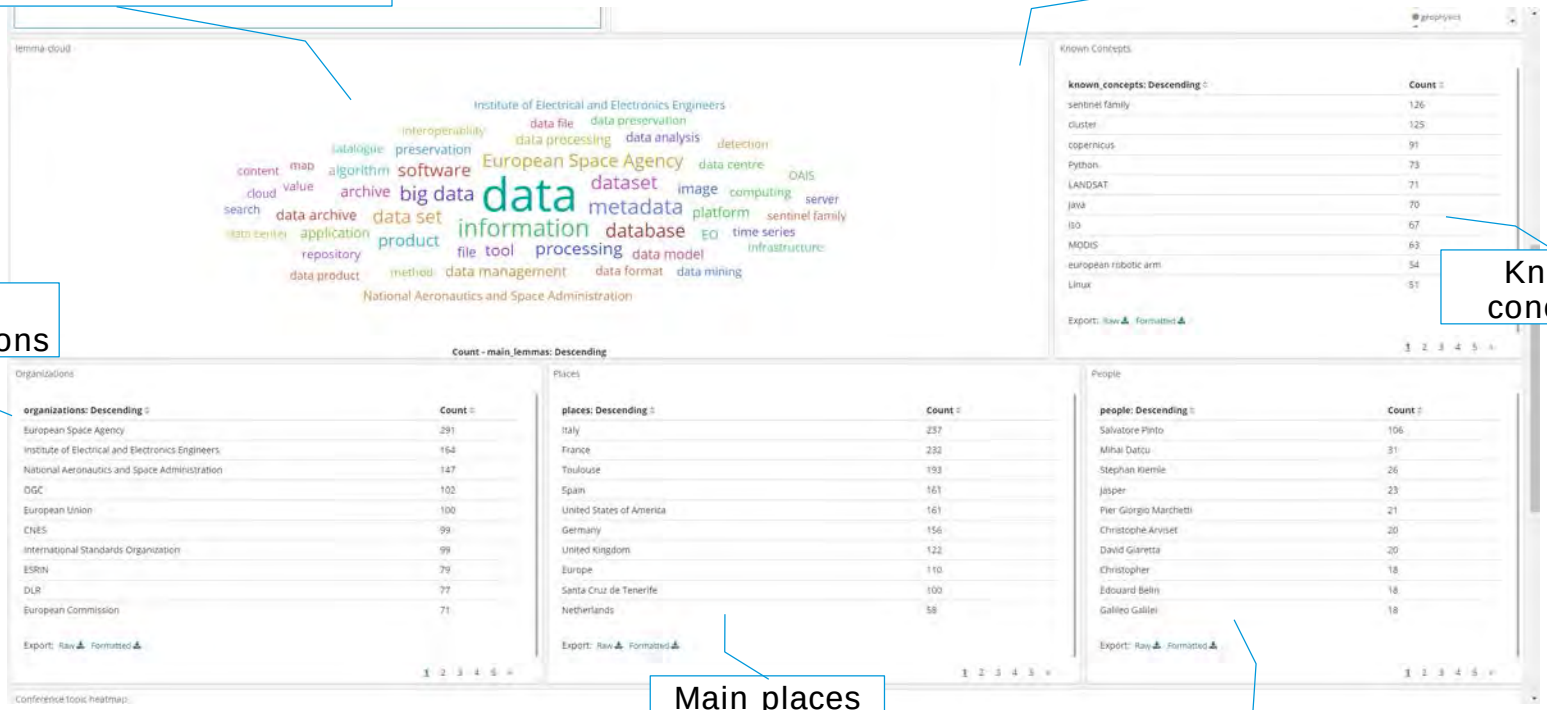
KNOWLEDGE NUMBERS



MAIN LEMMAS, CONCEPT AND TOP ENTITIES

Main lemmas per frequency

Sensigrafo domain customization optimized performance sensibly



Main organizations

Known concepts

Main places

Main persons

Lemma: canonical representations of words

RELEVANCE OF TOPICS PER CONFERENCE AND YEAR

OGC	102
European Union	100
CNES	99
International Standards Organization	99
ESRIN	79
DLR	77
European Commission	71

Export: Raw Formatted

1 2 3 4 5 »

Spain	161
United States of America	161
Germany	156
United Kingdom	122
Europe	110
Santa Cruz de Tenerife	100
Netherlands	58

Export: Raw Formatted

1 2 3 4 5 »

Jasper	23
Pier Giorgio Marchetti	21
Christophe Arviset	20
David Giaretta	20
Christopher	18
Edouard Belin	18
Galileo Galilei	18

Export: Raw Formatted

1 2 3 4 5 »


Thank
You
Maake
Asante
Dhanyavadagalu
Shukria
Kiitos
köszönöm
Vinaka
спасибо
Dankscheen
Kam Sah Hammida
Manana
Dankon
Matondo
Terima Kasih
Taiku
Tack
Diolch i Chi
Chokrane
Grazie
Arigato
Mochchakkeram
Gracias
Tingki
Paldies
Gratias Tibi
Obrigado
Kia Ora
Khun Khap
Kop
Djiere Dieuf
Eskerrik Asko
Najis Tuke
Go Raibh Maith Agat
Salamat
Merci
Di Ou Mèsi
Hvala
Welalin
Danke
ХВала
Nirringrazzjak
Grazas
धन्यवाद
cảm ơn bạn
Biyan
Mauruuru
Juspaxar
Dziekuje
Ngiyabonga
Blagodaram
Dank Je
Bedankt
Dakujem
Ua Tsaug Rau Koj
Děkuji
Suksama
Rahmat
Nuwun
Matur
Misaotra
谢
谢
நன்றி
감사합니다
ありがとう
ขอขอบคุณคุณ