Towards an event-centric knowledge graph approach for public administration

1st Dimitris Zeginis
University of Macedonia
Thessaloniki, Greece
zeginis@uom.edu.gr

2nd Konstantinos Tarabanis
University of Macedonia and
Information Technologies Institute
Centre for Research & Technology – Hellas
Thessaloniki, Greece
kat@uom.edu.gr

***Abstract*—Public administrations (PA) around the globe produce and handle a vast amount of data that are mainly the outcome of interactions of end-users. By evaluating the focus of PA one finds that most interactions involve only a few core entities such as the citizen or business. Usually, this information involving the core entities are scattered in numerous siloed databases developed by different departments and divisions, thus hindering PA to provide a comprehensive overview of their core entities and their interactions. Recently, knowledge graphs have been proposed for structuring large collections of data in a meaningful way, however they tend to represent a static state of the world and do not focus on the dynamics and changes over time. To address this, a new approach of event-centric knowledge graphs has been introduced that captures the dynamics of knowledge considering events as first-class entities for knowledge representation. The aim of this paper is to apply an event-centric knowledge graph approach for a holistic data governance of all data repositories in PA which models all interactions of PA related actors. We anticipate that the proposed approach will also facilitate PAs to adopt a data-centic orientation that can facilitate ubiquitous AI and data analytics.**

***Index Terms*—Semantic web, Knowledge graphs, event centric, public administration, life event, business event, CPSV-AP**

I. Introduction and motivation

Public administrations (PA) around the globe produce and handle a vast amount of data that are mainly the outcome of interactions of end-users. These interactions are usually related to the consumption of public services by citizens or businesses [1]. For example, a citizen uses a public service in order to register a birth at the responsible public authority or to register a new car, while a business uses a public service for regular/annual VAT declarations. All these interactions with public services produce data that are scattered in numerous siloed databases developed by different departments and divisions. The prevalent way of structuring these databases is based on the area of jurisdiction of the PA divisions. For example, different databases exist for storing general information about citizens (e.g. birth date, address), births and car registrations, each falling under the jurisdiction of different PA departments or divisions. By evaluating the focus of PA one finds that most interactions involve only a few core entities such as the citizen or business. However, the current siloed and fragmented way of structuring the data hinders PAs to provide a comprehensive overview of the core entities considering their interactions.

Knowledge graphs have gained increasing popularity in the past couple of years since they are able to structure large collections of data in a meaningful way. Typically, knowledge graphs represent a static state of the world including facts about persons or organizations (e.g. birth date, birth place, education, occupation) and do not focus on the dynamics and the changes over time (e.g. changes in occupation over time). Recently, a new approach of event-centric knowledge graphs [2] has been proposed to capture this dynamic knowledge, which requires considering "events" as first-class entities for knowledge representation.

From a PA perspective, events include interactions of citizens/businesses with the PA, mainly through the execution of public services. These events can be organized to broader categories called life-events (e.g. buy a new car) or business events (e.g. start a business) for citizens and businesses respectively. For example, the life event "buy a new car" requires a set of interactions with public services including the registration of the car, paying vehicle related taxes etc. This categorization of events offers a dimension to index events, which enables a more fine grained data exploration and analysis.

In reality however, public services have different "versions"/"variants" based on citizens/businesses profiles and circumstances, e.g. the service "registration of new car" has many versions/variants such as the "registration of a new private car", "registration of a new business car", "registration of a new heavy duty car" etc. which leads to different supporting documents, different costs and different outputs in each case. Thus, in order to represent the events it is crucial to pinpoint the specific public service version/variant (i.e. the specific supporting documents among the multiple valid options) that is to be consumed.

PAs can be benefited by adopting such a event-centric knowledge graph (ECKG) approach since it provides a comprehensive overview of citizens/businesses thus allowing PAs to adopt a citizen-friendly approach. But also, can facilitate ubiquitous data analytics and AI by enabling the integration of data not just in representation, but also in context and time.

At the core of current research in AI are Graph Neural Networks (GNN), a deep learning based method designed to gain information out of graphs by combining specific nodes and graph structural information. GNN recently gained increasing popularity in various domains, including knowledge graphs (KGs) [3]. Thus, the use of GNNs in event-centric knowledge graphs is very promising and offers a great potential.

The aim of this paper is to apply an ECKG approach for holistic data governance of all data repositories in PA in order to model all interactions of PA related actors. Towards this direction, the paper analyzes existing PA models and ontologies in order to: i) identify the main core entities and their interactions with PA, ii) identify concepts for public service versions/variants and iii) define an abstract ECKG model for PA. On-wards the paper applies these finding for a specific use case scenario and identifies potential problems that can be solved by using AI data analytics (e.g. GNNs). Thus the research questions addressed in this paper can be summarized as follows:

- How can ECKGs be applied at the PA?
- Which are the events and their variations at the PA?
- Which are the main core entities that participate at PA events?
- How to model core entities and events in the PA?

The rest of the paper is organized as follows, section II presents background information and related work, section III describes the proposed ECKG approach for PA, section IV presents a use case that applies the proposed approach and finally section V contains concluding remarks.

II. Background and Related Work

A. Event-centric knowledge graphs

Recently Knowledge Graphs (KGs) have been proposed to structure large collections of data in a meaningful way in the form of interrelated facts about entities. KGs rely on semantic web technologies such as RDF [4] (a model for data interchange on the Web) and SPARQL [5] (a language to express queries across diverse RDF data sources). These technologies use on URIs to name things and their relationships that are expressed as "triples" $< subject, predicate, object >$.

Existing popular open KGs include DBpedia [6], YAGO [7] and Wikidata [8] that contain facts about persons (e.g. name, birth date) or organizations (e.g. foundation date, owner). KGs are also broadly used to enhance the results provided by popular search engines such as Google [9] and Bing [10]. However, existing KG approaches focus on static facts and do not capture the dynamics of data [2] e.g. everyday interactions of citizens with the PA. From a data modeling perspective one can say that KGs represent organizational and master data, but do not handle transactional data. Information is centered around entities (e.g. person), thus the subject and object of RDF triples are often entities, and any information about interactions is captured through the predicate. Recently a new approach for event-centric knowledge graphs (ECKG) has been proposed that can capture the dynamics of data. For ECKGs, the subject of the triples is typically the event related to entities (e.g. place, actor) and is bound to time.

A set of models and approaches have already been proposed in the literature in order to represent events and their context. The Simple Event Model (SEM) [11] enables the representation of events (sem:Event) including the temporal (sem:Time) and spatial (sem:Place) dimension as well as the actors (sem:Actor) i.e. entities participating in the event. SEM has been applied in diverse domains including historic events [11] [12], events that appear in the news [2] and tourism related events [13]. Additionally, a multi-domain open event knowledge graph [14], that uses SEM, has been developed by integrating multiple ECKGs. A similar model to SEM is proposed by Scherp et. al [15], this model can represent time, space, objects and persons related to events, as well as relationships between events. Additionally, the Core Public Event Vocabulary (CPEV) [16] is proposed to model public events (e.g. conferences). The notion of public event is different from the notion of events used by the previous models. However, the core of CPEV is highly similar to them enabling the representation of time and space of events.

The increased popularity of ECKGs has enabled the development of applications, e.g. a question answering systems [17], as well the use of machine learning and deep learning techniques on top of them e.g. for script event prediction [18] and for a prediction system for COVID-19 [19].

Existing ECKGs use a graph-like structure to model knowledge. Although this is very flexible, it lacks performance when used at large scale. Additionally, organizations, including the PA, handle interactions (i.e. events) with only a few core entities (e.g. citizens, businesses). Thus, modeling the events based on these core entities is beneficial from a performance point of view and also enables organizations to provide a comprehensive overview of their core entities. Along this direction [20] has proposed the representation of entities and events as hierarchical trees instead of graphs.

B. Public services and Life events

From a PA perspective, events include interactions of citizens/businesses with PA through the execution of public services. The development of public services is based on underlying models and standards that are considered as main enablers for promoting interoperability and quality of public service provision. [21].

Embracing the need for a standard public service model, the European Commission has recently developed a public service model, termed Core Public Service Vocabulary (CPSV) [22]. CPSV is a simplified reusable and extendable data model, describing the characteristics of a public service in a structured and machine-readable way. In order to employ the CPSV model, an application profile was developed (CPSV-AP) that allows PAs to provide user-oriented information, grouped in life or business events. It consists of classes and attributes (mandatory or optional) defining a public service, its inputs, outputs costs, evidences and requirements. The current paper adopts the key points of the CPSV-AP v2.2.1 [23].

An important issue regarding the provision and execution of public services is their "versioning"/variants [24]. For example, the citizens' need to acquire a new car may require the fulfilment of different versions/variants of public services, according to different citizens' profiles, for example EU citizen vs. non-EU citizen. The profile of a citizen or other legal conditions define the mapping of a citizen's needs to a specific public service version/variant. For different versions/variants of a public service, citizens may have to submit different documents or the cost for the execution of the service may vary [25].

Finally, the literature has already identified "life events" and "business events" as situations of humans/businesses raising specific needs (e.g. to study, to start a professional activity, to get married, to have a baby, to travel) that require the execution of a group of public services in order to fulfil these needs [26]. For example, "having a baby" is a life event that requires the execution of a group of public services, e.g. "birth registration", "obtaining child allowance", etc. Thus, life events and business events can be considered also as groupings of ECKG events for PA.

C. Graph Neural Networks

Graph Neural Networks build a neural network based on the topology of the data graph i.e., nodes are connected to their neighbours per the data graph. GNNs can be used for diverse graph learning tasks [3] including node classification, node clustering, edge classification, link prediction, graph classification and graph matching. GNNs have already been applied in various problems including recommender systems, combinatorial optimization etc.

The use of GNNs on top of KGs is very promising since the structure of KGs form a graph thus GNNs can be directly applied out-of-the-box [27]. Regarding ECKGs, GNNs can be applied in order to predict events that may occur, as well as their their details.

III. An ECKG approach for the public administration

This section presents an ECKG approach for PA that is based on the tree-like structure proposed by [20] for the representation of events. For this purpose, section III-A presents an analysis of existing PA models and ontologies in order to identify the core entities and the types of events that need to be modeled in the context of PA. While section III-B proposes an abstract model for the representation of events as hierarchical trees in PA.

A. Identifying the core entities, event types and public service variants

The first step towards the identification of the core entities is the definition of the event types (i.e. types of interactions) that occur in PA. According to Capgemini [38] three main types of interactions occur in public administration:

- The citizen fulfils financial obligations to the PA e.g. the payment of income taxes, municipality taxes, road taxes health and social insurances.
- The citizen interacts with PA for non-financial reasons to e.g issue a passport, register a birth, register a new car.
- PA fulfils financial obligations of the citizen e.g. payments of social security benefits, unemployment benefits and pensions.

All the above three types of interactions occur through the execution of financial and non-financial public services. In this paper we consider PA events as the interactions of the core entities with a public service. For example, the payment of income taxes is done through the interaction of citizens with a financial public service, the registration of a birth is done through the interaction of a citizen with a non-financial public service, and the payment of a pension is done through the interaction of a PA entity with a financial public service. The entities involved in the interactions with the public services are the core entities of the hierarchical tree structure.

In order to identify all these core entities we conducted an analysis of existing PA and public service models and ontologies. This analysis included models and ontologies proposed in the academic literature (e.g. [28], [30]) as well as national models and standards applied in real-life settings (e.g. [31], [23]). The identification of relevant models and ontologies was based on existing reports and literature review paper. Specifically, national PA and public service models are reported in a study published by the European Commission [39] while academic ontologies and models are reviews in [1]. Table I presents the outcome of this analysis. Three main core entities have been identified namely: "citizen", "business" and "public administration". From an ECKG point of view, these core entities are the actors that participate in an event.

Another interesting outcome of the analysis is that events usually occur under the scope of broader categories called "life-events" and "business-events". This broader category can be used as a high-level event containing more sub-events.

Since PA events are considered as the interactions of the core entities with a public service, it is crucial to associate the event with the specific public service (and where applicable to the public service version/variant) that is consumed. The CPSV-AP model enables the description of the characteristics of a public service, however without distinguishing the variations introduced by the public service versions/variants. Specifically, in CPSV-AP models all variations are under the same public service definition, and thus, it is the role of the ECKG events to pinpoint the exact version/variant that is actually consumed.

For this purpose, it is fundamental to identify the CPSV-AP concepts that may vary amongst public service versions/variants as well as those that remain the same. Table II provides the name and description of the most important CPSV-AP concepts as well as their categorization as event aware (i.e. may vary amongst public service versions) or event agnostic (i.e. do not vary amongst public service versions) including a justification. In a nutshell, the event agnostic concepts include "Public Service", "Public Organization", "Legal Resource" and "Contact Point", while the event aware concepts include "Output", "Evidence", "Channel", "Cost" and "Criterion Requirement". This categorization of concepts as event aware and event agnostic is not absolute and may vary for different public

TABLE I: Core entities identified from existing public administration and public service models and ontologies

Public Administration model	Citizen	Business	Public Administration	Life/Business event
Governmental Markup Language (GovML) [28]	✗	✗	✗	✓
SmartGov model [29]	✓	✓	✗	✗
E-GOV Public Services Ontology (E-GOV PSO) [30]	✓	✓	✗	✓
Switzerland Data Model for Public Administration (DMPA) [31]	✓	✓	✓	✗
Governance Enterprise Architecture (GEA)	✓	✓	✗	✗
DIP model [32]	✓	✗	✗	✓
OneStopGov model [33]	✓	✗	✗	✓
Access-eGov model [34]	✗	✗	✗	✓
Government to Businesses Model (G2BM) [35]	✗	✓	✗	✗
eGovernment Knowledge Interoperability Ontology (eGKI) [36]	✓	✓	✓	✗
Life Event Ontology (LEO) [37]	✓	✗	✗	✓
Core Public Service Vocabulary (CPSV) [22], [23]	✗	✗	✓	✓

service implementations. For example, different "Public Organizations" may be responsible for delivering different public service versions/variants, in this case "Public Organization" would be categorized as an event aware concept. However, any changes to the categorization do not affect the basic approach and structure of the model proposed in section III-B.

B. An abstract ECKG model for PA

This section presents an abstract ECKG tree model for PA (figure 1) that can be used in order to populate event-centric graphs. The structure of the model is based on the outcomes of the analysis of existing PA models and ontologies (section III-A) and includes four main levels:

- Level 1: is the core entity that serves as the root of the tree. For PA the core entity can be either a citizen, or a business or a PA entity.
- Level 2: includes the high-level events such as "life-event" or "business-event". These high-level events contain more sub-events of level 3. This level is optional since some events may not be associated with a life-event.
- Level 3: includes the fine-grained events e.g. birth registration. These events are related to the interaction of core entities with a public service and include connections with the "event aware" concepts of CPSV-AP in order to distinguish the specific version/variant of the service that is consumed.
- Level 4: includes public service descriptions based on the CPSV-AP model. These descriptions form a knowledge base of public services that contains all the information needed to add context to the ECKG. The public service descriptions include both the event aware and event agnostic concepts of CPSV-AP.

Events of level 3 are directly associated: i) with a "Public Service" using the predicate "event type" and ii) with all the event aware concepts of CPSV-AP (e.g. Channel,Output). This way of modeling enables the association of events with the specific public service version/variant (including "Costs", "Channel", "Output", "Evidence" etc) that is consumed. All the event agnostic concepts (e.g. Public Organization responsible for delivering the public service) can be retrieved directly through the Public Service descriptions since they do not vary amongst the diverse public service versions/variants. All the events (level 2 and level 3) have a start and end-time expressed through the "Time" concept.

The concepts of the proposed model can also be associated with other external knowledge bases (i.e. ontologies, taxonomies and thesauri). The associated knowledge bases are not part of the tree, but are crucial to the operation of the ECKG since they contain all the domain dependent data needed to make the data in the tree meaningful. For example, the proposed model can use existing PA knowledge bases, documented at the CPSV-AP specification [23], to retrieve information about life-events, business-events, channels, evidences, legal resources etc.

IV. USE CASE

This section presents a use case that applies the proposed model in a real-life scenario: *"Mary has given birth to two children, one on 2018-10-10 and the other on 2020-11-11. After each birth she had to interact with PA using two public services: i) to register the birth, ii) to obtain a child allowance. Both services are offered by the Registry Office. The interactions with the registry office in 2018 were done at the physical location of the registry, while all the other interactions are done through online services. The amount of allowance obtained in 2018 is 3000\$, while in 2020 it was reduced to 2500\$."*. This scenario can be represented as a tree-like ECKG (figure 2). Some interesting remarks are:

- "Mary", that is a citizen, is the root of the tree. In this way, a citizen-centric data representation is enabled.
- The life-event "Birth" is used in order to group a set of related public services (Birth Registration Service and Child Allowance Service).
- The sub-events in level 3 are associated only with the values of the CPSV-AP event aware concepts that was actually used during the specific interaction. For example, the "Birth Registration Service" has two potential channels "Physical office" and "Online service", while the event "Birth Registration 1" is associated only with one of them ("Physical office"), that which was actually used.
- The descriptions of public services (e.g. "Birth Registration Service" and "Child Allowance Service"), including all event aware and event agnostic concepts, rarely change

TABLE II: Description and Categorization of CPSV-AP concepts as event aware (may vary amongst public service versions/variants)) and event agnostic (do not vary amongst public service versions/variants)

Name	Description	Category	Justification
Public Service	Represents the Public Service itself, as it is described in a public service catalogue	Event agnostic	The main characteristics of the public service (e.g. sector, keyword, language) do not vary amongst different versions/-variants
Public Organization	The responsible Agent for the delivery of the Public Service	Event agnostic	The responsible agent for delivering the public service does not vary amongst different versions/variants
Legal Resource	The legislation, policy or policies that lie behind the Rules that govern the service	Event agnostic	The legislation that defines the Rules that govern the service does not vary amongst different versions/variants
Contact point	Represents the contact information for a Public Ser-vice	Event agnostic	The contact information for a public service does not vary amongst different versions
Output	Any resource - document, artefact – anything **pro-duced** by the Public Service	Event aware	The output of the public service (e.g. the amount of child allowance) may vary amongst different versions/variants
Evidence	Any resource - document, artefact – anything **needed** in order to execute the Public Service	Event aware	The evidence required as input for a public service (e.g. for adults or minors) may vary amongst different versions/vari-ants
Channel	Represents the medium through which an Agent interacts with a Public Service e.g. online services, phone, walk-in centres	Event aware	The channel may vary for different public service versions/-variants due to the different channels used by the citizens
Cost	Any costs related to the execution of a Public Service that the consuming Agent needs to pay	Event aware	The cost may vary amongst different public service version-s/variants (e.g. due to additional processing steps required for complex service versions/variants)
Criterion Requirement	Describes the criteria for needing or using the service	Event aware	The Criterion Requirement vary by default amongst different public service versions/variants i.e. different service version-s/variants are defined by different criterion requirements

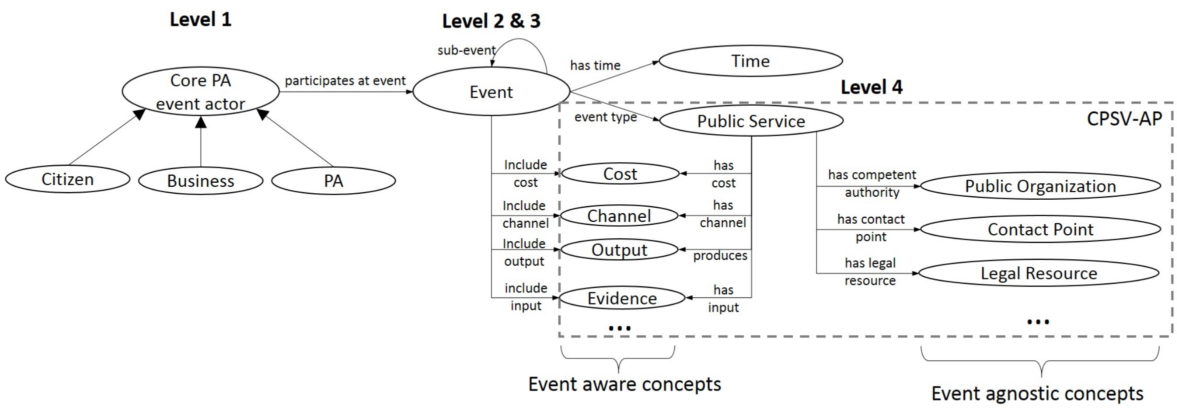

Fig. 1: An abstract ECKG model for PA

(e.g. when legislation changes) and act as a reference for the ECKG events.

- The "Registry Office" which is the public organization responsible for offering these two public services could be considered as a secondary-core entity of the tree.

In order to implement the scenario the paper uses the RD-F/TTL[1] representation. The first step for the implementation of the scenario is the mapping of classes/properties of the proposed ECKG model to existing models and vocabularies. The vocabularies used for the mapping are:

- The core vocabularies of the European Union[2] (pre-fix `cv:`) that define concepts related to people (`cv:Person`), businesses (`cv:LegalEntity`) and life events (`cv:LifeEvent`)
- The Core Public Service Vocabulary[3] (prefix `cpsv:`) of the European Union that defines concepts about public services (`cpsv:PublicService`) and their variations.
- The SEM vocabulary [11] (prefix `sem:`) that defines concepts about events (`sem:Event`, `sem:hasSubEvent`, `sem:hasTimeStamp`, `sem:eventType`)

In cases where no relevant concepts can be found from the existing vocabularies, the paper defines new concepts (prefix `ex:`). For example, the participation of a citizen in a life-event is modeled through the predicate `ex:faces`. Additionally, all relations of `sem:Event` to the event aware CPSV-AP concepts are expressed through the `ex:` prefix i.e. `ex:includeChannel` and `ex:includeOutput`.

[1]https://www.w3.org/TR/turtle/
[2]http://data.europa.eu/m8g
[3]https://ec.europa.eu/isa2/solutions/core-public-service-vocabulary-application-profile-cpsv-ap_en

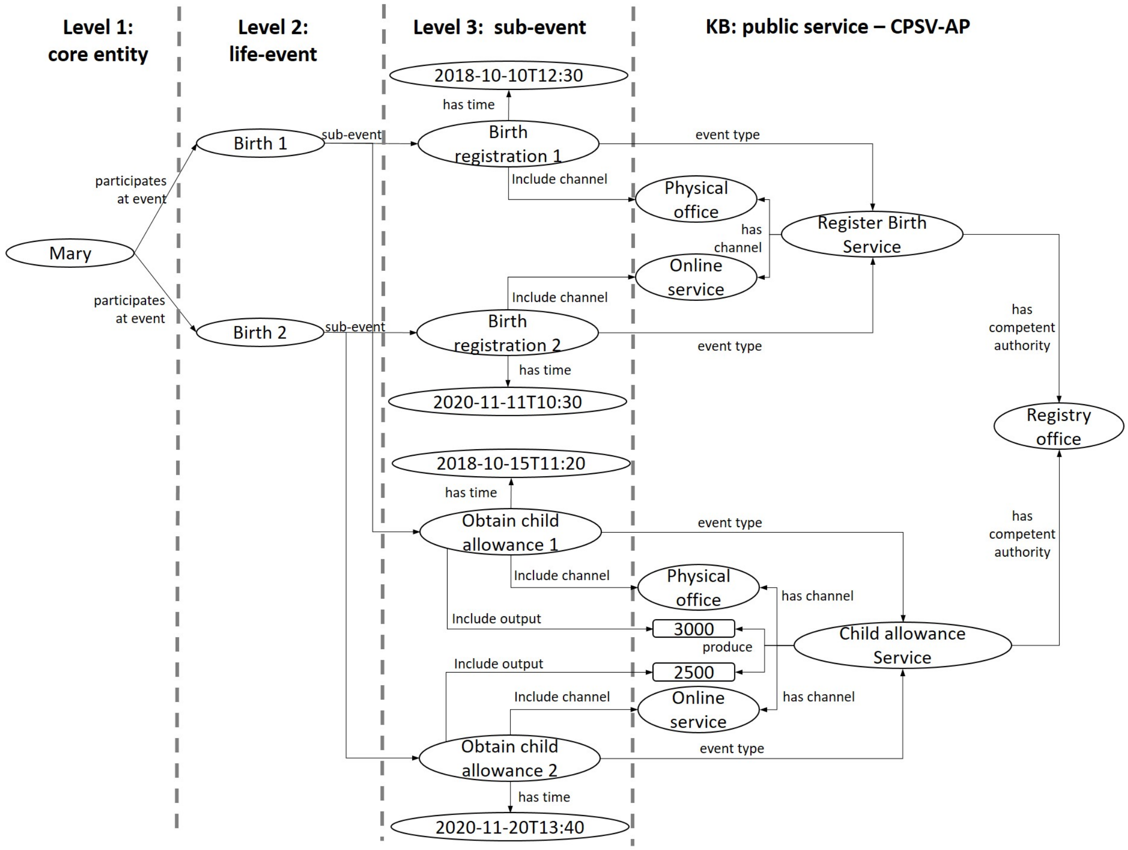

Fig. 2: Tree-like ECKG of the use case scenario

Listing 1: RDF/TTL code that represents the ECKG of the use case

```
ex:Mary a cv:Person;
    ex:faces ex:birth_1, ex:birth_2.
ex:birth_1 a cv:LifeEvent;
    dct:type ex:HavingAChild;
    sem:hasSubEvent ex:BirthRegistration_1,
                    ex:ChildAllowance_1.
ex:BirthRegistration_1 a sem:Event;
    sem:hasTimeStamp
        "2018-10-10T12:30:00"^^xsd:dateTime;
    sem:eventType ex:RegisterBirthPS;
    ex:includeChannel ex:PhysicalOffice.
ex:ChildAllowance_1 a sem:Event;
    sem:hasTimeStamp
        "2018-15-10T11:20:00"^^xsd:dateTime;
    sem:eventType ex:ChildAllowancePS;
    ex:includeChannel ex:PhysicalOffice;
    ex:includeOutput 3000.
ex:birth_2 a cv:LifeEvent;
    dct:type ex:HavingAChild;
    sem:hasSubEvent ex:BirthRegistration_2,
                    ex:ChildAllowance_2.
ex:BirthRegistration_2 a sem:Event;
    sem:hasTimeStamp
        "2020-11-11T10:30:00"^^xsd:dateTime;
    sem:eventType ex:RegisterBirthPS;
    ex:includeChannel ex:OnlineService.
ex:ChildAllowance_2 a sem:Event;
    sem:hasTimeStamp
        "2020-20-11T13:40:00"^^xsd:dateTime;
    sem:eventType ex:ChildAllowancePS;
    ex:includeChannel ex:OnlineService;
    ex:includeOutput 2500.
ex:RegisterBirthPS a cpsv:PublicService,
                     sem:EventType;
    cv:hasCompetentAuthority ex:RegistryOffice;
    cv:hasChannel ex:PhysicalOffice,
                  ex:OnlineService.
ex:ChildAllowancePS a cpsv:PublicService,
                      sem:EventType;
    cv:hasCompetentAuthority ex:RegistryOffice;
    cv:hasChannel ex:PhysicalOffice,
                  ex:OnlineService;
    cpsv:produces 3000, 2500.
ex:RegistryOffice a cv:PublicOrganisation.
ex:PhysicalOffice a cv:Channel.
ex:OnlineService a cv:Channel.
```

The RDF/TTL code in listing 1 represents the use case scenario. An interesting remark is that both the `cv:LifeEvent` and the `sem:Event` are associated with a type property (`dct:type` and `sem:eventType` respectively), which allows their linking to controlled vocabularies (i.e. controlled vocabularies about life-events and public services). For example, the statement `ex:birth_1 dct:type ex:HavingAChild` associates the life-event `ex:birth_1` to the type "Having a child". These controlled vocabularies contain domain specific information making the data in the tree meaningful.

Based on the RDF/TTL representation of the scenario a number of SPARQL queries can be executed providing a comprehensive overview of the core entity (i.e. Mary) and its interactions with PA as well as the secondary-core entity (i.e. Registry Office). Some example SPARQL queries include:

- Listing 2 presents a SPARQL query to retrieve the total amount of children allowances received by Mary.
- Listing 3 presents a SPARQL query to retrieve the number of interactions between Mary and the Registry Office (including all the public services offered by the Registry Office).
- Listing 4 presents a SPARQL query to retrieve the total number of citizen interactions with the Registry Office after 01/01/2020..

Other queries can also be expressed retrieving e.g. the number of online services consumed by Mary within a specific time range. These queries demonstrate the simplicity to retrieve information about core entities (e.g. citizen) and secondary-core entities (e.g. Public Organization). Such queries would be more complex in existing PA databases requiring information from multiple tables.

Listing 2: Total amount of children allowances received by Mary

```
PREFIX ex: <http://www.exampe.org/>
PREFIX sem: <http://semanticweb.cs.vu.nl/2009/11/sem/>

SELECT (SUM(?allowanceValue) AS ?total)
WHERE{ ex:Mary ex:faces ?lifeEvent.
       ?lifeEvent sem:hasSubEvent ?event.
       ?event sem:eventType ex:ChildAllowancePS.
       ?event  ex:includeOutput ?allowanceValue.}
```

Listing 3: Number of interactions between Mary and the Registry Office

```
PREFIX ex: <http://www.exampe.org/>
PREFIX sem: <http://semanticweb.cs.vu.nl/2009/11/sem/>
PREFIX cv: <http://data.europa.eu/m8g/>

select (COUNT(?event) as ?numInteractions)
WHERE {ex:Mary ex:faces ?lifeEvent.
       ?lifeEvent sem:hasSubEvent ?event.
       ?event sem:eventType  ?ps.
       ?ps cv:hasCompetentAuthority ex:RegistryOffice.}
```

Listing 4: Number of Registry Office service interactions after 01/01/2020

```
PREFIX ex: <http://www.exampe.org/>
PREFIX sem: <http://semanticweb.cs.vu.nl/2009/11/sem/>
PREFIX cv: <http://data.europa.eu/m8g/>

SELECT COUNT(?event) as ?numInteractions
WHERE{ ?event sem:eventType ?ps.
       ?ps cv:hasCompetentAuthority ex:RegistryOffice.
       ?event  sem:hasTimeStamp ?date.
       FILTER(?date >xsd:date("2020-01-01"))}
```

V. Conclusion and discussion

Data produced and handled by public administrations are currently scattered in numerous siloed databases hindering a comprehensive understanding of PA operation, including interactions amongst PA related actors (e.g. citizens, businesses). To address this issue, this paper proposes an event-centric knowledge graph approach for holistic data governance of all data repositories in PA, which models all interactions amongst PA related actors.

The proposed approach provides an easily understandable data model encompassing a uniform representation of all interactions in PA. This, among other things, facilitates the writing/execution of queries related to overall PA operation. Such queries would be more complex in existing siloed databases used by PAs. Additionally, it provides a comprehensive view of PA related actors through simple queries, and also enables event-level provenance by modeling events as first class entities.

A set of models and approaches have already been proposed in the literature in order to represent events and their context (e.g. [11], [15], [16]). However, these models are generic and do not capture the specific needs of the PA towards the creation of ECKGs. For example, they do not consider events (i.e. interactions with a public service) and their versions in the context of PA or do not define the core entities that interact with the PA. Towards this direction, the proposed model uses CPSV-AP to describe public services and their versions. Thus, the model relies on CPSV-AP to express any complexity relayed to public services e.g. the expression of hierarchies and internal structure of public organizations, contact points and legal resources.

The model is demonstrated through a use case example. As future work, the authors plan to validate the model: i) by using large synthetic datasets with PA events in order to evaluate the scalability of the model and the performance of queries and ii) through a human assessment evaluation (e.g. using questionnaires) by involving PA domain experts in order to evaluate the expressivity and completeness of the model.

The use case of the model presented in the paper is implemented using graph databases (i.e. RDF and SPARQL), however other implementation of the model are also possible e.g. relational databases. The use of graph databases enables reasoning over the data and allows the discovery of hidden information, thus answering complex queries. On the other hand relational databases are usually more efficient. A qualitative and quantitative comparison of different implementations of the models is also planed by the authors as future work.

In order to realize the ECKG approach at the PA a set of applications also need to be exploited including an application to manage/maintain the model and the potential external knowledge bases (e.g. about public services or public organizations), an application to create data conformant to the model and a database to store the ECKG. On top of the ECKG various added value graph-based analytics (e.g. GNN) can be applied, for example: i) the prediction of future interaction of citizens with PA, enabling resource planning e.g. reserving the appropriate amount of money to be given through allowances, ii) the prediction of the public service version/variant to be consumed in each interaction, enabling e.g. management of the channels offering the service. Other graph-based analytics tasks on top of ECKGs are also possible but their full potential still remains to be explored.

We anticipate that the proposed approach can foster ubiquitous AI and analytics by fusing scattered PA data, leading to the discovery of deep and explainable connections and patterns in PA interactions. As a whole, the ease of generating this knowledge by employing our model will benefit both PAs and society, which they serve.

REFERENCES

[1] A. Gerontas, V. Peristeras, E. Tambouris, E. Kaliva, I. Magnisalis, and K. Tarabanis, "Public service models: A systematic literature review and synthesis," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 2, pp. 637–648, 2021.

[2] M. Rospocher, M. van Erp, P. Vossen, A. Fokkens, I. Aldabe, G. Rigau, A. Soroa, T. Ploeger, and T. Bogaard, "Building event-centric knowledge graphs from news," *J. of Web Semantics*, vol. 37-38, pp. 132–151, 2016.

[3] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666651021000012

[4] R. Cyganiak, D. Wood, and M. Lanthaler, "Rdf 1.1 concepts and abstract syntax," W3C Recommendation, W3C, 2014, https://www.w3.org/TR/rdf11-concepts/.

[5] "Sparql 1.1 overview," W3C Recommendation, W3C, 2012, https://www.w3.org/TR/sparql11-overview/.

[6] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann, "Dbpedia - a crystallization point for the web of data," *Journal of Web Semantics*, vol. 7, no. 3, pp. 154–165, 2009, the Web of Data.

[7] F. M. Suchanek, G. Kasneci, and G. Weikum, "Yago: A core of semantic knowledge," in *Proceedings of the 16th International Conference on World Wide Web*, ser. WWW '07. New York, NY, USA: Association for Computing Machinery, 2007, p. 697–706.

[8] D. Vrandečić and M. Krötzsch, "Wikidata: a free collaborative knowledgebase," *Communic. of the ACM*, vol. 57, no. 10, pp. 78–85, 2014.

[9] A. Singhal, "Introducing the knowledge graph: things, not strings," Google blog, 2012, https://blog.google/products/search/introducing-knowledge-graph-things-not/.

[10] R. Qian, "Understand your world with bing," Microsoft blog, 2013, https://blogs.bing.com/search/2013/03/21/understand-your-world-with-bing/.

[11] W. R. van Hage, V. Malaisé, R. Segers, L. Hollink, and G. Schreiber, "Design and use of the simple event model (sem)," *Journal of Web Semantics*, vol. 9, no. 2, pp. 128–136, 2011.

[12] S. Gottschalk and E. Demidova, "Eventkg: A multilingual event-centric temporal knowledge graph," in *The Semantic Web*. Cham: Springer International Publishing, 2018, pp. 272–287.

[13] J. Wu, X. Zhu, C. Zhang, and Z. Hu, "Event-centric tourism knowledge graph—a case study of hainan," in *Knowledge Science, Engineering and Management*, G. Li, H. T. Shen, Y. Yuan, X. Wang, H. Liu, and X. Zhao, Eds. Cham: Springer International Publishing, 2020, pp. 3–15.

[14] S. Gottschalk, E. Kacupaj, S. Abdollahi, D. Alves, G. Amaral, E. Koutsiana, T. Kuculo, D. Major, C. Mello, G. S. Cheema *et al.*, "Oekg: The open event knowledge graph." in *CLEOPATRA@ WWW*, 2021, pp. 61–75.

[15] A. Scherp, T. Franz, C. Saathoff, and S. Staab, "F–a model of events based on the foundational ontology dolce+dns ultralight," in *Proceedings of the 5th Intern. Conference on Knowledge Capture*, ser. K-CAP '09. New York: Association for Computing Machinery, 2009, p. 137–144.

[16] DG-DIGIT/ISA-Programme, "D03.03.02 core public event vocabulary," European Commission, 2018.

[17] T. Souza Costa, S. Gottschalk, and E. Demidova, "Event-qa: A dataset for event-centric question answering over knowledge graphs," in *Proceedings of the 29th ACM International Conference on Information Knowledge Management*. New York, NY, USA: ACM, 2020, p. 3157–3164.

[18] Z. Li, X. Ding, and T. Liu, "Constructing narrative event evolutionary graph for script event prediction," *CoRR*, vol. abs/1805.05081, 2018. [Online]. Available: http://arxiv.org/abs/1805.05081

[19] X. Fu, X. Jiang, Y. Qi, M. Xu, Y. Song, J. Zhang, and X. Wu, "An event-centric prediction system for covid-19," in *2020 IEEE International Conference on Knowledge Graph (ICKG)*, 2020, pp. 195–202.

[20] J. Asman, "Entity event knowledge graphs for data centric organizations," Allegrograph, 2020, https://allegrograph.com/wp-content/uploads/2020/05/Entity-Event-Knowledge-Graphs-v5112020.pdf.

[21] A. Gerontas, V. Peristeras, E. Tambouris, E. Kaliva, I. Magnisalis, and K. Tarabanis, "Public service models: A systematic literature review and synthesis," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 2, pp. 637–648, 2021.

[22] DG-DIGIT/ISA-Programme, "D02.02 – definition and development of a data model for description of the services related to key business events, cpsv-ap, version 1.00," 2014.

[23] ——, "D02.02 – core public service vocabulary application profile 2.2.1," 2019.

[24] A. Stamatis, A. Gerontas, A. Dasyras, and E. Tambouris, "Using chatbots and life events to provide public service information," in *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance*, ser. ICEGOV 2020. New York, NY, USA: Association for Computing Machinery, 2020, p. 54–61. [Online]. Available: https://doi.org/10.1145/3428502.3428509

[25] A. Gerontas, D. Zeginis, E. Tambouris, and K. Tarabanis, "Towards personalised public service co-creation using core vocabularies," in *2021 6th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)*, 2021, pp. 1–7.

[26] E. Tambouris and K. Tarabanis, "Understanding and scoping life events," *International Journal of Electronic Governance*, vol. 1, no. 2, pp. 139–154, 2008. [Online]. Available: https://www.inderscienceonline.com/doi/abs/10.1504/IJEG.2008.017901

[27] Y. Gao, Y. Li, Y. Lin, H. Gao, and L. Khan, "Deep learning on knowledge graph for recommender system: A survey," *CoRR*, vol. abs/2004.00387, 2020. [Online]. Available: https://arxiv.org/abs/2004.00387

[28] G. Kavadias and E. Tambouris, "Govml: A markup language for describing public services and life events," in *Knowledge Management in Electronic Government*, M. A. Wimmer, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 106–115.

[29] N. Adams, J. Fraser, A. Macintosh, and A. McKay-Hubbard, "Towards an ontology for electronic transaction services," *Intelligent Systems in Accounting, Finance & Management*, vol. 11, no. 3, pp. 173–181, 2002.

[30] C. Vassilakis and G. Lepouras, "Ontology for e-government public services," in *Encyclopedia of E-commerce, E-government, and mobile commerce*. IGI Global, 2006, pp. 865–870.

[31] O. Glassey, "Developing a one-stop government data model," *Government Information Quarterly*, vol. 21, no. 2, pp. 156–169, 2004.

[32] V. Roberto, M. Rowlatt, R. Davies, A. Gugliotta, L. Cabral, and J. Domingue, "A semantic web service-based architecture for the interoperability of e-government services," in *Web Information Systems Modeling Workshop (WISM 2005) / 5th International Conference on Web Engineering (ICWE 2005)*, Sydney, Australia, July 2005.

[33] I. Trochidis, E. Tambouris, and K. Tarabanis, "An ontology for modeling life-events," in *IEEE International Conference on Services Computing (SCC 2007)*, 2007, pp. 719–720.

[34] R. Qian, "Deliverable 3.1: Access-egov platform architecture," The Access-eGov Project Consortium, 2006, available: https://intersoft.sk/accessegov/uploads/delivs_D3.1.pd.

[35] D. Barone, G. Viscusi, C. Batini, and P. Naggar, "A repository of services for the government to businesses relationship," in *Next Generation Information Technologies and Systems*, O. Etzion, T. Kuflik, and A. Motro, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 47–58.

[36] A.-M. Sourouni, F. Lampathaki, S. Mouzakitis, Y. Charalabidis, and D. Askounis, "Paving the way to egovernment transformation: Interoperability registry infrastructure development," in *Electronic Government*, M. A. Wimmer, H. J. Scholl, and E. Ferro, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 340–351.

[37] L. M. Álvarez Sabucedo, L. E. Anido Rifón, R. Míguez Pérez, and J. M. Santos Gago, "Providing standard-oriented data models and interfaces to egovernment services: A semantic-driven approach," *Computer Standards & Interfaces*, vol. 31, no. 5, pp. 1014–1027, 2009, specification, Standards and Information Management for Distributed Systems.

[38] Capgemini, "Rising to the new challenges of transactional services in the public sector," 2012.

[39] DG-DIGIT/ISA-Programme, "Definition and development of a data model for description of the services related to key business events," European Commission, 2014.