

SemWeb.Pro 2024, Paris

# Données ouvertes du Parlement Européen

- retour d'expérience sur la mise au point et la diffusion d'un "graphe de connaissances" institutionnel

Laura Liudvinavičiūtė, Thomas Francart – 21 November 2024

# Overview

- 1 EP open data
- 2 Context
- 3 Our ambitions
- 4 Journey towards semantics
- 5 Implementation insights
- 6 SHACL as a foundation for the KG





# EP Open Data Portal: at a glance

- **Dataset Catalogue**

*Features:* DCAT dataset metadata, downloadable dumps

(multilingual RDF, monolingual CSV, ...)

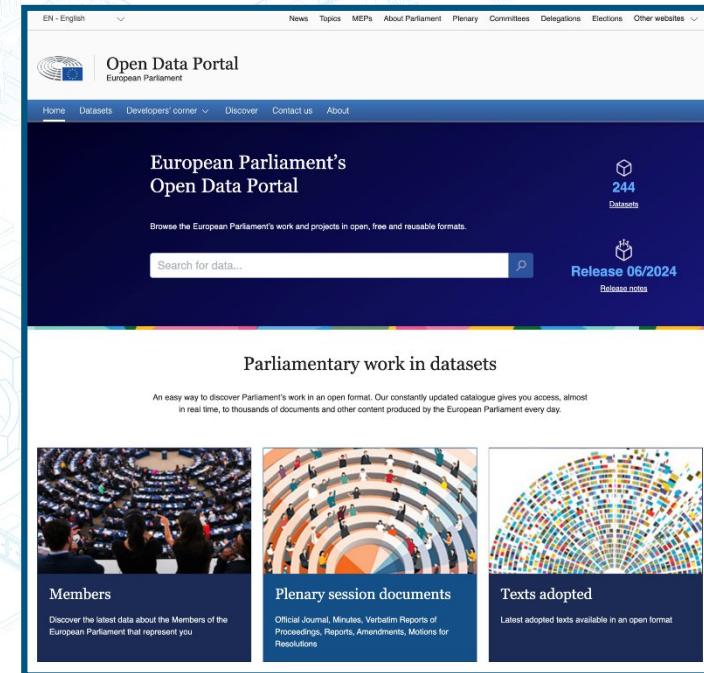
*Key users:* non-IT users, researchers, data scientists, journalists

- **Near-real time RESTful APIs & Persistent URIs**

*Features:* data in multiple formats (RDF, json-ld, ...), Notification (ATOM feed) based on provenance

*Key users:* semantic web specialists, web developers (non-RDF)

- **Human & machine readable documentation**



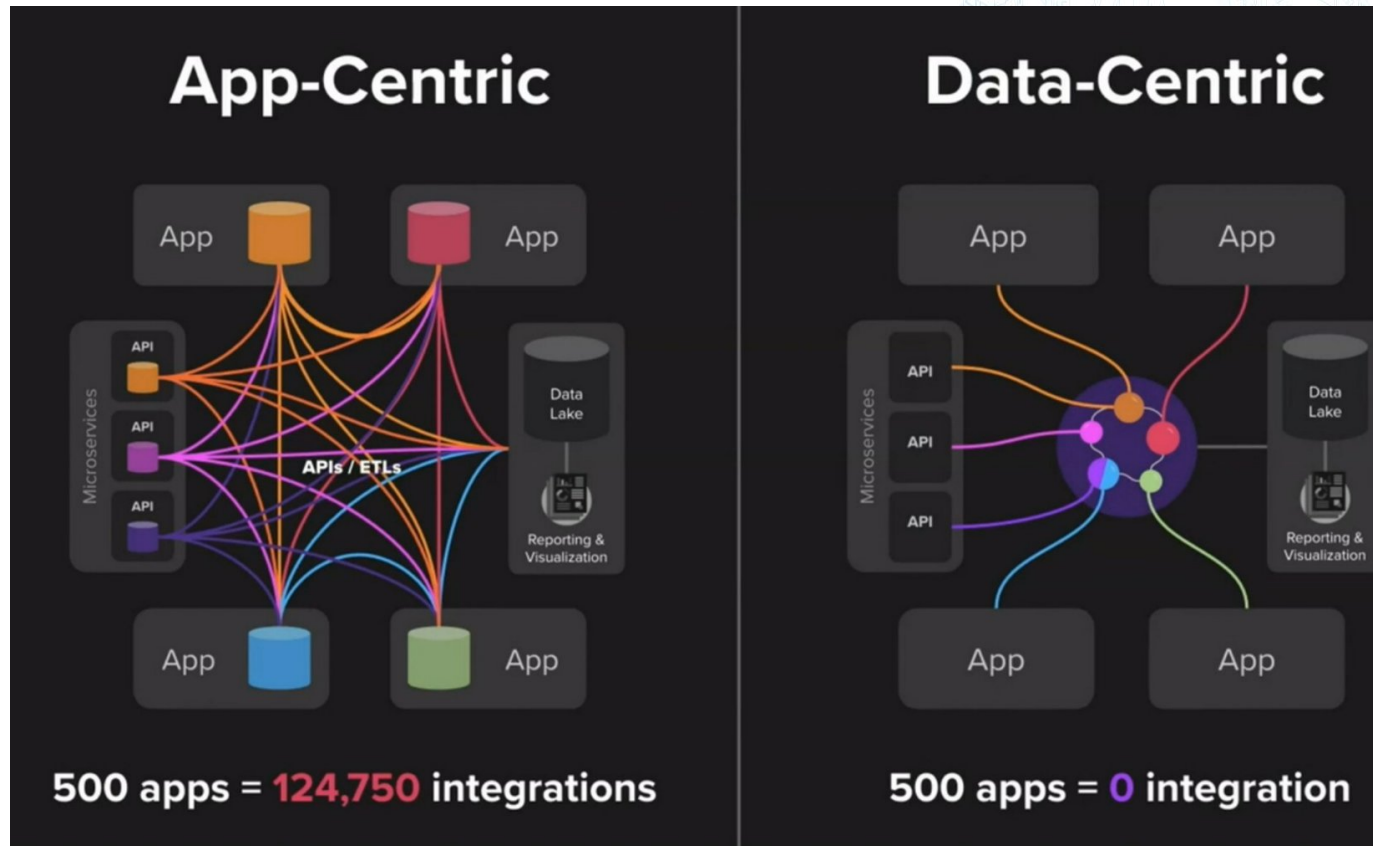
<https://data.europarl.europa.eu>

# Interoperability challenges

- **Scattered & duplicated data**
- **Data model silos** unique to each IT system
- **Multiple data sources** (20+ production systems) with **diverse connectors & complex workflows** for managing core business processes (parliamentary committees, plenary sessions, ...)
- **Diverse input formats** (word, pdf, json, sgm, xml, akn, ...)
- **Multiple & diverse data sources & formats for legacy data** (from ~1979)
- **Multiple (24) EU languages**

# Shift towards data-centric approach

Design **data models** remaining **valid long after the application is gone.**

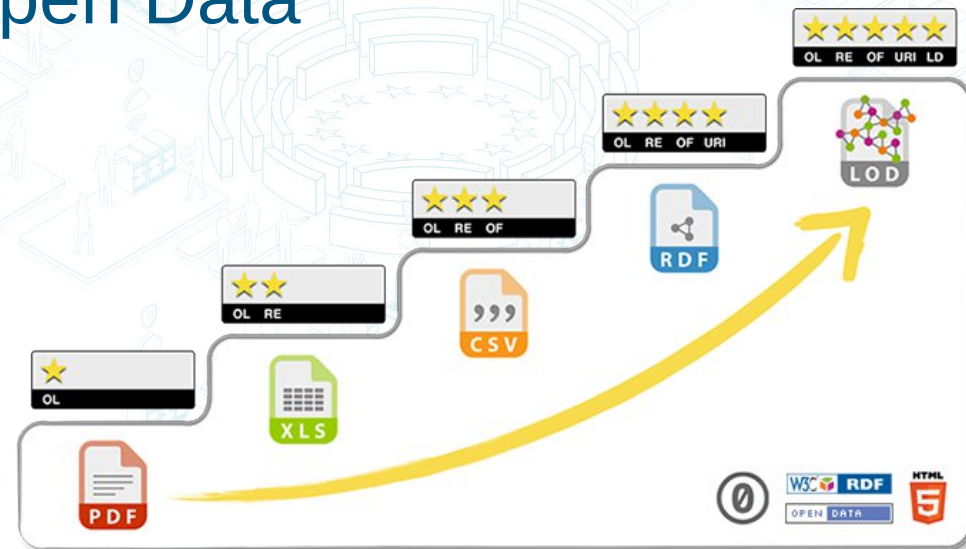
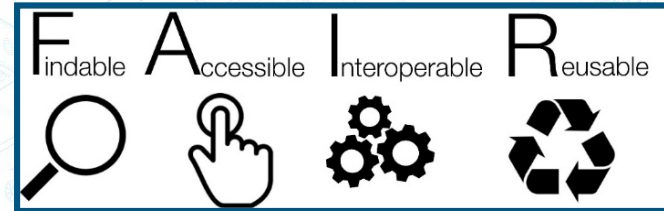


[twitter.com/SimenMunter](https://twitter.com/SimenMunter)



# Be open, findable, interoperable & linked

- Compliance with F.A.I.R. data principles
- Compliance with 5-star Linked Open Data (LOD) framework



 [5stardata.info](https://5stardata.info)

# Our semantic modelling journey

## *Open World Assumption*



## *Closed World Assumption*

**Ontologies (OWL):** Semantic Web language designed to **represent rich and complex knowledge** of:

- cross-domains at web-scale (e.g. *Dublin Core, FOAF, VCARD*)
- one business domain (e.g. *ELI, ELI-DL, ORG*)
- extensions to specific business needs (e.g. *ELI-impact*)

**Application Profiles (SHACL):** W3C language designed to **define the structure of a knowledge graph** for specific applications

### Advantages:

- Reuse existing ontologies
- Enforce specific constraints
- Enables data validation



# Foundation for application profiles

## Re-use (80-90%)

well-established ontologies:

- Dublin Core
- DCAT
- ORG
- FOAF
- SKOS
- ELI (*European Legislation Identifier*)
- ELI-DL (*ELI Draft Legislation*)

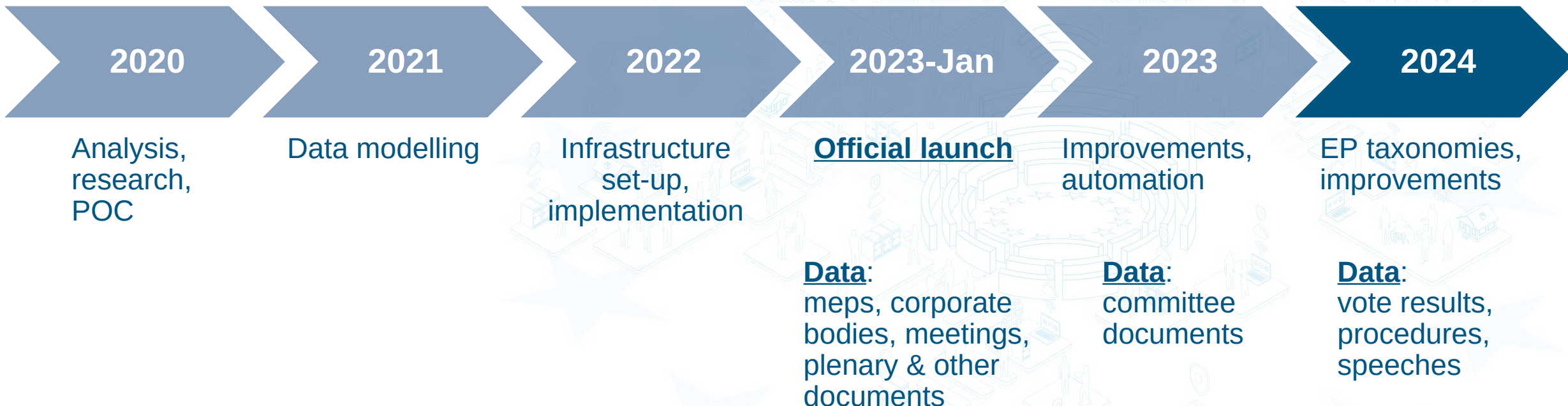


## Create (10-20%)

- **EP ontology** for highly **specific metadata** for which no existing standards were suitable:

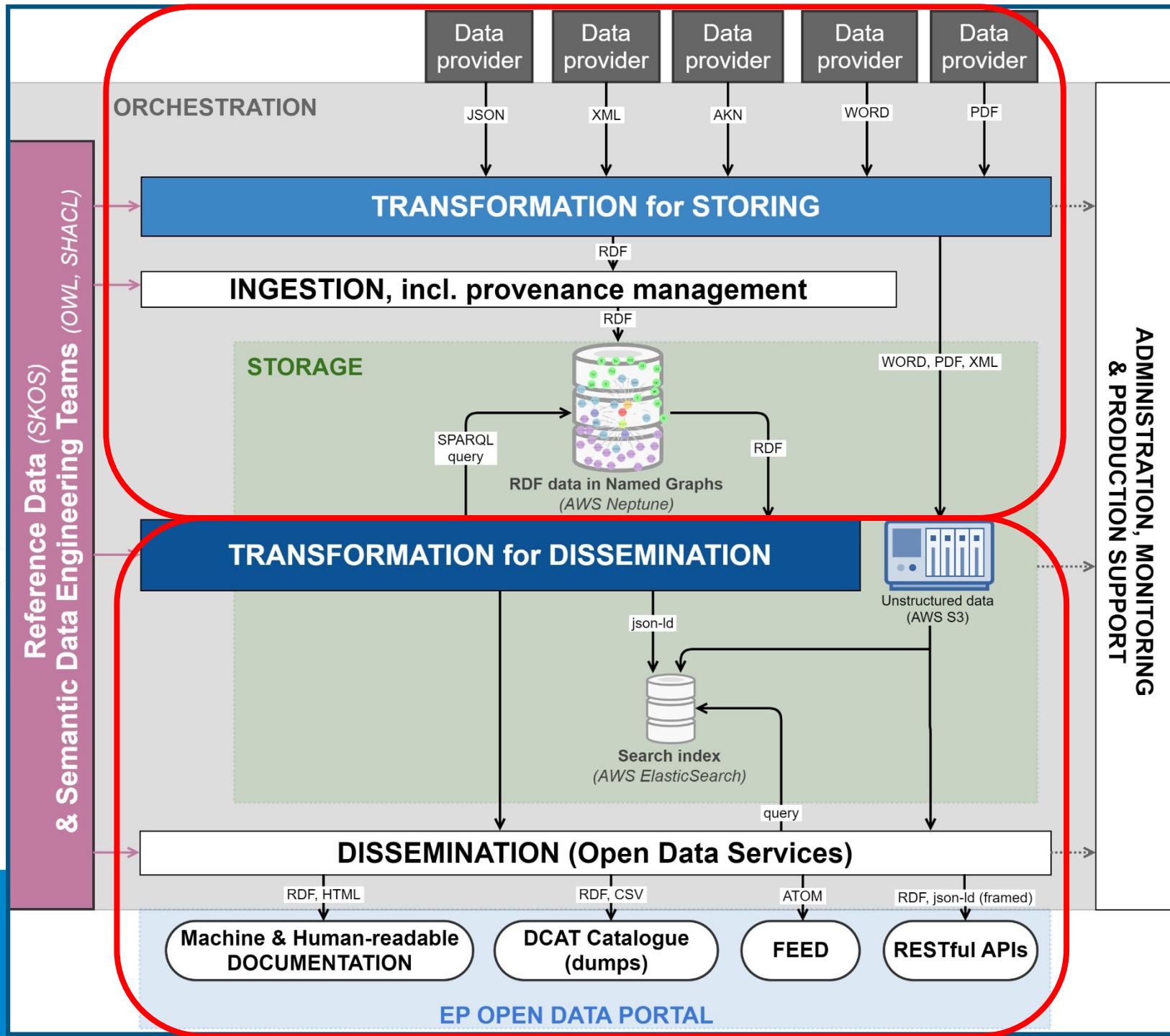
- EPVOC

# EP Open Data Project timeline



## 5. Implementation insights

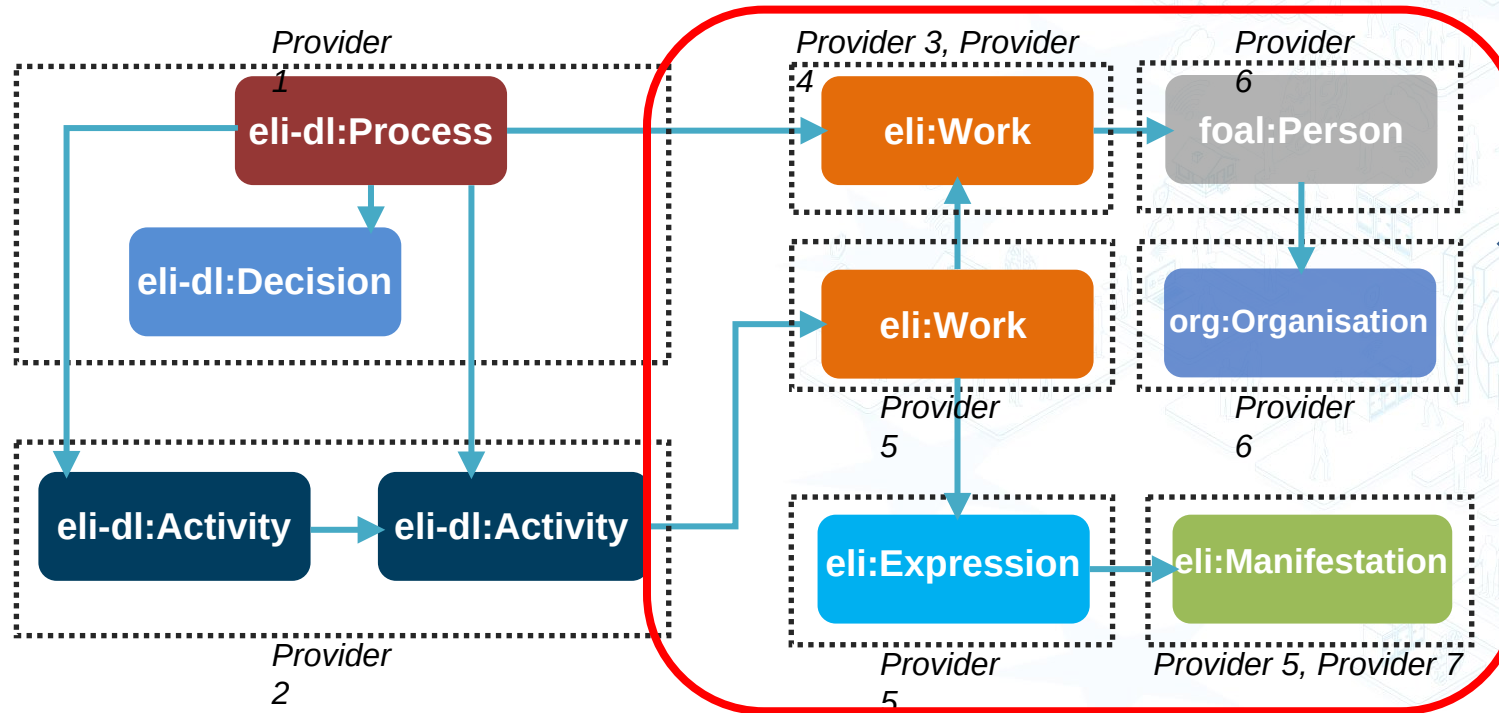
# Data flows architecture





# One knowledge graph, multiple data models

EP named graphs - storage (SHACL)

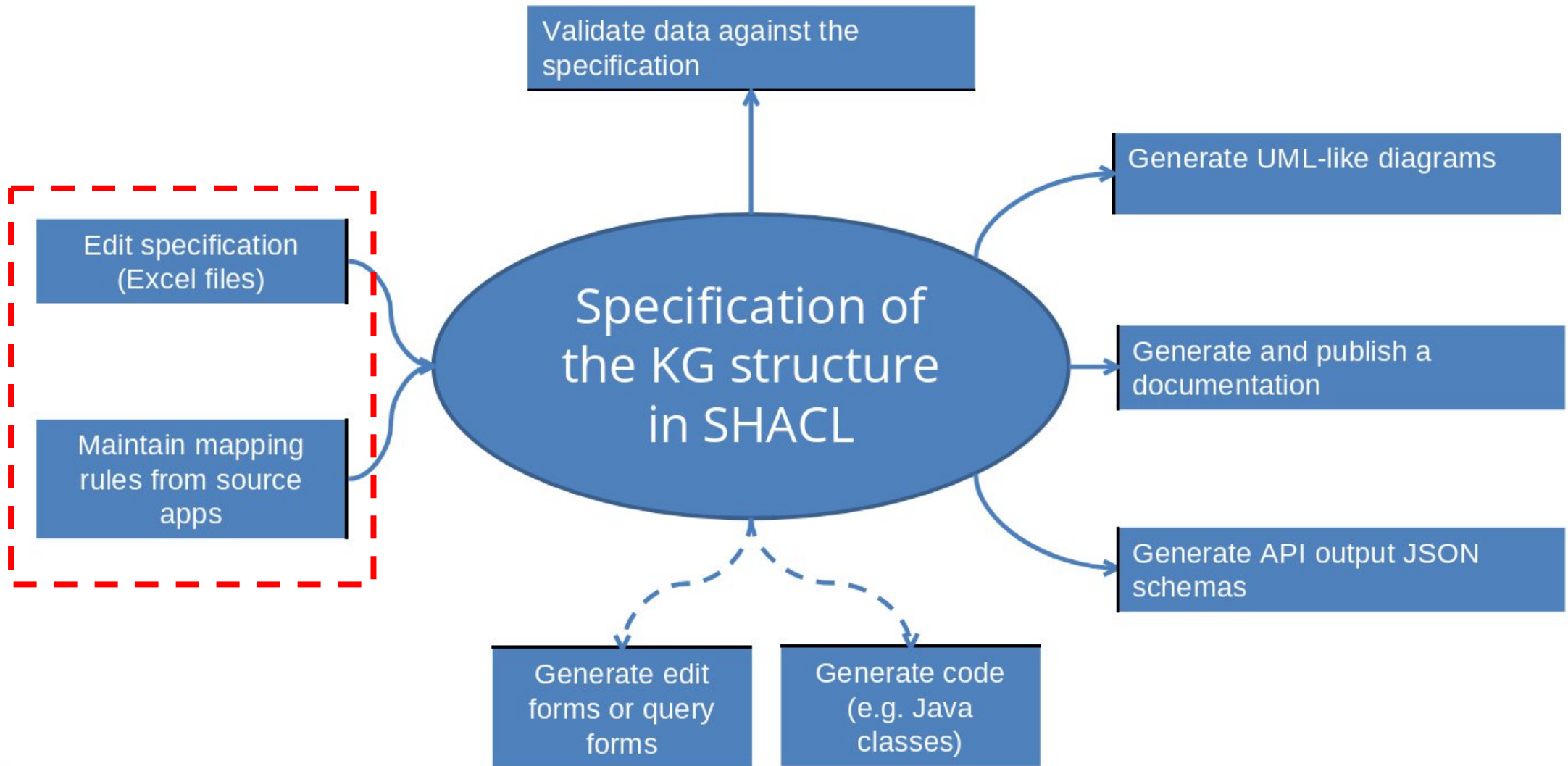


EP datasets - dissemination (SHACL)

- meps
- corporate-bodies
- committee-documents
- plenary-documents
- parliamentary-questions
- adopted-texts
- procedures
- plenary-session-documents
- meetings
- speeches
- vote-results

SHACL comme langage de spécification de la structure du graphe permet d'avoir une réelle approche « **model driven** » du projet.

## 6. SHACL as the foundation for the KG





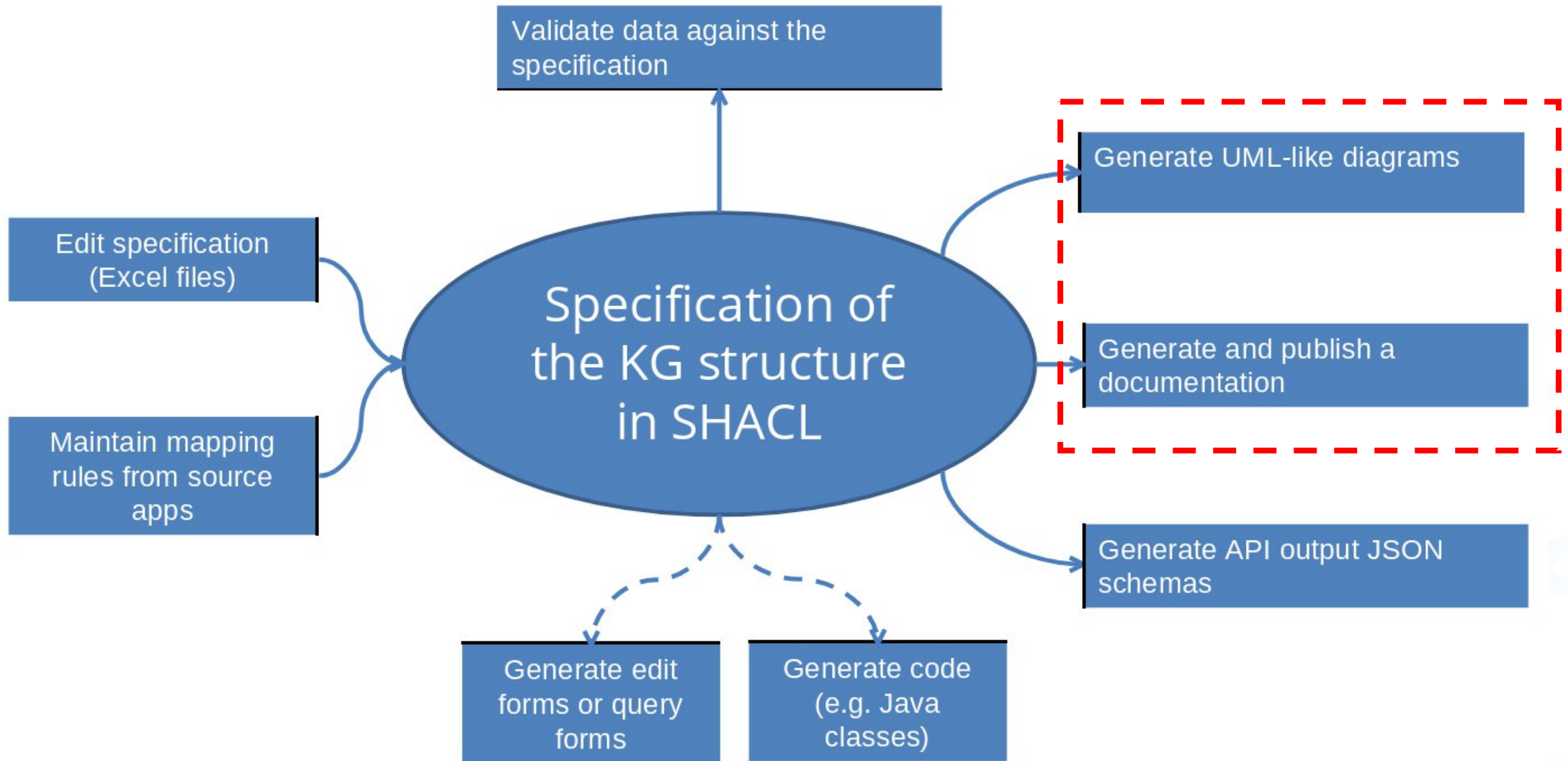
## 6. SHACL as the foundation for the KG

			entities validated by this Shape	Textual description of the Shape		Typical Shape (al)	
26	NodeShape URI	Type of Shape (always NodeShape)			Example URIs for the shape		
27	URI	rdf:type(separator="," )	h:targetClass	rdfs:comment @en	skos:example^^xsd:string		
28	eli:Work	sh:NodeShape, rdfs:Class			<a href="https://data.europarl.europa.eu/eli/dl/doc/A-9-2019-0001">https://data.europarl.europa.eu/eli/dl/doc/A-9-2019-0001</a>	sh	
29	eli:ComplexWork	sh:NodeShape, rdfs:Class			<a href="https://data.europarl.europa.eu/eli/dl/doc/OJ-9-2020-10-21">https://data.europarl.europa.eu/eli/dl/doc/OJ-9-2020-10-21</a>	sh	
30	eli:Expression	sh:NodeShape, rdfs:Class			<a href="https://data.europarl.europa.eu/eli/dl/doc/A-9-2019-0001/en">https://data.europarl.europa.eu/eli/dl/doc/A-9-2019-0001/en</a>	sh	
31	eli:Manifestation	sh:NodeShape, rdfs:Class			<a href="https://data.europarl.europa.eu/eli/dl/doc/A-9-2019-0001/pdf">https://data.europarl.europa.eu/eli/dl/doc/A-9-2019-0001/pdf</a>	sh	
32	eli-dl:Activity	sh:NodeShape, rdfs:Class			<a href="https://data.europarl.europa.eu/eli/dl/event/MTG-PL-2020-10-23">https://data.europarl.europa.eu/eli/dl/event/MTG-PL-2020-10-23</a>	sh	
33	eli-dl:ForeseenActivity	sh:NodeShape, rdfs:Class			<a href="https://data.europarl.europa.eu/eli/dl/event/MTG-PL-2020-10-23-TF-9000">https://data.europarl.europa.eu/eli/dl/event/MTG-PL-2020-10-23-TF-9000</a>	sh	
34	eli-dl:Participation	sh:NodeShape, rdfs:Class			<a href="https://data.europarl.europa.eu/eli/dl/participation/A-9-2021-0226_112611">https://data.europarl.europa.eu/eli/dl/participation/A-9-2021-0226_112611</a>	sh	
35	epvoc:Room	sh:NodeShape, rdfs:Class			<a href="https://data.europarl.europa.eu/room/LOWHEM">https://data.europarl.europa.eu/room/LOWHEM</a>	sh	
36	eli-dl:ParliamentaryTerm	sh:NodeShape, rdfs:Class			<a href="https://data.europarl.europa.eu/org/ep-9">https://data.europarl.europa.eu/org/ep-9</a>	sh	
37	eli:LegalResource	sh:NodeShape, rdfs:Class				sh	
				Remark: Process data will not be yet available. For now, this inverse relation will be provided only in adopted-texts dumps and APIs.			
38	eli-dl:Process	sh:NodeShape, rdfs:Class			<a href="https://data.europarl.europa.eu/eli/dl/proc/2019-0001">https://data.europarl.europa.eu/eli/dl/proc/2019-0001</a>	sh	
39							
40	<b>References to external controlled vocabularies</b>						
41	eli:WorkType	sh:NodeShape, rdfs:Class			<a href="http://publications.europa.eu/resource/authority/resource-type/ADOPT_TEX1">http://publications.europa.eu/resource/authority/resource-type/ADOPT_TEX1</a>	sh	
42	eli-dl:ParticipationRole	sh:NodeShape, rdfs:Class			<a href="http://publications.europa.eu/resource/authority/role/RAPP">http://publications.europa.eu/resource/authority/role/RAPP</a>	sh	
43	dcterms:LinguisticSystem	sh:NodeShape, rdfs:Class			<a href="http://publications.europa.eu/resource/authority/language/LIT">http://publications.europa.eu/resource/authority/language/LIT</a>	sh	
44	dcterms:MediaTypeOrExtent	sh:NodeShape, rdfs:Class			<a href="http://publications.europa.eu/resource/authority/file-type/PDF">http://publications.europa.eu/resource/authority/file-type/PDF</a>	sh	
45	dcterms:MediaType	sh:NodeShape, rdfs:Class			<a href="https://www.iana.org/assignments/media-types/application/pdf">https://www.iana.org/assignments/media-types/application/pdf</a>	sh	
46	euvoc:Country	sh:NodeShape, rdfs:Class			<a href="http://publications.europa.eu/resource/authority/country/FRA">http://publications.europa.eu/resource/authority/country/FRA</a>	sh	
47	dcterms:Location	sh:NodeShape, rdfs:Class			<a href="http://publications.europa.eu/resource/authority/place/FRA_SXB">http://publications.europa.eu/resource/authority/place/FRA_SXB</a>	sh	
48	euvoc:FileStatus	sh:NodeShape, rdfs:Class			<a href="http://publications.europa.eu/resource/authority/file-status/ADOPTED">http://publications.europa.eu/resource/authority/file-status/ADOPTED</a>	sh	
49	org:Organization	sh:NodeShape, rdfs:Class			<a href="http://publications.europa.eu/resource/authority/corporate-body/EP">http://publications.europa.eu/resource/authority/corporate-body/EP</a>	sh	
50	org:Site	sh:NodeShape, rdfs:Class			<a href="http://publications.europa.eu/resource/authority/site/ASP">http://publications.europa.eu/resource/authority/site/ASP</a>	sh	
51	skos:Concept	sh:NodeShape, rdfs:Class			<a href="http://eurovoc.europa.eu/434743">http://eurovoc.europa.eu/434743</a>	sh	
52	eli-dl:ActivityType	sh:NodeShape, rdfs:Class			<a href="http://publications.europa.eu/resource/authority/event/EP_PLENARY_PART">http://publications.europa.eu/resource/authority/event/EP_PLENARY_PART</a>	sh	
53							

## 6. SHACL as the foundation for the KG

Properties on eli-dl:Activity							
51	eli-ep:P51	eli-dl:Activity	eli-dl:had_activity_type	1,1	eli-dl:Activitytype	activity type	The type of a process. The type of a process often determine the workflow being followed to create the law.
52	eli-ep:P52	eli-dl:Activity	eli-dl:activity_date	1,1	xsd:date	activity date	The date of an activity, to indicate a single date. Alternatively, if an activity spans a time range, "activity_start_date" and "activity_end_date" can be used to indicate the beginning and end of the activity.  Dates can be expressed either at a date precision (yyyy-mm-dd), or with time precision (yyyy-dd-mmYhh:mm:ss)
53	eli-ep:P53	eli-dl:Activity	eli-dl:activity_start_date	1,1	xsd:dateTime	activity start date	The date at which an activity started.  Dates can be expressed either at a date precision (yyyy-mm-dd), or with time precision (yyyy-dd-mmYhh:mm:ss)
54	eli-ep:P54	eli-dl:Activity	eli-dl:activity_end_date	1,1	xsd:dateTime	activity end date	The date at which an activity ended.  Dates can be expressed either at a date precision (yyyy-mm-dd), or with time precision (yyyy-dd-mmYhh:mm:ss)
55	eli-ep:P55	eli-dl:Activity	eli-dl:documented_by_a_realization_of	0,n	eli:Work	documented by a realization of	Indicates that the activity was or will be conducted according to the plan described by a Work. Typically these works are agendas that foreseen how future activities will take place.
56	eli-ep:P56	eli-dl:Activity	eli-dl:recorded in realization of	0,n	eli:Work	recorded in realization of	Indicates that an activity was recorded in a related work.  Note : this property is shortcut for an implicit FRBR "Recording Event" that would have created the Recording of the event.
57	eli-ep:P57	eli-dl:Activity	eli-dl:consists of	0,n	eli-dl:Activity	consists of	Indicates a more fine-grained activity that is taking place within this one. This allows to nest activities inside each other, at an arbitrary level of precision.  Example : "Legislative process number 384" consists of "Initiation of the process 384", "First lecture in assemblée nationale of legislative process 384", "First lecture in senate of legislative process 384", "joint committee reading of legislative process 384" and "enactement of legislative process 384".
58	eli-ep:P58	eli-dl:Activity	epvoc:scheduledIn	0,n	eli-dl:ForeseenActivity	scheduled in	Indicates a more fine-grained foreseen activity that is taking place within this one (ex:time slot).
59	eli-ep:P59	eli-dl:Activity	eli-dl:parliamentary_term	0,1	eli-dl:ParliamentaryTerm	parliamentary term	The parliamentary term in which the activity took place
60	eli-ep:P60	eli-dl:Activity	vcard:hasLocality	0,1	dcterms:Location	locality	Used to support property parameters for the locality data property

## 6. SHACL as the foundation for the KG





## EPVOC Ontology

The EPVOC ontology defines specific properties and classes used in the application profiles managed by the European Parliament (DCAT-EP, ELI-EP, ORG-EP).

> [Documentation](#)

## Controlled vocabularies

Controlled vocabularies used throughout the European Parliament's Open Data Portal, following the recommendations of the latest version of DCAT-AP to ensure a minimum level of interoperability.

> [View all available vocabularies \(datasets\)](#)

## Release notes

Discover the latest changes, improvements and bug fixes.

> [View the release notes](#)

## DCAT-EP

DCAT-EP is an extension of the "[DCAT Application Profile for European data portals](#)" ([DCAT-AP](#)), designed and used for documenting datasets and data services of the European Parliament.

As DCAT-AP, DCAT-EP is based on the [W3C Data Catalog Vocabulary \(DCAT\)](#), a metadata standard designed to facilitate interoperability between data catalogues published on the Web.

> [Documentation](#)

## Application Profiles

Application profiles for ORG-EP (Persons, Bodies) and ELI-EP (Documents and Activities) and all related dataset-specific profiles described in human readable documentation and in SHACL.

### ELI-EP Application profile:

> [Documentation](#)

- Texts Adopted: [Documentation](#)
- Parliamentary Questions and Answers: [Documentation](#)
- European Parliament Meetings: [Documentation](#)
- Plenary Documents: [Documentation](#)
- Plenary Session Documents: [Documentation](#)
- Results of votes: [Documentation](#)
- Committee Documents: [Documentation](#)
- External Documents: [Documentation](#)
- Procedures: [Documentation](#)

### ORG-EP Application profile:

> [Documentation](#)

- Members of the European Parliament: [Documentation](#)
- European Parliament Bodies: [Documentation](#)

### SKOS-EP Application profile:

> [Documentation](#)

- Controlled Vocabularies: [Documentation](#)

## 6. SHACL as the foundation for the KG

### Activity

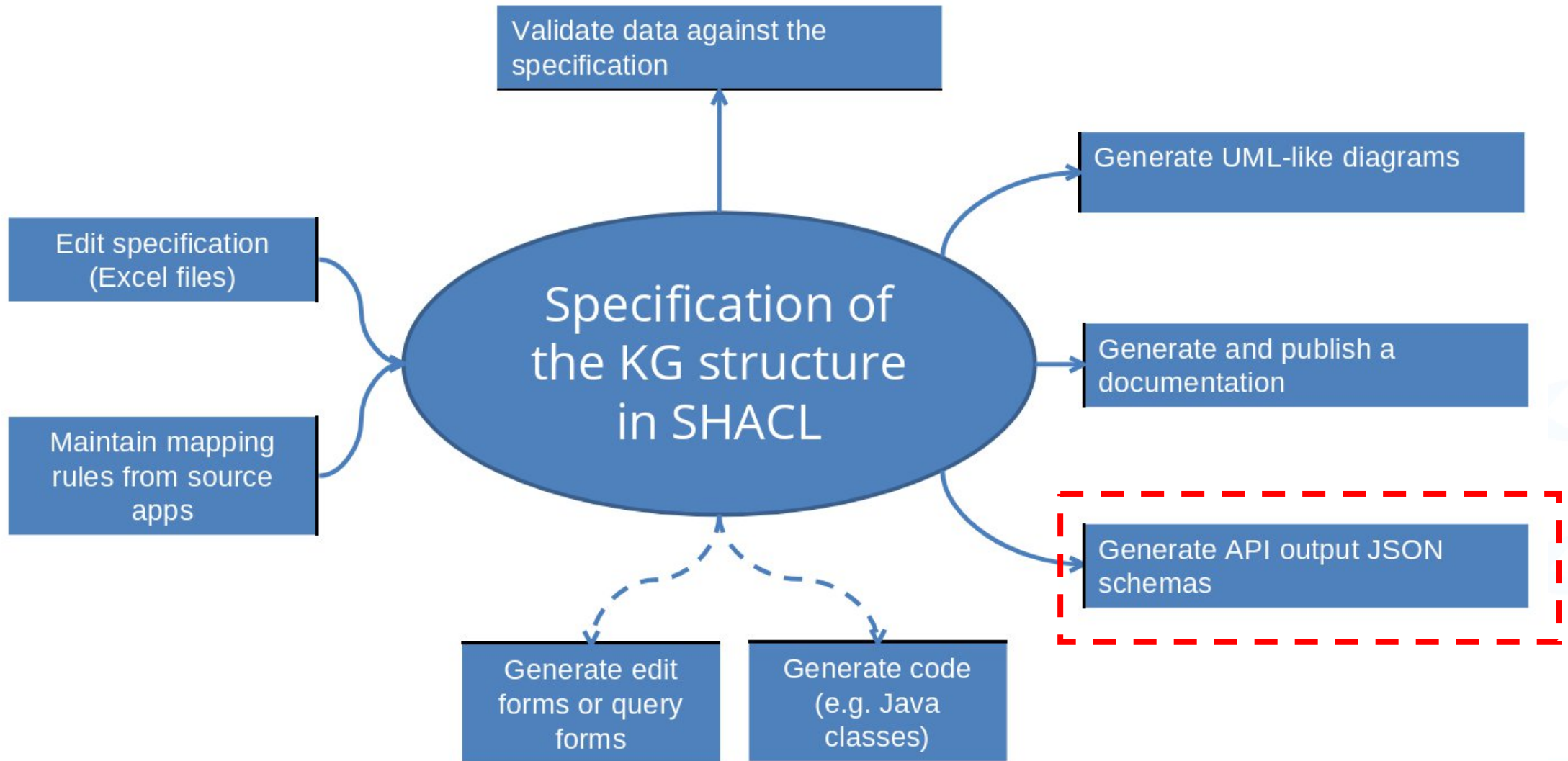
<http://data.europa.eu/eli/eli-draft-legislation-ontology#Activity>

- Nodes: IRI
- Closed shape
- URI pattern: `^https://data.europarl.europa.eu/eli/dl/event/[A-Za-z0-9\-\_]+\$`
- Example: `https://data.europarl.europa.eu/eli/dl/event/MTG-PL-2020-10-23`

Property name	URI	Expected value	Card.	Description
activity date	<a href="#">eli-dl:activity_date</a>	xsd:date	1..1	The date of an activity, to indicate a single date. Alternatively, if an activity spans a time range, <code>"activity_start_date"</code> and <code>"activity_end_date"</code> can be used to indicate the beginning and end of the activity.
activity end date	<a href="#">eli-dl:activity_end_date</a>	xsd:dateTime	1..1	The date at which an activity ended.
activity id	<a href="#">eli-dl:activity_id</a>	xsd:string	0..1	An identifier for the process. This is meant to capture how the process is identified by law makers. A process can have multiple identifiers for different institutions, e.g. an identifier for the government, and an identifier for the parliament. Note that a subproperty <code>process_external_id</code> exists to capture the identifier of another institution, depending on the point of view.
activity label	<a href="#">eli-dl:activity_label</a>	rdf:langString	0..*	A label on the activity, describing what is being carried out.
activity start date	<a href="#">eli-dl:activity_start_date</a>	xsd:dateTime	1..1	The date at which an activity started.
activity type	<a href="#">eli-dl:had_activity_type</a>	eli-dl:Activitytype	1..1	The type of a process. The type of a process often determine the workflow being followed to create the law.
based on a realization of	<a href="#">eli-dl:based_on_a_realization_of</a>	Work	0..*	Indicates that an Activity used some Work .The works are seen as the "input" of the activity. Example : the work of a committee "used" a specific version of the draft as input to produce an opinion document. Note : the naming "used a realization of" is taken from FRBRoo. Since a work is an idea, existing only in people's mind, an activity cannot be simply



## 6. SHACL as the foundation for the KG





# La spécification SHACL contrôle la **structure JSON de résultat des APIs.**

*Le schema JSON est aligné sur la structure de la spécification SHACL.*

*(Conversion SHACL vers JSON schema + context JSON-LD, avec ajout d'annotations pour savoir « quand imbriquer » dans le JSON)*

## 6. SHACL as the foundation for the KG

```
{
  "id": "eli/dl/proc/2022-0269",
  "type": "Process",
  "consists_of": [
    {
      "id": "eli/dl/event/2022-0269-ANPRO-2022-10-06",
      "type": "Activity",
      "activity_date": "2022-10-06",
      "activity_id": "2022-0269-ANPRO-2022-10-06",
      "had_activity_type": "def/ep-activities/REFERRAL"
    },
    {
      "id": "eli/dl/event/MTG-PL-2024-04-22-PVCRE-ITM-19",
      "type": "Activity",
      "activity_date": "2024-04-22",
      "activity_id": "MTG-PL-2024-04-22-PVCRE-ITM-19",
      "had_activity_type": "def/ep-activities/PLENARY_ACTIVITY",
      "recorded_in_a_realization_of": [
        "eli/dl/doc/PV-9-2024-04-22-ITM-19"
      ]
    }
  ]
}
```

### Procedure

This node shape refers to the class `eli-dl:Process`

- Applies to: `eli-dl:Process`
- Nodes: IRI
- Closed shape
- URI pattern: `^https://data.europarl.europa.eu/eli/dl/proc/[A-Za-z0-9\-\_]+$`
- Example: `https://data.europarl.europa.eu/eli/dl/proc/2022-0147`

Property name	URI	Expected value	Card.	Description
consists of	<code>eli-dl:consists_of</code>	IRI	0..*	
created a realization of	<code>eli-dl:created_a_realization_of</code>	Reference to a Work	0..*	
current stage	<code>eli-dl:current_stage</code>	Reference to a Process stage	1..1	
identifier year	<code>ept:epvoc#identifierYear</code>	xsd:string	1..1	

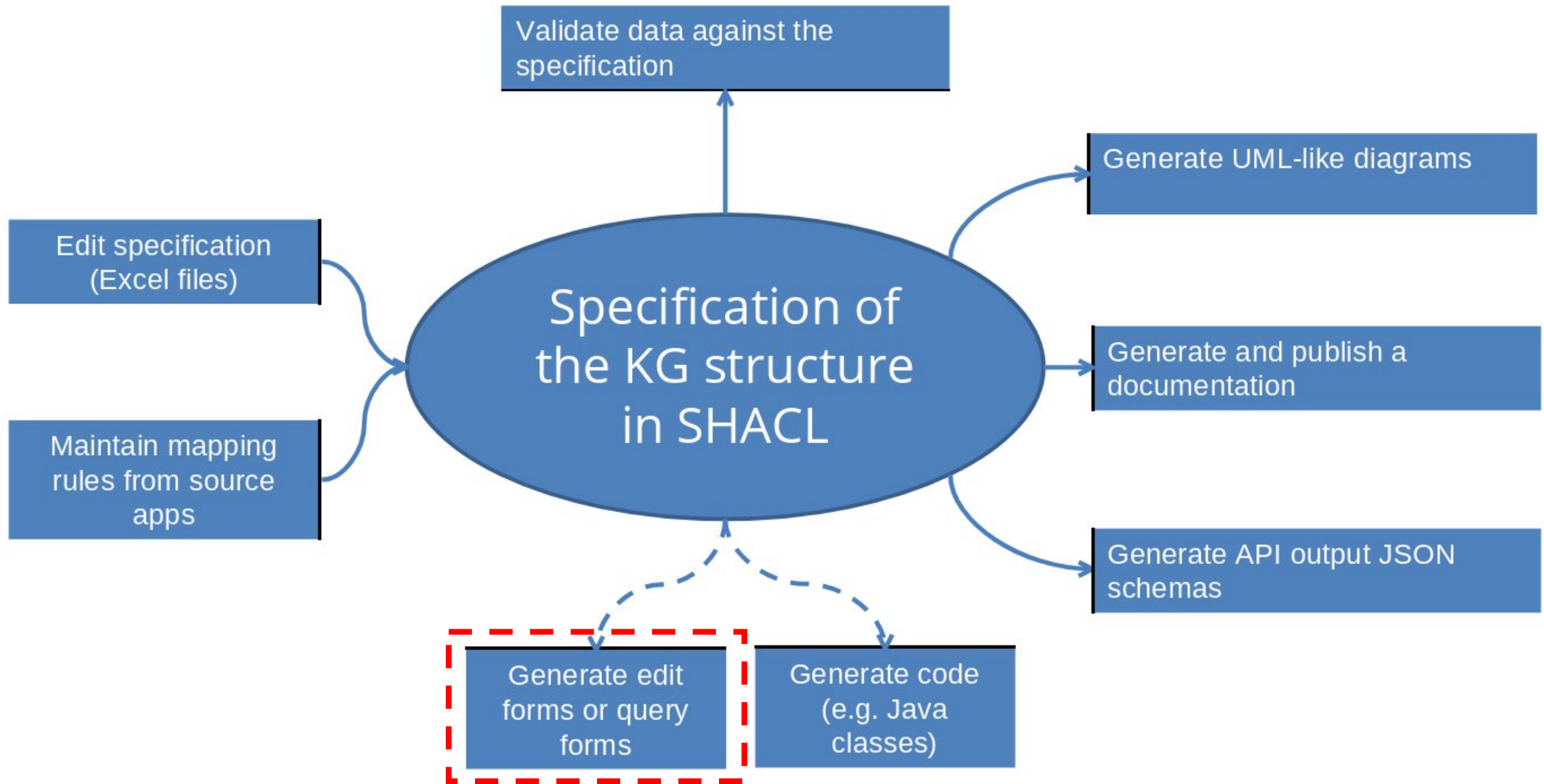
### Activity

This node shape refers to the class `eli-dl:Activity`

- Nodes: IRI
- Closed shape
- URI pattern: `^https://data.europarl.europa.eu/eli/dl/event/[A-Za-z0-9\-\_]+$`
- Example: `https://data.europarl.europa.eu/eli/dl/event/2023-2121-ANPRO-2023-07-12`

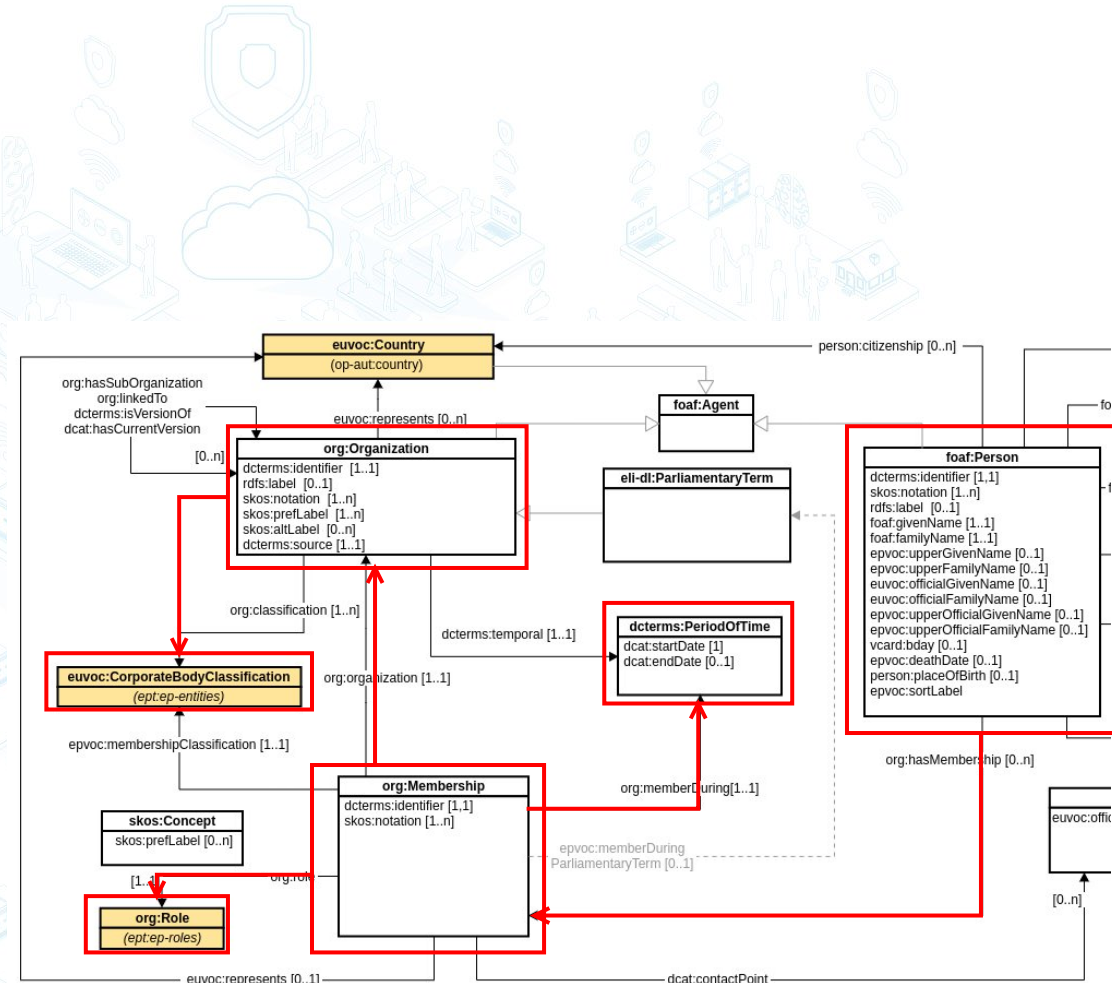
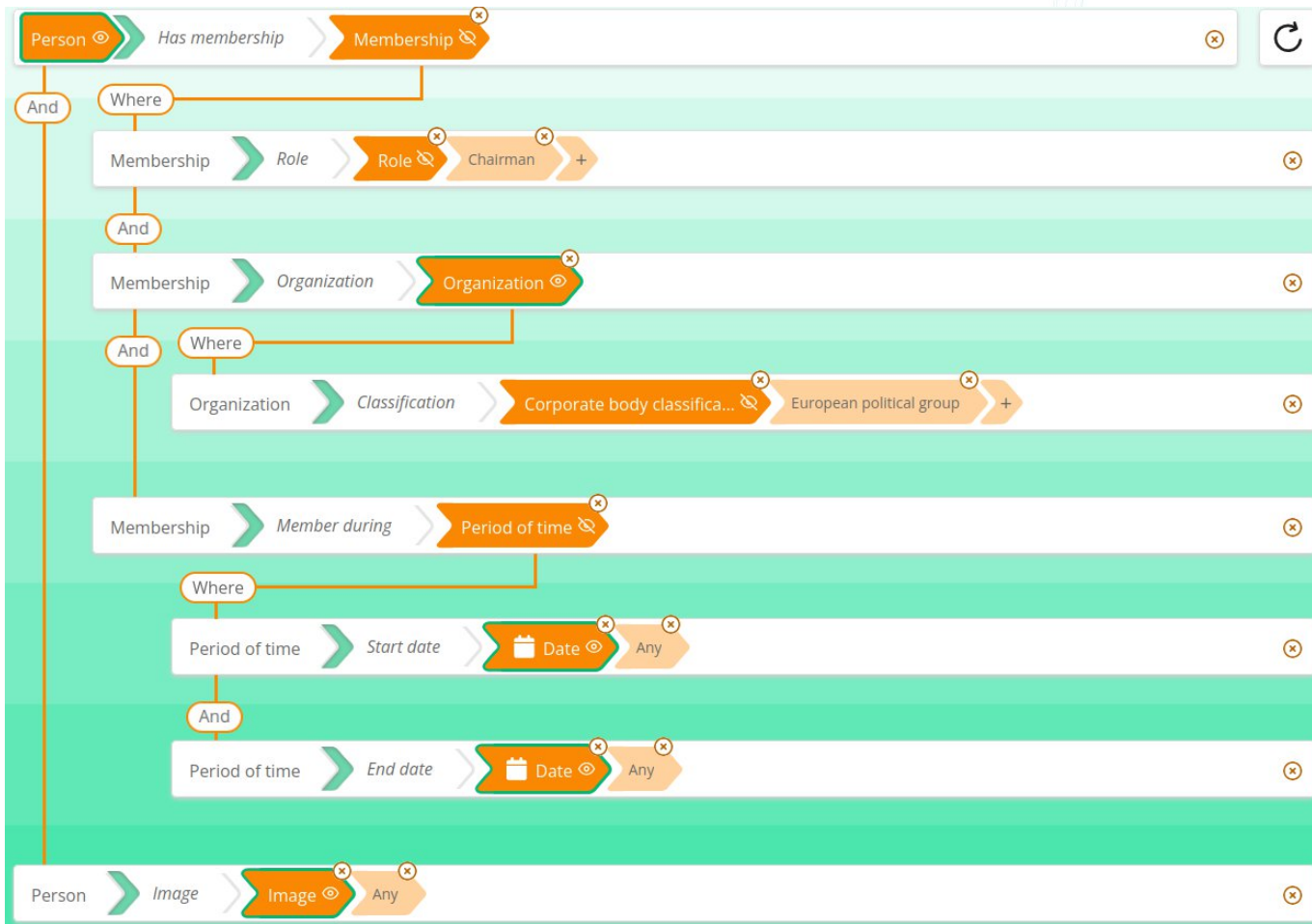
Property name	URI	Expected value	Card.	Description
activity date	<code>eli-dl:activity_date</code>	xsd:date	1..1	
activity id	<code>eli-dl:activity_id</code>	xsd:string	1..1	
activity type	<code>eli-dl:had_activity_type</code>	Reference to an Activity type <code>ept:ep-activities/REFERRAL, ept:ep-activities/TABLING_PLENARY, ept:ep-activities/PUBLICATION_OFFICIAL_JOURNAL, ept:ep-activities/SIGNATURE, ept:ep-activities/PLENARY_DEBATE, ept:ep-activities/PLENARY_VOTE, ept:ep-activities/PLENARY_ACTIVITY</code>	1..1	

## 6. SHACL as the foundation for the KG





## 6. SHACL as the foundation for the KG



Editeur de requêtes [Sparnatural](#) configuré automatiquement à partir de la spécification SHACL ORG-EP + ELI-EP : « tous les présidents d'un groupe politique du parlement, avec la période de leur présidence, et leur photo »

## 6. SHACL as the foundation for the KG

**Person**

**Jan ZAHRADIL**

- Has membership :
  - Organization : **ECR** (Organization)
- Member during :
  - Start date : 2011-03-08
  - End date : 2011-12-12

**Person**

**Syed KAMALL**

- Has membership :
  - Organization : **ECR** (Organization)
- Member during :
  - Start date : 2014-07-01
  - End date : 2017-07-04

**Person**

**Ryszard Antoni LEGUTKO**

- Has membership :
  - Organization : **ECR** (Organization)
- Member during :
  - Start date : 2022-10-13
  - End date : 2023-02-13

**Person**

**Gabriele ZIMMER**

- Has membership :
  - Organization : **GUE/NGL** (Organization)
- Member during :
  - Start date : 2012-03-14
  - End date : 2014-06-30
- Has membership :
  - Organization : **GUE/NGL** (Organization)
- Member during :

**Person**

**Manfred WEBER**

- Has membership :
  - Organization : **PPE** (Organization)
- Member during :
  - Start date : 2019-07-02
  - End date : 2024-07-15

**Person**

**Gianni PITTELLA**

- Has membership :
  - Organization : **S&D** (Organization)
- Member during :
  - Start date : 2014-07-01
  - End date : 2018-03-19

**Person**

**Manfred WEBER**

- Has membership :
  - Organization : **PPE** (Organization)
- Member during :
  - Start date : 2014-07-01
  - End date : 2019-07-01

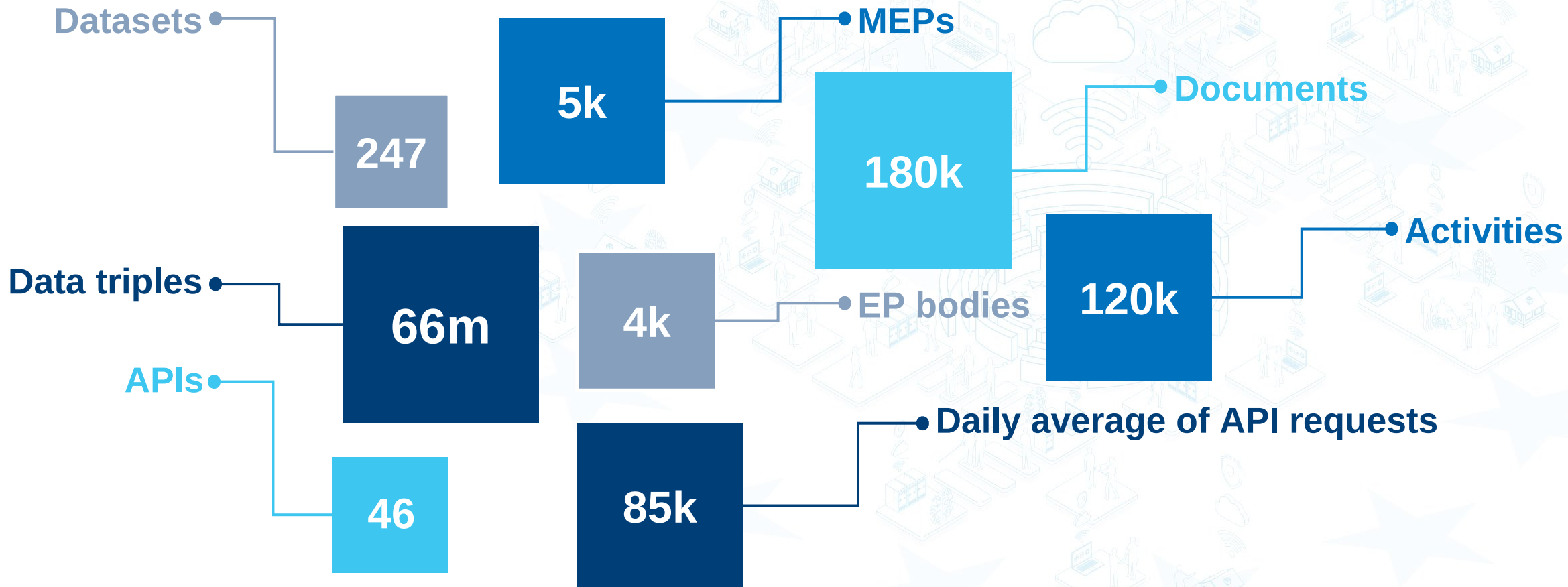
*Affichage du résultat de la requête précédente : la structure de la requête est reprise dans la structure des encadrés de résultats.*



# Thank you!



# Interesting figures (2024)



# EP Open Data in practice (1/2)

## Minutes - Results of votes

183k 158k 49k

Wednesday, 13 March 2024 - Strasbourg

[+ Abbreviations and symbols](#)

1. European Media Freedom Act \*\*\*I

Minutes - Item 8.1

Report: Sabine Verheyen (A9-0264/2023)

Subject	Am No	Author	Remarks	Vote	In favour, against, abstentions
Request to vote on the amendments instead of the provisional agreement					
Request to vote on the amendments to the draft legislative act		The Left	RCV	-	56, 535, 18
Provisional agreement					
Provisional agreement	332	committee	RCV	+	464, 92, 65
Draft legislative act					
Text as a whole	332CP1	committee		↓	
Article 4, § 2, point c	334D	The Left	RCV	↓	
	332CP2	committee		↓	
After Article 4	335	The Left	RCV	↓	
After recital 16	333	The Left	RCV	↓	
Commission proposal			RCV	↓	

Requests for roll-call votes

The Left: amendments 333, 334D, 335; request to vote on the amendments to the draft legislative act

## 4. Implementation insights

# EP open data in practice (2/2)

# EU Law Tracker

(Publication Office of the EU)

 <https://law-tracker.europa.eu/>

2022/0406(COD)

Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on multiple-vote share structures in companies that seek the admission to trading of their shares on an SME growth market

Stage

✓ Proposal

✓ First reading

End of procedure

Responsible institutions and bodies

European Parliament

✓ Council of the European Union

✓ European Commission

Latest events

Full timeline

Last event

08/10/2024

Approval of the EP's first reading position by the Council (adoption of the legislative act)

Previous step

24/04/2024

EP position at first reading

Stage: First reading

Documents: P9\_TA(2024)0352 (Texts adopted)

ITEC

Directorate-General  
for Innovation  
and Technological Support