International Journal for Quality Research 13(1) 131–144
ISSN 1800-6450

Iñaki Salinas
Isaac Lera [1)]
Carlos Guerrero
Carlos Juiz
Enes Sukic

Article info:
Received 14.09.2018
Accepted 12.12.2018

UDC – 004.415
DOI – 10.24874/IJQR13.01-08

TOWARDS A SEMANTIC RESEARCH INFORMATION SYSTEM BASED ON RESEARCHERS' CV DOCUMENTS

Abstract: *Curricula vitae are widely used as the main mechanisms for evaluation of researchers. They also reflect a fraction of the data about the research activity of an organization and geographic area. The nature of the academic world causes researchers to have to continuously keep the information updated. There are tools to help manage this information, usually offered by an organization such as a university or the national government. However, the reports generated by these tools have a local scope when in most of the cases they are used outside the organization. We explore Semantic Web as a solution to this problem because of its benefits on interoperability, inference capabilities when merging data from different sources, and expedience when responding to semantic queries made by evaluators. We design a software solution using the existing tools, able to get data from a CV document, insert it into a semantic system, and then build a document CV from the existing amount of semantic data of a researcher in that system.*

Keywords: *Curriculum vitae, Semantic web, Research management, Linked data*

1. Introduction

Human capital is a key element for any research (Caire & Becker, 1967). Scientific method requires a combination of knowledge and talent in order to think appropriate questions, formulate hypothesis, develop and implement methodologies, understand observations, and share and discuss results with peers. However, human capital is difficult to measure, and thus, to evaluate. The usual evaluation mechanism is the Curriculum Vitae (CV): a well-structured text document with ordered lists of merits. For researchers, the CV contains abundant and varied types of evidence: education degrees, teaching experiences, faculty positions, participations in scientific meetings, staying in outer institutions, involvement on research projects, and its outcomes: publications, innovations, patents, awards, etc. This diversity hampers objective evaluations of CV candidates.

Universities and governments use the CV for evaluations. Even, with an enormous and ever-growing public source of information, as editorials or institutional web portals, research decision makers still ask for the CV when it comes to decisions. Nevertheless, CV documents are not a practical evaluation mechanism for all kind of decisions. They are for those, which affect individuals or groups, like funding projects, appointing positions or rewarding based on merits. But when it comes to planning or monitoring for the whole institution or nation, CV evaluation becomes

[1] Corresponding author: Isaac Lera
Email: isaac.lera@uib.es

a hard and complex task. Decision makers need integrated and summarized information for analysis, an overview of the situations and the relevant indicators. Current Research Information Systems (CRIS) are designed for these tasks. A CRIS is any information tool dedicated to provide access to and disseminate research information (Schöpfel et al., 2017). Most CRIS systems have been set up to serve national needs. In our opinion, according to the paradigm of Linked Data (LD) (Linked Data, 2017), CRIS might be a system with research information interconnected with the world, with profile management services for researchers, and the capability to provide interesting information to research decision makers. LD paradigm presents more benefits instead of dividing software systems according to the service they provide to researchers (Lorenzo-Escolar& Pastor-Ruiz, 2012; Shanks & Arlitsch, 2016). CRIS systems rarely include CV document import/export capabilities. They are either migrated from a database of the interested organization, or introduced by the researcher manually. The lack of normalized, standard-defined, CV formats hinders the import/export implementations on CRIS. Currently a European Union Recommendation to the member states, the Common European Research Information Format (CERIF), is being adopted quite widely and it encourages interoperation (Asserson, 2010). However, it does it on a broad scope, focusing on all possible data that the CRIS manages, including, but not limited to, CV data. It means CERIF does include CV data elements in its definition, but does not include any CV document format to import from or export to.

In Spain, there is a normalized standard CV for researchers, named Curriculum Vitae Normalizado (CVN) (Curriculum, 2017). It was defined as part of a national CRIS, designed to store CV data in databases of universities, share that data with the national CRIS, and compose printable documents with it. It simplifies the interaction with the administration, for example, in submissions of projects to national calls. CVN is also an example of a vocabulary definition for national interoperability: it fulfilled the integration of various previous standards for CV and research activity reporting in Spain (Lacunza & Arellano, 2011) and started a process to achieve European interoperability; expressing CVN to fulfil CERIF (Simons, 2013), enabling the importation of all CV generated as CVN in a CRIS that is CERIF compliant, like euroCRIS.

In our research, we analyse how could a university -or any organization that manages research-leverage their curriculum data of researchers with the Semantic Web technologies. Both research decision makers and researchers should benefit from it. We take the advanced degree of standardization and interoperability of Spanish CVN to demonstrate the potential of Semantic Web technologies. In addition, we use CV data from the University of Balearic Islands (UIB) where a previous related work has been done to migrate its data to semantic format (Lera et al., 2014). We propose a set of tools that would allow the university to achieve interoperability with the Semantic Web, and revalorize the CV of the researchers as evaluation mechanisms. For simplicity issues, our approach is called uCV that stands for university curriculum vitae.

The main contribution of this approach is the establishment and analysis of essentials services that a CRIS should have from a Semantic Web perspective to exploit academic CV.

2. Related work

The work presented here builds upon (Lera et al., 2014) done at the University of Balearic Islands for five years of curriculum data transformation, from relational into Semantic Web data. Its purpose was to manage data of researchers owned by the university in order to obtain useful knowledge in governance tasks, transparency and better curricular services. This university has also adopted the

CVN standard into its CV management system called GREC. This adoption ensures a common format for the CV of the researchers, and a machine-readable definition of that format.

Information systems specialized in managing research information have been developed lately from a national government scope, such as NARCIS (Dijk et al., 2006) in Netherlands, and DeGois (Fontes et al., 2018) in Portugal. There are others, which started with an open and worldwide perspective, such as VIVO (Börner et al., 2012), software that uses a semantic approach for managing research information. It has become popular: there are currently over 140 VIVO implementations in the United States, and international VIVO projects in over 25 countries, as VIVO site registry shows. Significant partners include CASRAI (Consortium Advancing Standards in Research Administration Information), EuroCRIS (Current Research Information Systems), and the ORCID (Open Researcher and Contributor ID) Initiative. First steps towards expressing CERIF -the euroCRIS format- in VIVO have been taken and will be continued (Vestdam, 2013). Then, all CERIF formats, like CVN shortly, would be compatible with Semantic Web through VIVO implementations.

Regarding tools for migrating data from non-semantic to semantic domain, (Bohring & Auer, 2005) (Kramer et al., 2015) works are based on the use of XML and the XML Schema. Both require human supervision to create the XML to OWL translator.

OWL ontologies have been developed for domains related to research, primarily publication related, like Bibliographic Ontology (BIBO) (Dimić et al., 2012) or Event Ontology (Raimond & Abdallah, 2007). And about people and its social information, like vCard (Iannella & McKinney, 2014), or relations, like FOAF (Finin et al., 2005). The Integrated Semantic Framework of VIVO leverages those ontologies in a unified, semantic structure.

There are others not included in VIVO and related to research, like Citation of Publications (Shotton, 2009a; 2009b). Management of scholarly products is an emerging research area in the Semantic Web field known as Semantic Publishing. Some examples are Semantic Conference (Nuzzolese et al., 2016), or DBLP (Ley, 2002). There is also a huge quantity of research data on the web that are not semantic, but otherwise retrievable: services such as Google Scholar, Microsoft Academic Search, CiteSeerX, etc.

3. Methodology

We have divided in three main issues the enhancement of CV data with Semantic Web technologies: data migration, CRIS improvement, and CV composition.

3.1. Data migration

We cannot make any assumptions about the format of that legacy data. The database could be already semantic, such as in (Lera et al., 2014), or in a well schemed XML format, such as CVN, or in CERIF-compliant, such as a euroCRIS implementation. All these formats simplify the migration to semantics, but relational models could be a chaotic or poor documented relational database. As we focus on the CV, the data in the database of the university would be a minor loss for uCV since each CV has just the data we are interested about.

Researchers usually make CV on a human-readable format, though not easily machine-readable from any parser: pdf, odt or xdoc. Moreover, each researcher uses a different visual composition of the information in its own CV document. Parsing those documents to interpret its data semantically is not feasible because of the need of human guidance. To avoid this, CRIS have been increasingly including CV document management services: mostly some kind of entry CV data interface, and sometimes the possibility to get the CV of a researcher as a document. Whether these services include

CV document exportation to a format that allows its data to be interpretable will be a key to uCV. That is the case of CVN, where its CRIS offers web services to get CV documents in XML format, although to universities only.

Most of CV documents have no machine-understandable semantic data. Its data is semantically interpretable by humans only, by means of the structure, headers and disposition of it (i.e., we can imply whether a date is for a starting job experience despite there is not any written assertion that describes it so). If these metadata are somehow tied to each element of the CV, and they are machine-readable, then we can build some application to interpret the data in a CV document, in order to generate its machine-understandable semantic counterpart. A document format that allows us to do so is the eXtensive Markup Language (XML), along with its XML Schema Definition (XSD).

A well-formed XML CV document is enough to convert its data to an OWL instance (Bohring & Auer, 2005). However, the supporting OWL ontology would be changing as new XML documents with new CV elements arrived to be converted to OWL format. For that reason, an XML Schema Definition for the CV document provided by the owners of the CRIS would be preferable. Then, we could ensure the generation of an OWL ontology suitable for all valid XML CV documents. On the context of XML, "valid" implies the compliance with some document type definition, in this case an XML Schema for the researchers' CV of this CRIS. Because the CRIS would generate valid XML, uCV could use that XML Schema to easily convert the whole CV data of the XML document into an OWL instance.

Thus, data migration to uCV from a university adopter of CVN is possible. The data of their researcher's CV could be retrieved as XML document, and then converted to OWL instances. As soon as we have the data in OWL, it can be directly stored as RDF triples without any pre-processing.

3.2. CRIS improvement

Firstly, we focus our attention on the ontology as it is the key to data sharing. Having an OWL ontology based on an XML Schema Definition for our domain of researchers' CV does not ensure interoperability with other research-related ontologies. We need a common ontology that allows us to infer equivalent meanings. As far as we know, there is not a W3C standard for CV description. VIVO uses a set of ontologies focused on scholarship domain, easily extended to support additional domains of scholarly activity. Besides, VIVO is also a web platform that supports recording, editing, searching, browsing and visualizing scholarly activity. Its semantic approach encourages research discovery, expert finding, network analysis and assessment of research impact. All its data is linked within the context of a person, and that data can lead to other valuable sources.

Another searched improvement is the ability to infer. The main benefit is that users entering new data can make minimal assertions based on what they know at the time that new information is added to the application. Discovering additional statements has the potential to provide novel services to a Semantic Web application.

Finally, VIVO offers some interesting features for a CRIS. It can be customized through its CRIS Components to extend the presentation, tools and functionalities. It is able to draw a graph about network collaborations between Researchers using data in the CRIS Entities: Projects, OrgUnits, Dynamic Objects, etc. Its statistics system is able to collect and visualize statistics about the CRIS Entities and allow further extension. An implemented extension uses scripts that query external bibliometric databases for citation counts: Scopus and Web of Science. The system is able to use the existent metrics to build derivative metric aggregating the information to an upper level, for example the total number of citation received by all the researcher's publications. This is the publication citation count metric aggregated to the researcher level. Other than aggregation it is possible to calculate average, maximum, minimum, standard deviation, variation in a week and month and also sum different metrics together.

We focused on the problem of the research information management that will differentiate uCV because of the use of Web Semantic technology. In particular, uCV can integrate other sources of data from the web into its data collections. And it takes advantage of the logic constructs of the ontology to infer statements about data stored on uCV. From the researcher's point of view, we want to minimize the amount of work necessary to have a completed CV up-to-date. From the research decision makers' point of view, we want: updated and extensive information about the research activity and the human capital of the university; fast aggregated information building; and CV evidences correction and verification capabilities.

3.3. CV composition

We expect to have the ability to deliver CV documents of researchers, composed by following appropriate schemas and style sheets. For that, we need some document schema definition. Knowing which data to put where, we can set the semantic queries to get data and build the CV document properly. This work needs to be made by human domain experts. However, it is not expected the document schema definition to vary frequently. And, if it does, semantic technology adapts well to this kind of changes. Only we will need the experts again to supervise the changes made.

We also need a transformation tool from pure data to a human-friendly presentation format, like PDF or HTML. That tool would be probably offered by the same organization that makes the call, along with the document schema definition.

We use CVN as schema for uCV to comply with when composing CV. CVN, as official standard, is commonly required on public research calls in Spain.

4. uCV procedures

uCV implementation deals with the interpretation of data of several CV models; with transformation to semantic models; with reasoning ability to enhance new assumptions; with harvesting in order to obtain external data; with a building process, and finally with the composition to generate CV in a customizable style sheet. All these procedures set up uCV and they are explained in the next subsections.

4.1. Interpretation

This procedure consists of applying a transformation to a CV document such that, for each data or group of related data elements on it, one or more RDF semantic assertions are created.

Our incoming CV document must be a well-formed XML document and valid against an XSD. This XSD will probably define a normalized standard CV for some nation, organization or call. This is the case of CVN.

Interpretation consists of three stages: document parsing, assertions generation, and graph loading. In Figure 1. we show the process. On thinnest lines, we see a CV document in XML format transformed into a programming language object by a parser. Then, the calling functions of that object do the RDF assertion generation depending on the known meaning of each object that composes the CV document object. Finally, the graph loading is a simple process in which each of those RDF triples are stored in the graph for that CV-standard. Bolder lines represent files, tools and actions required once it is the time for each CV standard. The XSD document shall be provided in advance in order to generate (N) the corresponding XML parser for those CV documents. We use

Kuhlman library (Kuhlman, 2017) as the XML parser generator. This software also generates data structures in Python representing the elements in the XML document described by its XSD (L). In our example, we call the generated Python class: CVN.py, and it is generated (N) by means of a simple command:

$ python3 generateDS -o CVN.py CVN.xsd

CVN.py has the data structures and the parser function.

XSD is also used at (K), to create an ontology for the data, which that CV standard is designed for. For test purposes, we created the ontology for some of the CVN elements using an ontology editor like Protégé (Knublauch et al., 2014). For (J) and (I) it is almost inevitable the need for a domain expert to tailor the ontology for the CV-standard (J) and the ontology to translate that ontology to uCV and vice versa (I).

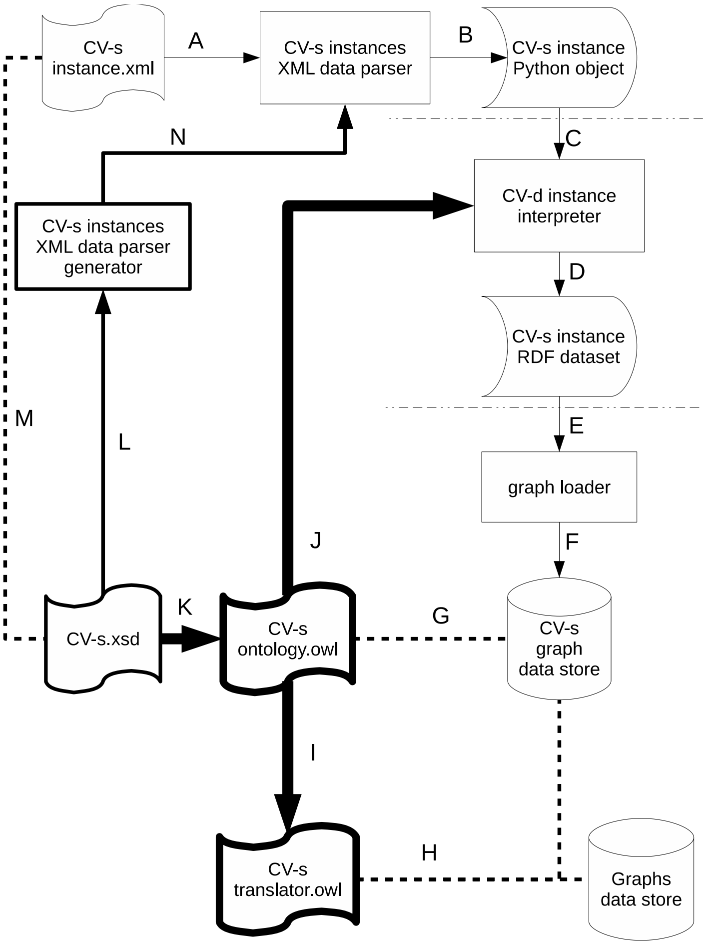

Figure 1. Software tools, model files, and instance files for the interpretation process of a CV document that follows an specific standard: CV-s. Medium bold lines are actions and tools used once for each standard. Boldest lines are actions that require the intervention of a domain expert. Thinnest lines are for each CV document interpretation

In order to manage name entity recognition, we add a mapping function to uCV ontology. If element names are expressive enough, we can use the implicit information that the structure of the document provides to interpret the semantics of each element. And thus, what semantic assertion in OWL language could be equivalent or fairly similar at least. In the contrary case, it is necessary for the expertise participation to define mapping rules. For example, an element GivenName nested in an element Author, in turn nested in an element CvnItem, is semantically different than another element GivenName, nested in an element PersonalIdentification, in turn nested in the Agent element.

Finally, for simplicity reasons, we use SPARQL as the language to create the RDF assertions.

4.2. Translation

Whether the university already had data about its faculty members in semantic format, it should be integrated within the uCV database. In order to do so, having that data in RDF format would be required. Thus, we can directly import all RDF triples to a new graph on the uCV database. The database import interface of VIVO accepts the most common syntaxes for RDF: XML, Turtle, N-Triples and JSON-LD. Probably, the ontology used by the university will use some classes and properties specifically created for its particular domain. Therefore, a problem of semantic interoperability arises: we will need to merge both ontologies. The strategy followed by the merging process is to consider the ontology of the university more specific than the ones used on uCV. Those -FOAF, BFO, BIBO, DC, SKOS...-have a global aspiration, and thus its definitions are broader, more inclusive and expecting specific definitions to extend it further. On the other hand, an ontology of a university probably has specific definitions, not expecting any ontology to extend its concepts. Therefore, we use owl constructors such as: subclass and subproperty for classes and properties of the ontology of the university with respect to uCV ontologies. Equivalency for classes and properties is used with caution because of the implications the inference engine could reason (Halpin et al, 2010).

4.3. Reasoning

The reasoning engine used is the embedded reasoning service of VIVO. It currently consists of a combination of Pellet for reasoning on the class and property hierarchies and a custom-built reasoner that maintains certain types of inferred statements about individuals. Thus, the VIVO application currently depends mainly on class hierarchy reasoning based on subclass and equivalent class relations. When an individual is asserted to be a member of a given class, the reasoner adds additional statements to show that it is also a member of any parent or equivalent classes. This allows data to be entered at a lower level of the class hierarchy (e.g., InvestigatorRole) and to be queried and discovered at a higher level (e.g., Role). This principle also applies when aggregating VIVO data involving an institution-specific ontology extension defining a new class of a VIVO core class such as uib: Researcher. Through inference, all such researchers will be recognized as members of the core Role class and found as such in search. Class hierarchy reasoning is also used to organize content on the index and menu pages, to provide facets to search results, and to ensure that properties defined for parent classes are offered for population during child class editing.

Most object properties in uCV are defined to be bi-directional through the explicit declaration in the ontology of an inverse property opposite in meaning. Bi-directional relationships allow users in the VIVO application to navigate from a person to a related department or publication while also supporting lists of department members on department pages or researchers on publication pages. An end user can easily see and navigate to contextual information from wherever they first arrive in uCV. If one direction of a property with a declared inverse has been populated, VIVO will add the inverse statement.

4.4. Harvesting

Harvesting consists of read and transforms data from external data sources and ingest it in a semantic platform. VIVO has a library of tools designed for harvesting. It is a collection of small Java tools that make the Harvester architecture extremely versatile, offering customizable and repeatable processes. A harvest process has these steps:

- Fetch. The first step is to get your data from a foreign source. VIVO has tools to fetch from relational databases using drivers like JDBC, or from OAI repositories.
- Translate. The fetched data will be in its own format, and this needs to be converted into uCV. If the input is an XML format, this can be done using a VIVO tool called XSLTranslator. Otherwise, VIVO UI is able to prepare SPARQL Constructs that will take in RDF data and transform it into the uCV ontology.
- Match. It may be needed to match incoming data with data already in uCV. VIVO does this comparison in two steps: first via a tool named Score, which ranks two values; second, a tool named Match will look at those numbers and compare them to a threshold value. Input entities compared by Scored that meet or exceed the threshold would have identities changed to the URI of the researcher in uCV, so that when the data is finally pulled into uCV the new data will be linked to the existing data.

Problem is publishers in general do not offer a direct access to their database, so the JDBC tool of Fetch is useless for that type of target source. They do offer an API, such as Semantic Scholar, Google Scholar, DBLP, or even a Linked Data API, like Elsevier and Springer. Some semantic publishers offer an SPARQL engine to answer queries, such as Scholarly Data.

However, VIVO has not implemented yet those API in its Fetch tool to interact with. An alternative would be a web crawler, which wrote the raw data collected from different sources to a relational database for foreign data sources. And then the harvester would fetch to that database in order to get the data of interest. Both alternatives are outside the scope of this work, so harvesting has not been tested.

4.5. Building

The schema definition sets which data to put where. Therefore, following that definition, uCV can ask the database for the corresponding data and put the response on the appropriate place. For easy implementation, we use XML as the destination CV document format uCV must compose to, and XSD as the schema definition for that CV document. This is the case of CVN. Then CVN apply a style sheet to the document using XSL in order to create a printable document on PDF format.

In Figure 2, we show the harvesting procedure. H-module creates a CV instance XML builder complying with the CVN schema. This sends the SPARQL queries to the data storage, and builds the XML with the data gotten. For that, we use a Python library based on (Kuhlman, 2017) already used on for XML parsing.

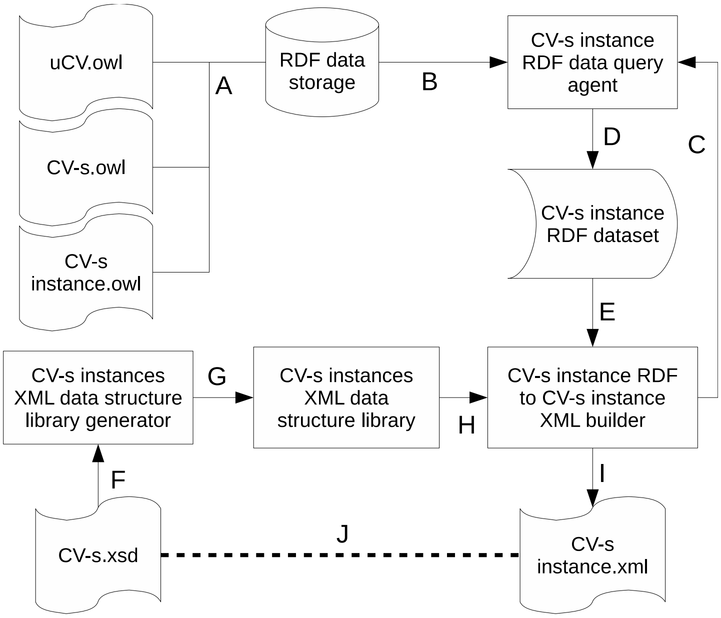

Figure 2. Software tools, model files, and instance files for the process of build a CV document that follows CVs standards.

In this case, we use it because of its capability of creating an object-oriented data structure (G) from an XML Schema (F). It also includes a function to build an XML compliant with that XML Schema (I). Then, we use this library to, instead of getting the data from parsing an XML document; get them from the RDF data storage (E). Therefore, for each object in the CV structure, we implement the corresponding SPARQL query to get the proper data (C).

4.6. Composition

Once we have an XML document built and compliant with the XSD, we should apply it a transformation to get a human-friendly presentation of its data. In Figure 3. this procedure: the transformation tool gets the XML built with the CV data (A), and uses some transformation rules to create a new document with a style sheet applied (B), in order to offer a human-friendly presentation of the CV document. As the XML is compliant with an XSD (C), a straightforward transformation would use XSL technologies, defining a document (D) with the transformations to apply to any XML document that would have that XSD schema.

However, we do not consider this as part of uCV. Thus, it should be provided for the organization behind. Furthermore, the inclusion of it is easier now due to the evolution of web services as open services on the Internet easily accessible through REST APIs. Then, uCV would only need to access that REST API with the XML document, and get the PDF as a result of consuming the web service. That is the way CVN works, with a web service that returns a PDF transformation of the XML CVN-compliant sent. However, Spanish government does not offer this web service as an open service, but only to those universities that have passed a validation process.

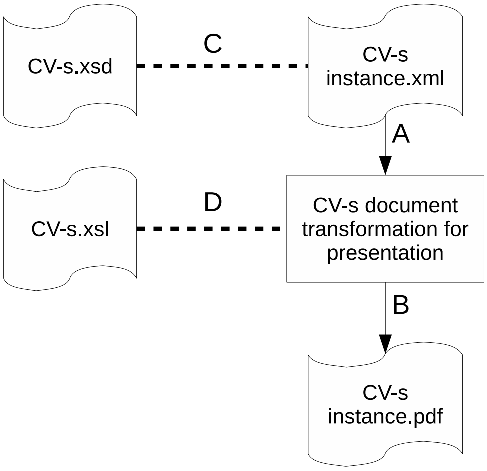

Figure 3. Process of composing a CV document that follows a specific standard, into a human-friendly format like PDF

5. Discussion

According to (Morzinski & Schubot, 2000) main drawbacks in CV evaluation are incompletion and self-reporting. Incompletion would be easier as the web grows in available research data. Regarding self-reporting, Semantic Web along with Open Link Data ease the contrasting process with other sources which in turn requires some trust management. As more data sources were included in uCV, more relevant trust management would become. The strategy proposed by (Richardson et al., 2007) could be applied to uCV.

Other strategies that could benefit from semantics would be author inference detection. Existing Software, as a Service for researchers such as ResearchGate (Thelwall & Kousha, 2015), implements the suggestion of authorship to researchers. A similar service could be implemented on uCV: with semantics embedded in both authors and publications, the prediction algorithm could be improved, i.e. by adding the degree of relationships between subjects (using SKOS or BFO ontologies) or authors (using FOAF ontology).

The information system proposed, uCV, strongly depends on the existence of normalization processes for CV documents on the universities. Despite these conditions, its compliance should not be unusual: W3C is an international reference organization for standardization, and its XML family is the dominant standard for information exchange on the Web. And normalization processes are the rule for those organizations, which must interoperate and integrate; we saw it with the European Community and some countries in regard to the research data they are interested to manage. As long as the normalization process has an open data purpose, it would surely involve XML. As the universities are expected to have a sharing and "open to the world" purpose, uCV is expected to be of interest for this type of organizations.

6. Conclusion

CV documents are the current mechanisms to evaluate human capital from scientific projects and programs. In this article, we have presented an exhaustive analysis to transform classical CV management tools into the paradigm of Semantic Web. This adaptation generates a system able to simplify governance tasks, researchers' CV adaptations to call programs, validations and trust processes, and mainly, tasks regarding evaluations and comparisons. Our approach, uCV, is designed to improve current systems and not start with a new system from scratch. For that reason, we have considered three key issues: better use of existent resources through a data migration, an enhancer of CRIS services using Semantic Web standards, and improving CV composition for generation of reports. uCV is composed of the following procedures or phases: interpretation, translation, reasoning, harvesting, building, and composition.

uCV proves to be valuable with only two data sources -a university database and a CVN of a researcher- and one target format. The more data sources, the more data could be inferred, both for every CV and about the university. And the more destinations are included to export CVs to, the more valuable uCV for researchers would be.

Harvesting data about publications and presentations is an important functionality to leverage the Semantic Web. The technologies and the information allow including it. It should be the next step of improvement on uCV. A more distant but relevant issue is how space and time can impact the representation of and reasoning about individuals and classes on organizations. Besides, it can be utilized for ontology modularization, evolution, and the handling of vague and contradictory knowledge

References:

Asserson, A. (2010). CERIF Best Practice. *Data Science Journal*, *9*, CRIS32–CRIS38. DOI: http://doi.org/10.2481/dsj.CRIS5

Bohring, H., & Auer, S. (2005). Mapping XML to OWL Ontologies. *Leipziger Informatik-Tage*, *72*, 147-156.

Börner, K., Conlon, M., Corson-Rikert, J., & Ding, Y. (2012). VIVO: A semantic approach to scholarly networking and discovery. *Synthesis lectures on the Semantic Web: theory and technology*, *7*(1), 1-178.

Caire, G., & Becker, G. (1967). Human Capital, A Theoretical and Empirical Analysis with Special Reference to Education. *Revue économique*, *18*(1), 132-133.

Curriculum Vitae Normalizado CVN (2017). Fundación Española para la ciencia y la tecnología. Retrieved October 18, 2017, from https://cvn.fecyt.es/

Dijk, E., Baars, C., Hogenaar, A., & van Meel, M. (2006, June). NARCIS: The Gateway to Dutch Scientific Information. In *Proceedings of the ELPUB2006 Conference on Electronic Publishing*, (pp. 49-58).

Dimić Surla, B., Segedinac, M., & Ivanović, D. (2012, September). A BIBO ontology extension for evaluation of scientific research results. In P*roceedings of the Fifth Balkan Conference in Informatics* (pp. 275-278). ACM.

Finin, T., Ding, L., Zhou, L., & Joshi, A. (2005). Social networking on the semantic web. *The Learning Organization*, *12*(5), 418-435.

Fontes, M., Assis, J., Araújo, E. R., Bento, S., Henriques, L., Pirralha, A., & Santos, L. D. D. (2008). *Building new indicators for researchers' careers and mobility based on electronic curriculum vitae: the case of the DeGóis platform* (pp. 1-13). INETI/Dinâmia.

Halpin, H., Hayes, P. J., McCusker, J. P., McGuinness, D. L., & Thompson, H. S. (2010). When owl: sameas isn't the same: An analysis of identity in linked data. In In*ternational Semantic Web Conference* (pp. 305-320). Springer, Berlin, Heidelberg.

Iannella, R., & McKinney, J. (2014). *vCard Ontology-for describing People and Organizations.* Retrieved from https://www.w3.org/TR/2014/NOTE-vcard-rdf-20140522/

Knublauch, H., Fergerson, R. W., Noy, N. F., & Musen, M. A. (2004). The Protégé OWL plugin: An open development environment for semantic web applications. In *International Semantic Web Conference* (Vol. 3298, pp. 229-243).

Kramer, T. R., Marks, B. H., Schlenoff, C. I., Balakirsky, S. B., Kootbally, Z., & Pietromartire, A. (2015). *Software tools for XML to OWL translation*. http://dx.doi.org/10.6028/NIST.IR.8068

Kuhlman, D. (2017). *generateDS -- Generate Data Structures from XML Schema*. Davekuhlman.org. Retrieved October 18, 2017, from http://www.davekuhlman.org/generateDS.html

Lacunza, I., & Arellano, D. (2011). *Curriculum Vitae Normalizado. Jornada técnica sobre CRIS y repositorios.* Fundación Española para la Ciencia y la Tecnología (FECYT).

Ley, M. (2002). The DBLP Computer Science Bibliography: Evolution, Research Issues, Perspectives. In: A. H. F. Laender & A. L. Oliveira (Eds.), *String Processing and Information Retrieval*. SPIRE 2002. Lecture Notes in Computer Science, vol 2476. Berlin, Heidelberg: Springer.

Lera, I., Guerrero, C., & Juiz, C. (2014). Indicators and functionalities of exploitation of academic staff CV using semantic web technologies. In P*roceedings of the 23rd International Conference on World Wide Web* (pp. 1221-1226). ACM.

Linked Data - Design Issues. (2017). Retrieved October 18, 2017, from https://www.w3.org/DesignIssues/LinkedData

Lorenzo-Escolar, N., & Pastor-Ruiz, F. (2012). Un análisis de los principales sistemas de identificación y perfil para el personal investigador. *Aula Abierta*, *40*, 107-118.

Morzinski, J. A., & Schubot, D. B. (2000). Evaluating faculty development outcomes by using curriculum vitae analysis. *Family Medicine*, *32*(3), 185-189.

Nuzzolese, A. G., Gentile, A. L., Presutti, V., & Gangemi, A. (2016, October). *Generating Conference Linked Open Data in One Click.* Paper presented at the 15th International Semantic Web Conference, Kobe, Japan.

Raimond, Y., & Abdallah, S. (2007). *The event ontology*. Retrieved from http://motools.sourceforge. net/event.

Richardson, M., Agrawal, R., & Domingos, P. (2003). Trust management for the semantic web. *Lecture Notes in Computer Science*, *2870*, pp. 351-368.

Schöpfel, J., Prost, H., & Rebouillat, V. (2017). Research Data in Current Research Information Systems. *Procedia Computer Science*, *106*, 305-320. http://dx.doi.org/10.1016/j.procs.2017.03.030

Shanks, J., & Arlitsch, K. (2016). Making Sense of Researcher Services. *Journal of Library Administration*, *56*(3), 295-316. http://dx.doi.org/10.1080/01930826.2016.1146534

Shotton, D. (2009a). Semantic publishing: the coming revolution in scientific journal publishing. *Learned Publishing*, *22*(2), 85-94.

Shotton, D. (2009b). *CiTO, the Citation Typing Ontology, and its use for annotation of reference lists and visualization of citation networks*. Paper presented at Bio-Ontologies 2009 Special Interest Group meeting at ISMB.

Simons, E. (2013). *euroCRIS, CERIF and CRIS*. Paper presented at Workshop on Empirical Foundations of Science and Innovation Policy.

Thelwall, M., & Kousha, K. (2015). ResearchGate: Disseminating, communicating, and measuring Scholarship? *Journal of the Association for Information Science and Technology*, *66*(5), 876-889.

Vestdam, T. (2013). *The future of CRIS systems-an interplay with VIVO*. Elsevier.

Iñaki Salinas
Universitat de les Illes Balears. Department of Computer Science,
Palma,
Spain
inyakisalinas@gmail.com

Isaac Lera
Universitat de les Illes Balears. Department of Computer Science,
Palma,
Spain
Isaac.lera@uib.es

Carlos Guerrero
Universitat de les Illes Balears. Department of Computer Science,
Palma,
Spain
Carlos.guerrero@uib.es

Carlos Juiz
Universitat de les Illes Balears. Department of Computer Science,
Palma,
Spain
cjuiz@uib.es

Enes Sukic
University for Business Engineering and Management,
Banja Luka,
Bosnia
sukic.enes@gmail.com