The Design Engineering Assistant

Applying Ontology Learning to the generation of a Space Mission Ontology

Audrey Berquand, University of Strathclyde (UK)
Annalisa Riccardi , University of Strathclyde (UK)
25th June 2019,
Space Systems Ontology Brainstorming Workshop

Scotland from the ISS- Image Credit: ESA/NASA

Background

Motivations

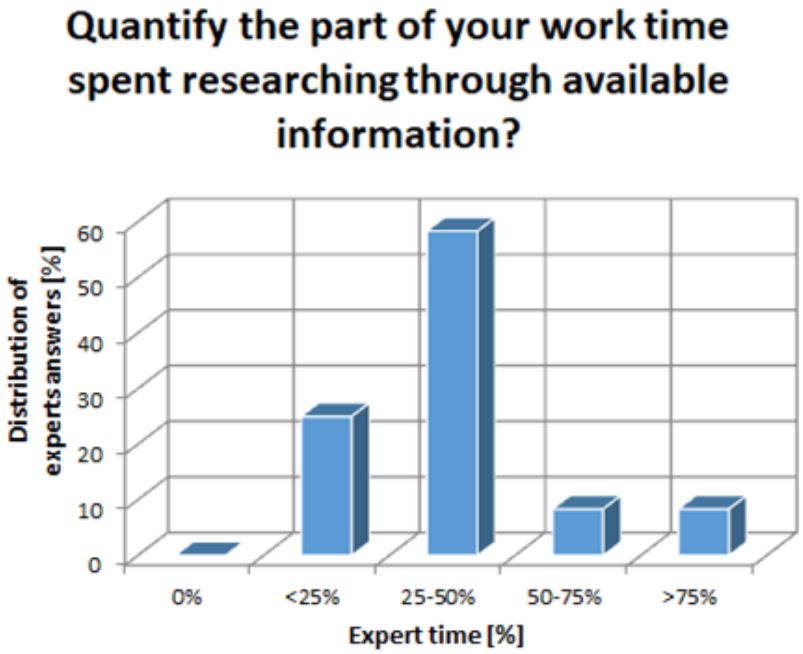
1. Large amount of accumulated data on spacecraft design, part of it unstructured

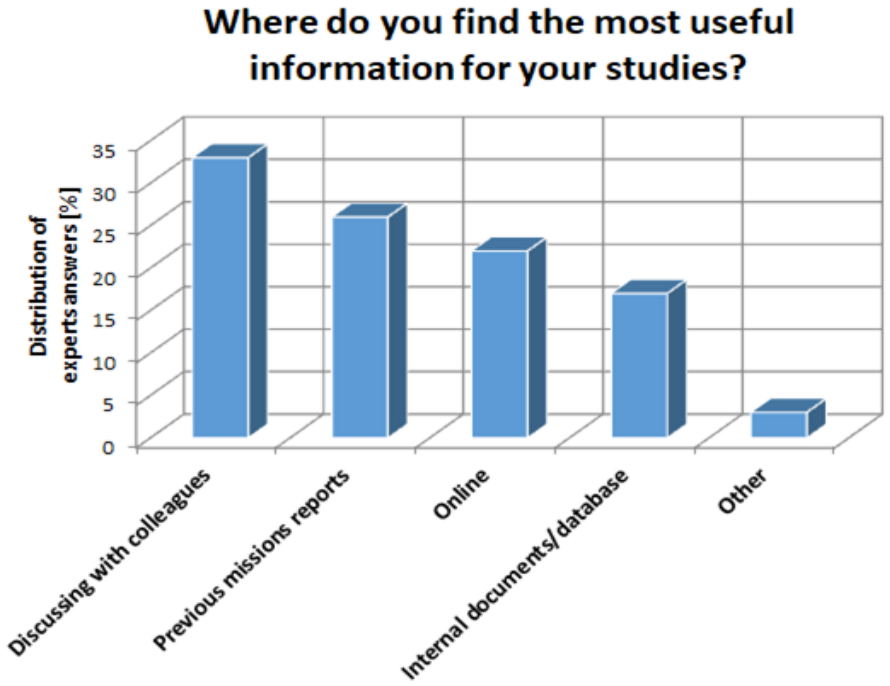
2. Main risk: Poor Knowledge Management leads to "corporate amnesia"

3. Wealth of information with potential to be reused to support current space mission design and accelerate feasibility studies preparation

Expert Interviews

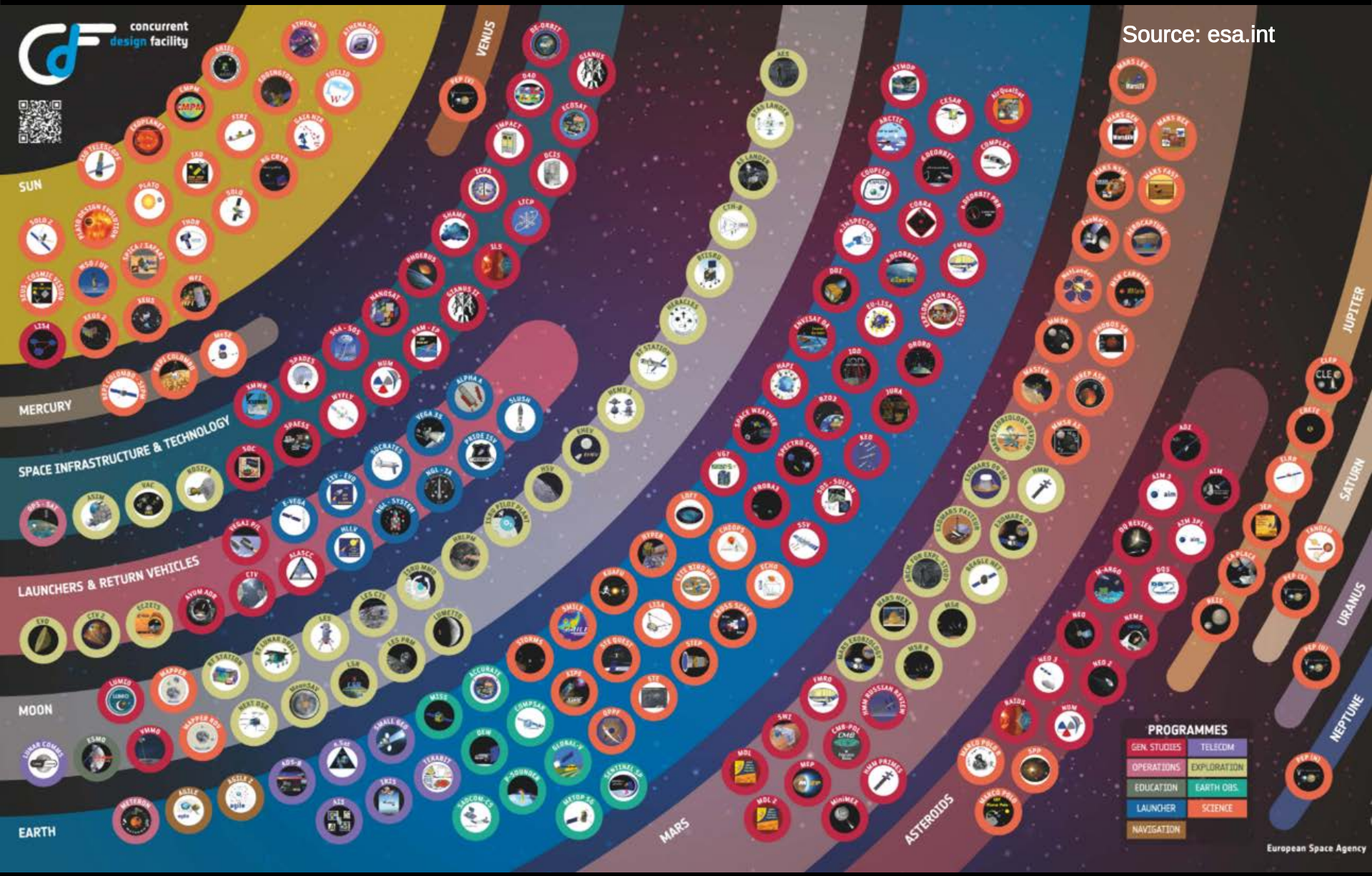
Organisation of interviews with ESA experts involved in Concurrent Design studies, to better understand the Users' needs and work processes. In total, 47 experts were involved :





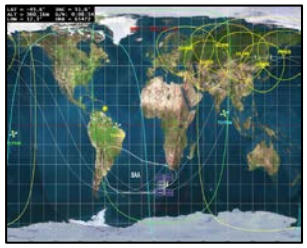
Available Data

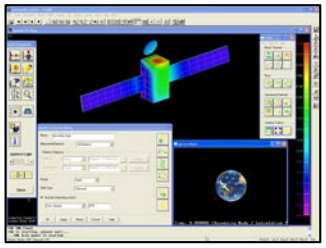
20 years of studies at the CDF



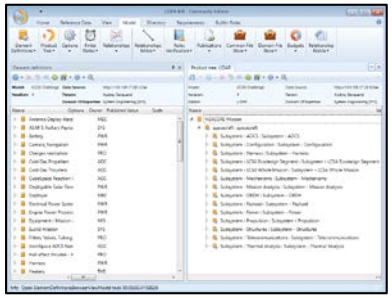
Example of data generated during a feasibility study

Excel tradeoffs, power points





Simulations



Engineering Model based on ECSS-E-TM-10-25A Annex A

Excel calculation sheets, power points

Final report

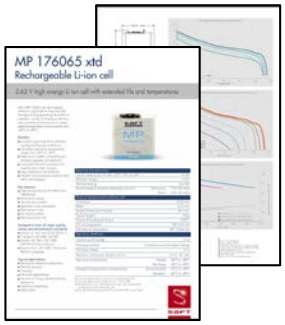
Final presentation

Preparation → **Study** → **Post-Study**

Three classic phases of a feasibility study at ESA Concurrent Design Facility

Beyond the data produced during a study :

Feasibility reports

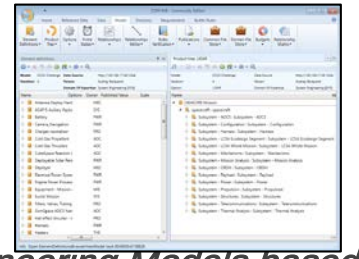


Data sheets

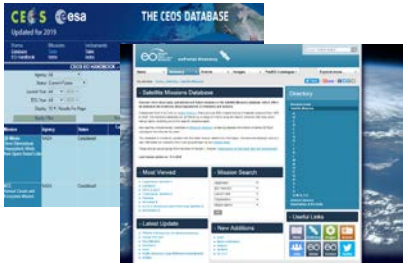
Publications

Textbooks

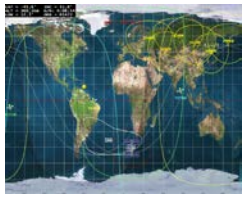
Unstructured Data

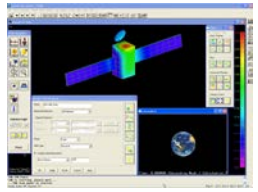


Engineering Models based on ECSS E-TM-10-25A Annex A



Web





Simulations

Semi-structured Data

Insight on the Complexity of Unstructured Data

Why is it so complex to transfer information into machine -readable data?

- Data available in many different formats (and languages)
- Natural language is in nature ambiguous and context dependent (i.e., synonyms, acronyms)
- Humans are not helping (e.g., do not respect templates, use screenshots)

Example of chapters' name inconsistencies over 7 reports :

"Navigation"

"GNC"

"AOGNS and Relative Metrology"

"AOCS"

"Attitude Control System"

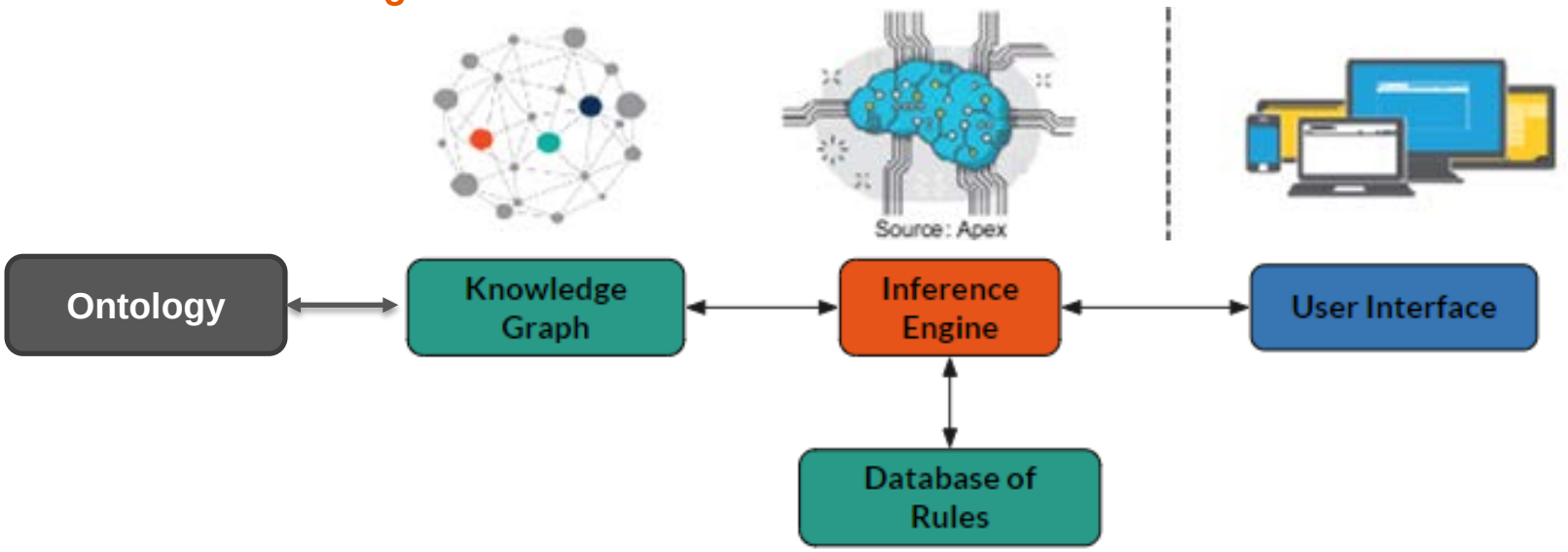
"Attitude and Orbit Control"

The Design Engineering Assistant (DEA)

An expert system for space mission design: DEA

The Design Engineering Assistant (DEA)
=
an expert system to support decision-making at the early stages design of spacecraft, a Knowledge Engine for mission design, facilitating Knowledge Management and Reuse

An Expert System captures Human expertise in a computer program and mimics Human reasoning.



Ontology Learning

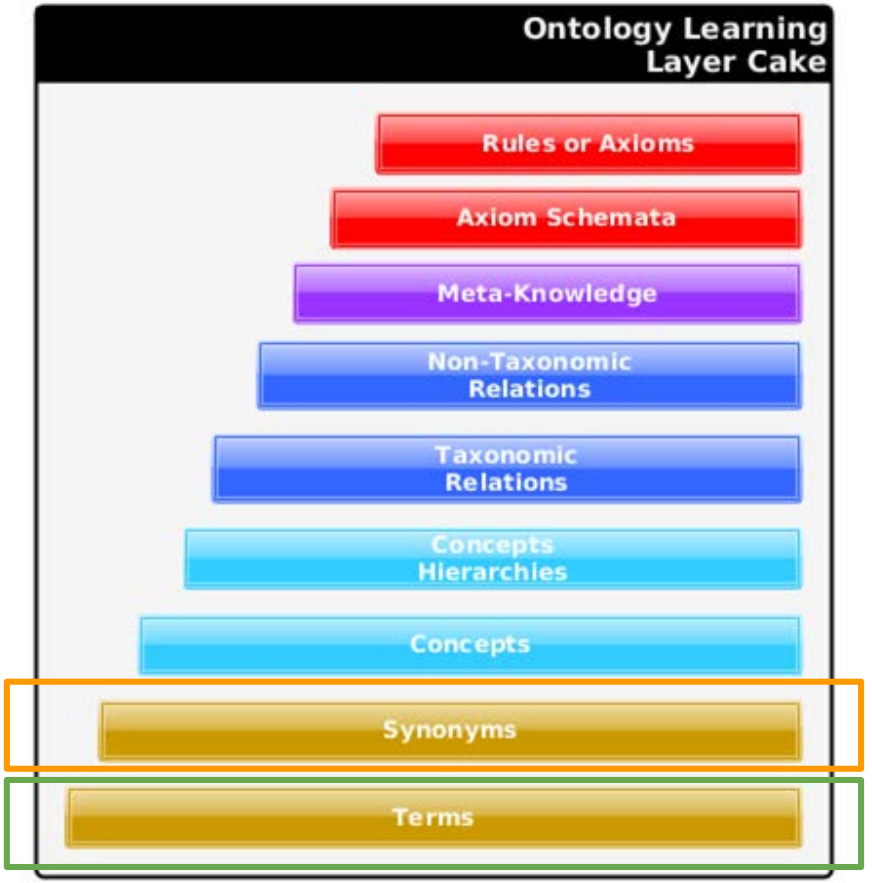
Ontology Learning = automatic/semi - automatic generation of ontologies

Coined in 2001 by Alexander Maedche and Steffen Staab

Set of methods and techniques used for building an ontology in a semi-automatic fashion using several sources, involving:

- extraction of domains and relationships between concepts from a natural language corpus
- encoding the concepts in an ontology language.

Ontology Learning Cake = steps of Ontology Learning



Current Results

Current status of work

Step 1: Natural Language Processing Pipeline- NLTK Python Library

Classic tasks of preparing the input text for further analysis: tokenization, removal of stop words, abbreviation expansion, lemmatization

Step 2: Candidate Entities Identification

generate a domain-specific vocabulary from the unstructured text

Step 3: Merge Similar/Synonyms Entities

Via **word embedding**, a set of Natural Language Processing methods, allowing to map the context of a term into vectors.

Entities Identification Methodology

Generate a domain-specific vocabulary from the unstructured text

1. Filter based on word frequency

1. Second filter based on Weirdness Index: Compare the use of a word, based on its frequency, between a domain-specific corpus and a set of corpus representing the general language. In this case, the British National Corpus (BNC).

$$W = \frac{N_G f_S}{(1 + f_G) N_S}$$

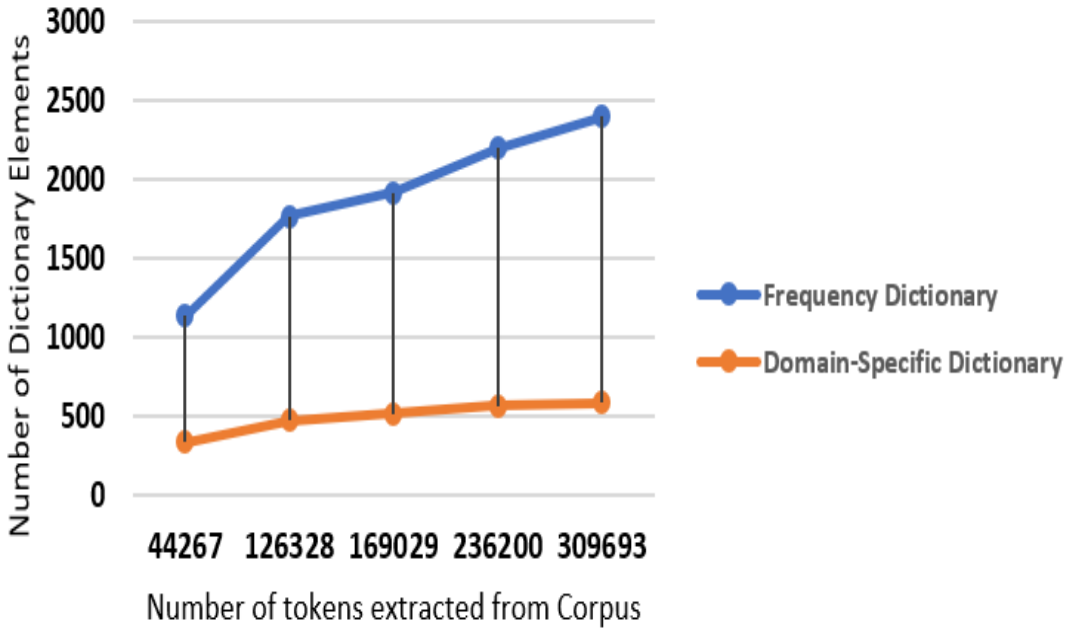
where fs is the frequency of the word in the specialized corpus, fg its frequency in the general corpus, the BNC, and Ns, Ng are respectively the number of tokens in the specialized and general corpus

(reference Ahmad, K., and Gillam, L., "Automatic Ontology Extraction from Unstructured Texts," Vol. 3760, No. October, 2005. doi:10.1007/11575771)

Results

Results over 9 CDF feasibility reports (available publicly)

Frequency Dictionary	Domain-Specific Dictionary
study	cdf
report	spacecraft
design	esa
cdf	data
mission	telescope
mass	orbit
system	baseline
requirement	payload
spacecraft	thermal
power	thruster



** corresponds to 1, 3, 5, 7 and 9 reports*

Comparison with WordNet

WordNet: lexical database, gathers similar concepts into synsets. developed by Princeton University, Accessible via NLTK Python library.

Noun

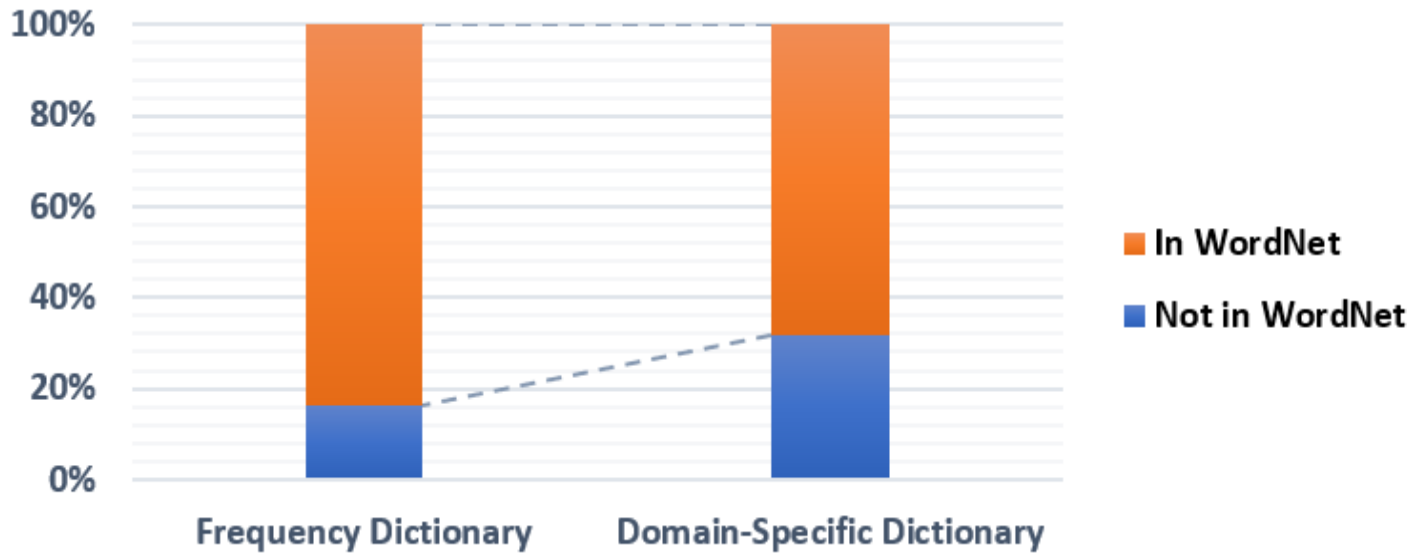
- S: (n) **satellite**, artificial satellite, orbiter (man-made equipment that orbits around the earth or the moon)
- S: (n) **satellite**, planet (a person who follows or serves another)
- S: (n) **satellite** (any celestial body orbiting around a planet or star)

Noun

- S: (n) **spacecraft**, ballistic capsule, space vehicle (a craft capable of traveling in outer space; technically, a satellite around the sun)

Comparison with WordNet

WordNet: lexical database, gathers similar concepts into synsets. developed by Princeton University, Accessible via NLTK Python library.



**results generated with 9 feasibility reports publicly available*

Comparison with WordNet - ECSS exercise

Used the ECSS Dictionary of terms , containing 2,130 terms:

- 20% entities were found in WordNet:

  e.g., *'fail-safe', 'system', 'decision', 'metric', 'migration', 'operational',- 'real time', 'software', 'validation', 'verification', 'port', 'unit', "actuator', 'cleanliness'*

- 80% entities were not found in WordNet

e.g., 'derating', 'performance', 'rating', 'concurrent engineering', '"configurable', 'controllability', 'convective', 'coprocessor', 'entropic', 'interfaced', 'interferometry', 'ionospheric', 'radiometric', 'realignment', 'spaceport', 'superconductor'

Word Embedding Method

Word embedding = a set of NLP methods, allowing to map the context of a term into vectors.

Currently implemented for the DEA project:

Hyper Analogue to Language (HAL) space + cosine similarity

HAL space method:

Sentence: *"The spacecraft was launched from Kourou on an Ariane."*

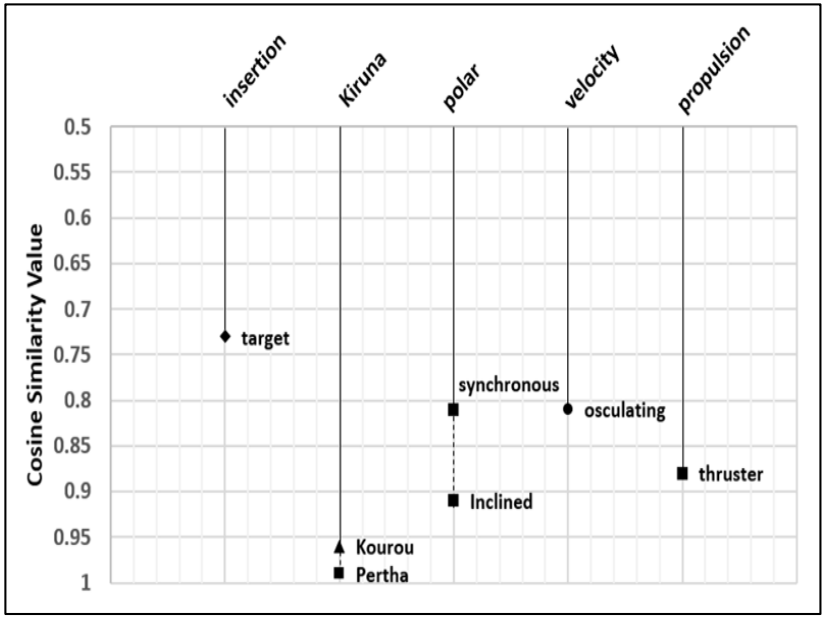
		Launched	Kourou	…	n
Co-occurrence vectors:	Spacecraft	(w1	w2	…	wn)

Cosine Similarity in between vectors to identify concepts used in similar contexts.

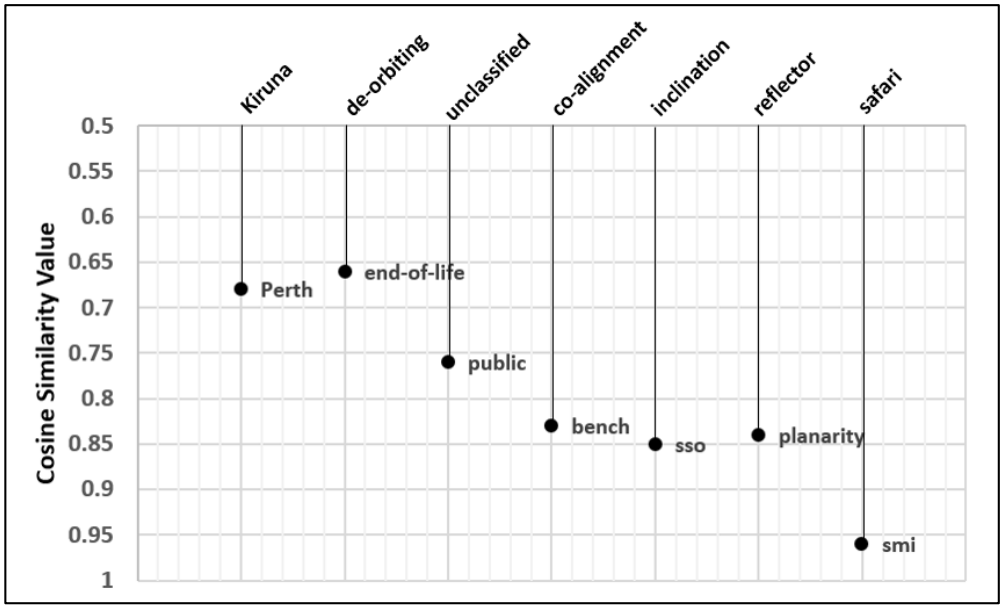
Results

Results generated with 9 feasibility reports publicly available,
Example of Extracted Similar Concepts based on:

Mission Analysis Chapters only



Executive Summaries only

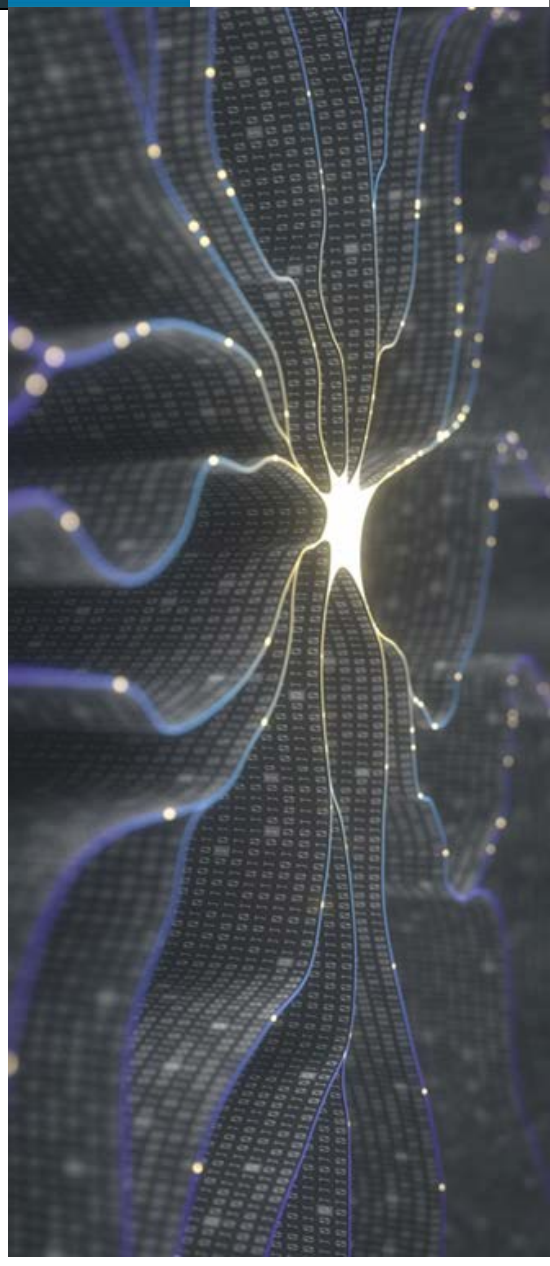


Conclusion

Encouraging preliminary results in the generation of candidate entities and identification of synonymous concepts

Results to be generated with finalised NLP pipeline, an extended corpus and implementation of advanced method (word 2vec)

Ontology Learning today needs Human validation



Thank you for your attention, questions?

contact:

audrey.berquand@strath.ac.uk

annalisa.riccardi@strath.ac.uk

Research gate: https://www.researchgate.net/project/Design-Engineering-Assistant-DEA-for-Space-Mission-Design

Scotland from the ISS- Image Credit: ESA/NASA

University of
Strathclyde
Glasgow

Extra slides

Publications

Published :

Towards an Artificial Intelligence based Design Engineering Assistant for the Early Design of Space Missions *A. Berquand (UoS), F. Murdaca (UoS), Dr. A. Riccardi (UoS), T. Soares (ESA) S. Gerené (RHEA), N. Brauer (AIRBUS), K. Kumar (satsearch),* ***IAC 2018, Bremen, Germany***

Ontology -Based Information Extraction from Datasheets of space parts
F. Murdaca (UoS), A. Berquand (UoS), K. Kumar (satsearch), Dr. A. Riccardi (UoS), T. Soares (ESA), S. Gerené (RHEA), N. Brauer (AIRBUS), ***SECESA 18, Glasgow, UK***

SECESA 18: Artificial Intelligence for Early Design of Space Missions in support of Concurrent Engineering sessions
F. Murdaca (UoS), A. Berquand (UoS), Dr. A. Riccardi (UoS), T. Soares (ESA), S. Gerené (RHEA), N. Brauer (AIRBUS), ***SECESA 18, Glasgow, UK***

Artificial Intelligence for the Early Design Phases of Space Missions
A. Berquand (UoS), F. Murdaca (UoS), Dr. A. Riccardi (UoS), T. Soares (ESA), S. Gerené (RHEA), N. Brauer (AIRBUS), K. Kumar (satsearch), ***IEEE Aerospace 2019, Montana, US***

All publications:

Strathclyde Knowledge Portal: https://pureportal.strath.ac.uk/en/projects/design-engineering-assistant-dea-for-space-mission-design

Research gate: https://www.researchgate.net/project/Design-Engineering-Assistant-DEA-for-Space-Mission-Design

An expert system for space mission design: DEA

The Design Engineering Assistant (DEA)
=
an expert system to support decision-making at the early stages design of spacecraft

Our vision:
Create a Knowledge Engine for Space Mission Design, facilitating Knowledge Management and Reuse

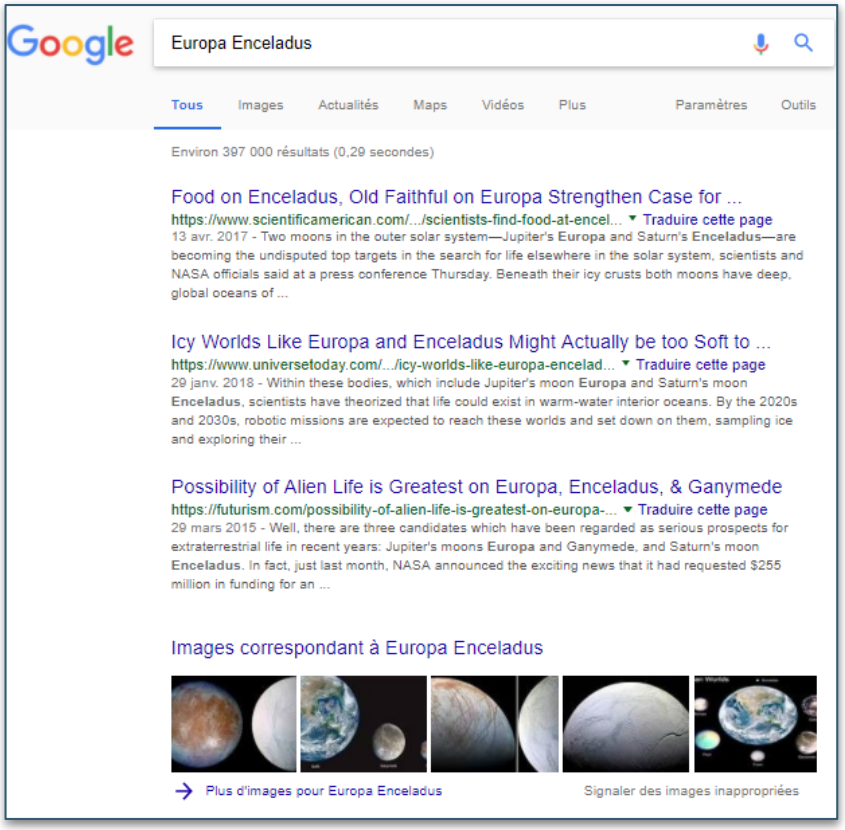
Development steps:

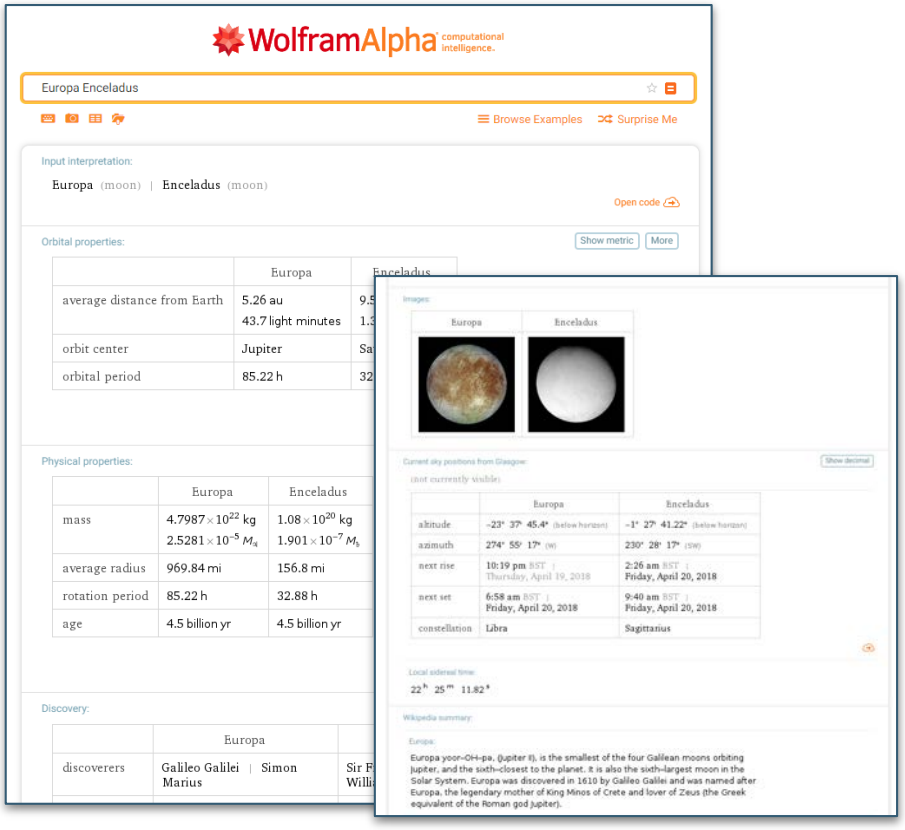
- ➔ Build an ontology from unstructured data using Natural Language Processing and Ontology Learning
- ➔ Assimilate and merge heterogeneous data into the DEA Knowledge Base
- ➔ Develop a User Interface

Project duration: January 2018 - December 2020

Partners:

Search Engine vs Knowledge Engine





Output of Google Search and Wolfram Alpha for the same query: “Europa Enceladus”

Comparison with WordNet

WordNet: lexical database, gathers similar concepts into synsets. developed by Princeton University, Accessible via NLTK Python library.

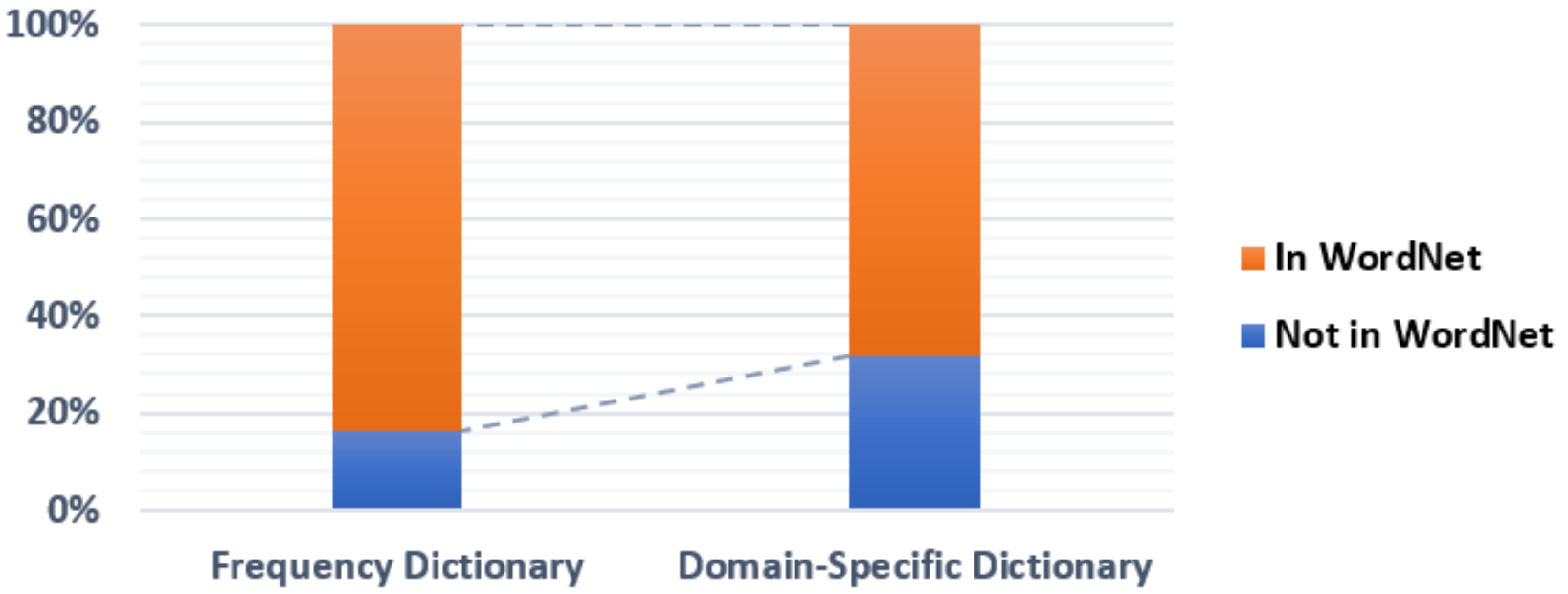
Noun

- S: (n) **satellite**, artificial satellite, orbiter (man-made equipment that orbits around the earth or the moon)
- S: (n) **satellite**, planet (a person who follows or serves another)
- S: (n) **satellite** (any celestial body orbiting around a planet or star)

Noun

- S: (n) **spacecraft**, ballistic capsule, space vehicle (a craft capable of traveling in outer space; technically, a satellite around the sun)

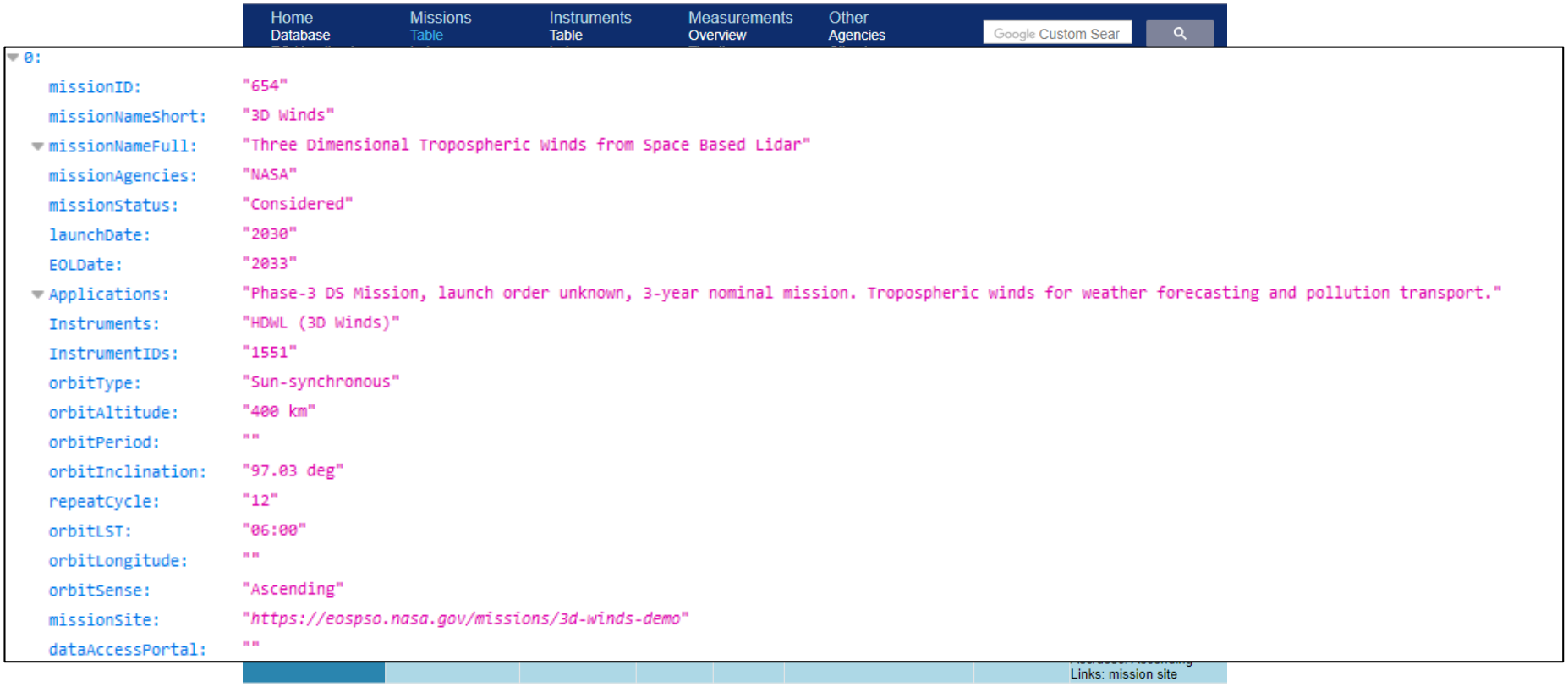
Comparison with WordNet



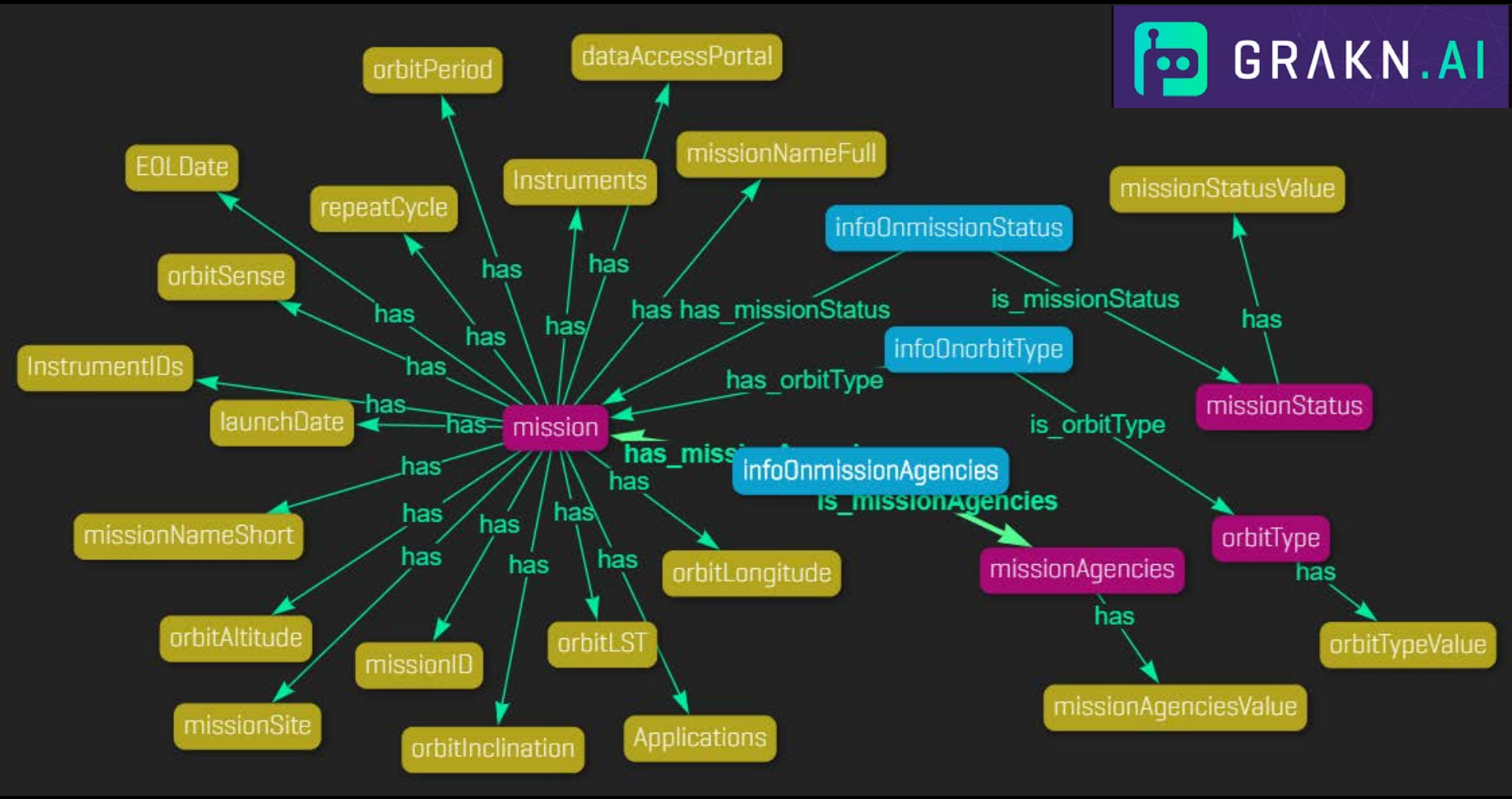
**results generated with 9 feasibility reports publicly available*

Example - the CEOS database

Committee on Earth Observation Satellites database:
Survey of Earth Observation space missions, updated every years.
http://database.eohandbook.com/

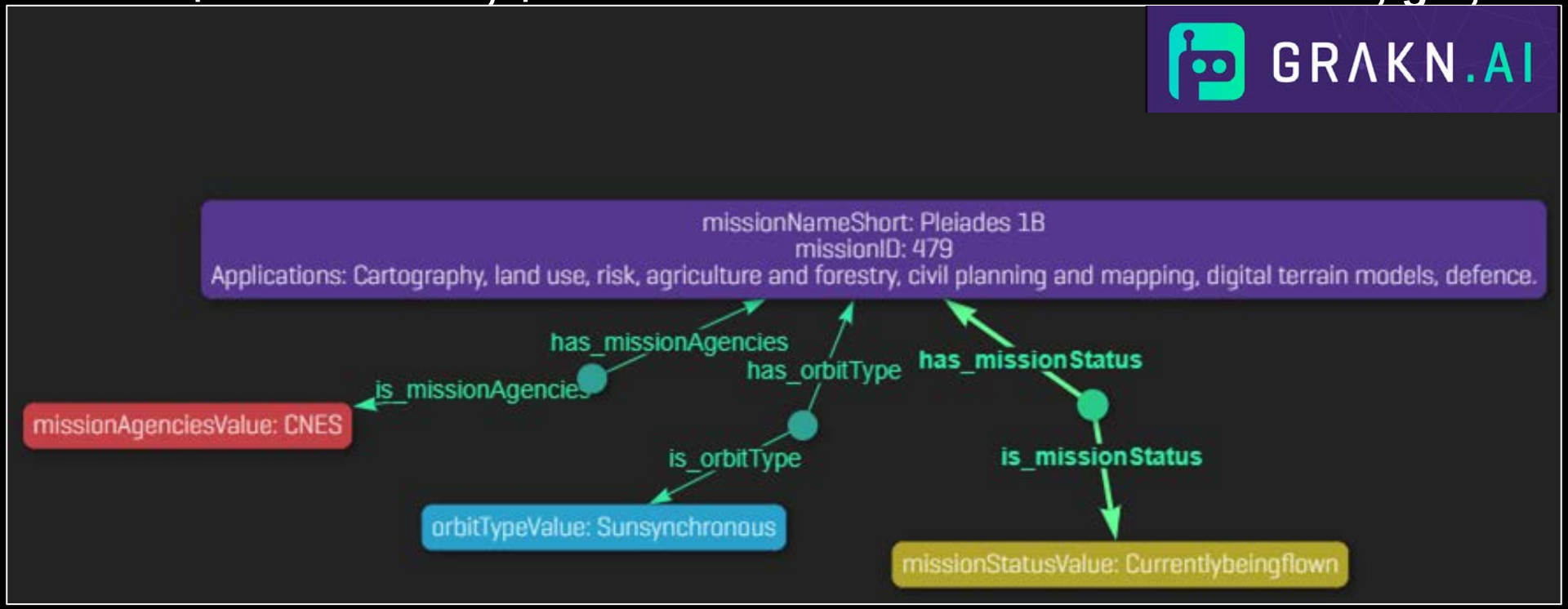


Schema Layer Visualisation



Query Example: Show all information on Pleiades 1B

```
match $x isa mission; $x has missionNameShort "Pleiades 1B"; get;
```



Comparison with WordNet - ECSS exercise

Used the ECSS Dictionary of terms , containing 2130 terms:

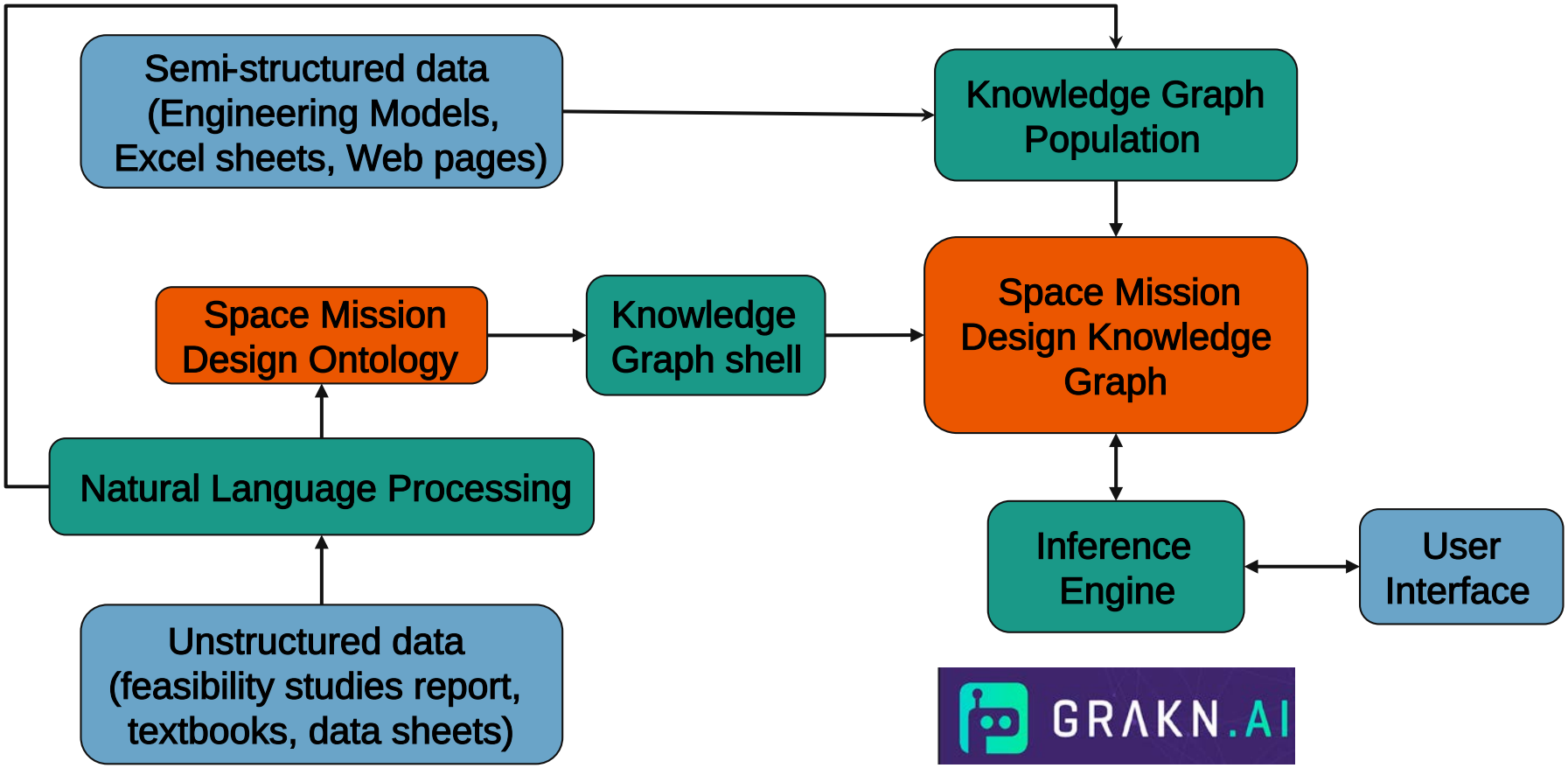
- 20% entities were found in WordNet:

  e.g., *'fail-safe', 'system', 'decision', 'metric', 'migration', 'operational', 'real-time', 'software', 'validation', 'verification', 'port', 'unit', "actuator', 'cleanliness'*

- 80% entities were not found in WordNet

  e.g., 'derating', 'performance', 'rating', 'concurrent engineering', "configurable', 'controllability', 'convective', 'coprocessor''entropic', 'interfaced', 'interferometry', 'ionospheric', 'radiometric', 'realignment', 'spaceport', 'superconductor'

The architecture of the DEA



Natural Language Processing (NLP):

A branch of AI which enables computers to understand, interpret, process and manipulate human (natural) language.

Ontology

Ontology = modelisation of the domain knowledge

Provide a **common vocabulary field and concepts definition** to facilitate the communication between experts of a same field extending the communication to human-machine and machine-machine.

The most popular definition of ontology in information technology and the AI community, provided by Tom Gruber (1993), is: "*An ontology is a formal, explicit specification of a sharedconceptualization*".

consensus on abstract representation of concepts accepted by different communities

An ontology should be machinereadable.

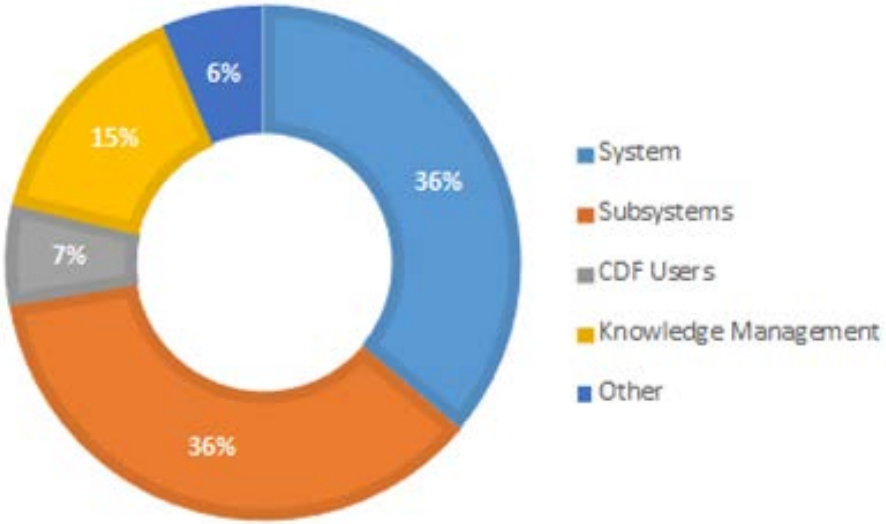
concepts definition and the constraints on their use are explicitly defined

The **basic elements**, or representational primitives, of an ontology are :

- **classes**: representing a "thing" or a "concept"
- **attributes** (or properties)
- **relationships** : linking the different classes

User-centred approach

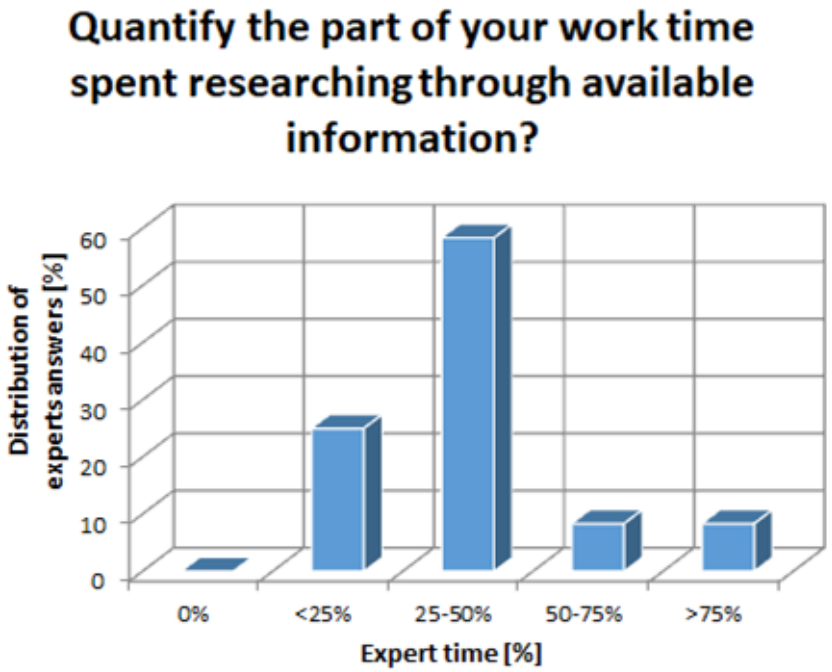
Organisation of interviews with ESA experts involved in Concurrent Design studies,
To better understand the Users' needs and work processes. In total, 47 experts were involved:

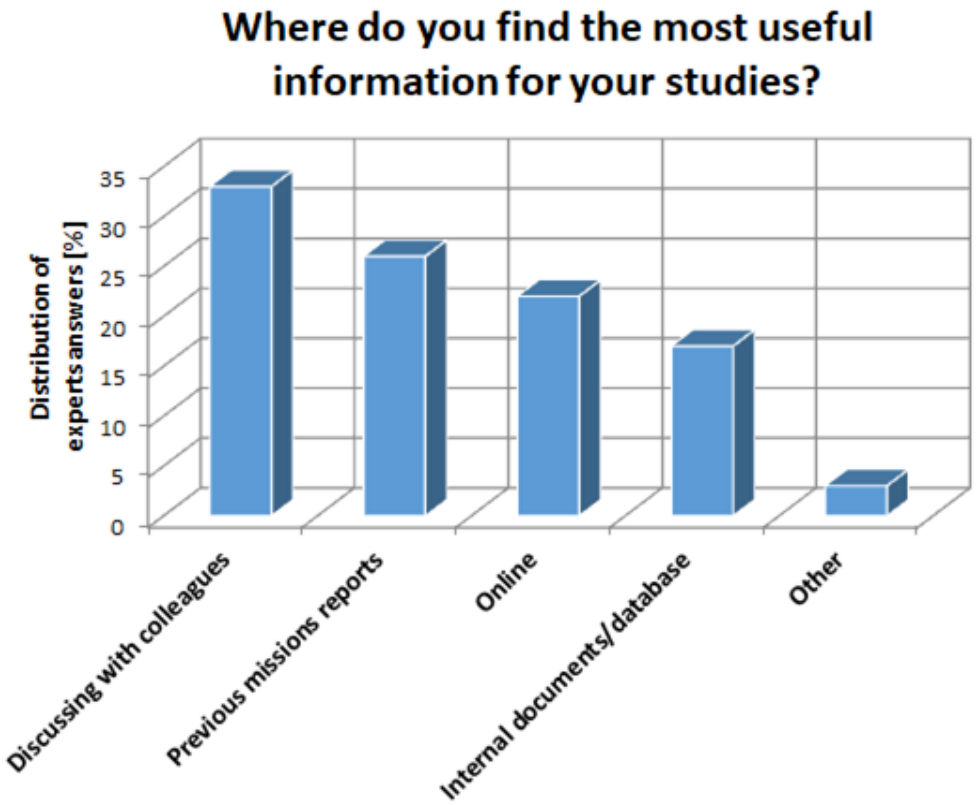


*The full interviews results are presented in: "Towards an Artificial Intelligence based Design Engineering Assistant for the Early Design of Space Missions" *A. Berquand (UoS), F. Murdaca (UoS), Dr. A. Riccardi (UoS), T. Soares (ESA) S. Gerené (RHEA), N. Brauer (AIRBUS), K. Kumar (satsearch), IAC 2018, Bremen, Germany*

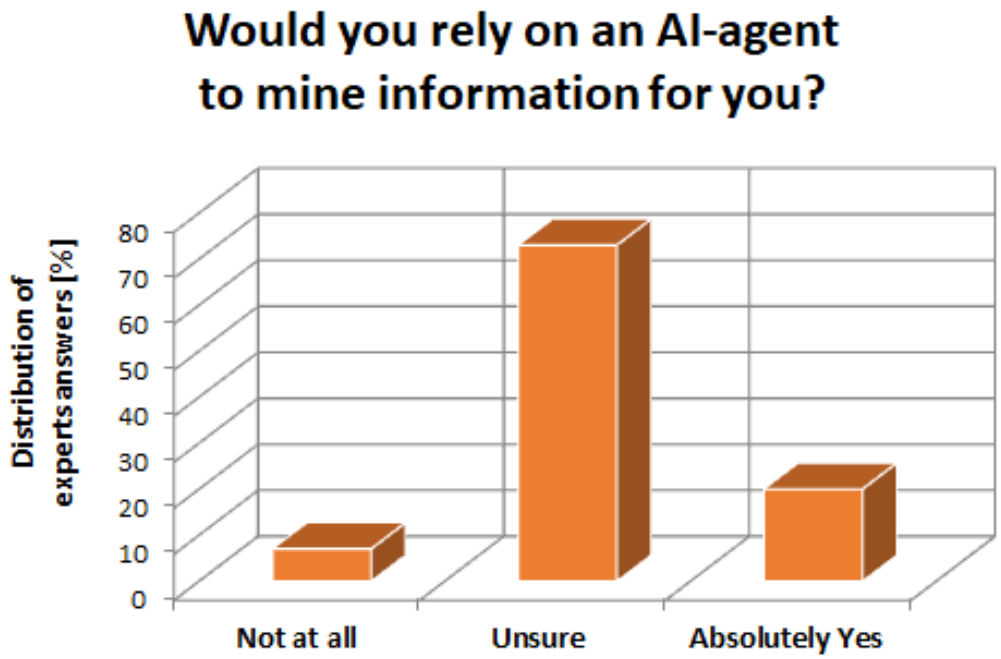
**The DEA requirements resulting from the user-centred approach are presented in: "Artificial Intelligence for the Early Design Phases of Space Missions" *A. Berquand (UoS), F. Murdaca (UoS), Dr. A. Riccardi (UoS), T. Soares (ESA) S. Gerené (RHEA), N. Brauer (AIRBUS), K. Kumar (satsearch), IEEE Aerospace 2019, Montana, US*

Expert Interviews Main Outcomes (1/2)





Expert Interviews Main Outcomes (2/2)



Grakn in a nutshell (1/2)

Grakn is an open-source, distributed Knowledge Graph.
Graql is Grakn's reasoning and analytics query language.
→ software architecture trade -off explained in Part III

A Grakn Knowledge Graph is a schema layer and a data layer.

define
entity: object
attribute: attributes associated with domain instances
relationship: between different domain instances
role: roles involved in specific relationships

insert
entity: instances of entity types (e.g. insert $x isa person, with 'person' an entity type)
resources: instances of attribute types (e.g. has name "John Doe")
relations: instance of relationships types

Grakn in a nutshell (2/2)


```
define

 "ElementDefinition" sub entity
    has name;

"Spacecraft" sub ElementDefinition
    plays TopElement;

 "Equipment" sub ElementDefinition
    plays Element
    plays ElementWithParameter;

"Parameter" sub entity
    plays ParameterofElement
    has valueSet # e.g. 160
    has parameterType # e.g. Data Volume
    has scale; # e.g. Gibit

"name" sub attribute datatype string;
"valueSet" sub attribute datatype string;
"parameterType" sub attribute datatype string;
"scale" sub attribute datatype string;

"ElementUsage" sub relationship
    relates TopElement
    relates Element;

"ParameterRelationship" sub relationship
    relates ElementWithParameter
    relates ParameterofElement;
```

Schema layer

Data layer

```
insert
$doveSC isa Spacecraft, has name "DoveSpacecraft";
$HRCamera isa Equipment, has name "HRCamera";
$dataVolume isa Parameter, has parameterType "Data Volume", has valueSet "160",has scale "GBit";
(TopElement: $doveSC, Element: $HRCamera) isa ElementUsage;
(ElementWithParameter:$HRCamera,ParameterofElement:$dataVolume) isa ParameterRelationship;
```

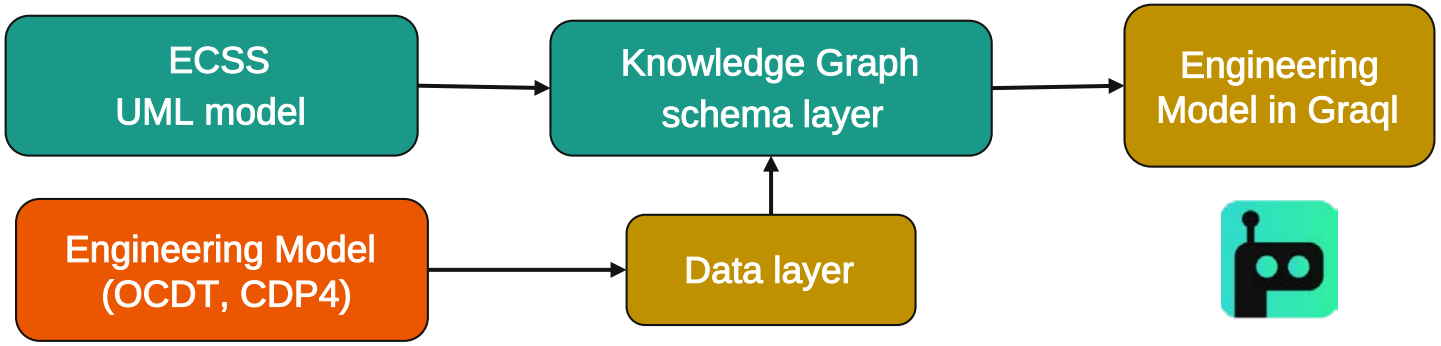
Semi-structured Data Extraction

A source of structured data is the Engineering Models, containing all the information about the design iterations and options.

The Engineering Model structure is based on the ECSS-E-TM-10-25A Annex A UML model and can be exported as JSON files.

How to make the Models usable by the DEA?

Migrate them into a Knowledge Graph.



Output

132 Entities
108 Attributes
194 Relationships