Linked Data, JSON-LD and the Semantics of Cultural and Scientific Heritage

Ovidiu Turcoane

Academy of Economic Studies, Bucharest, Romania
turcoaneovidiu06@stud.ase.ro

Abstract. This paper discusses the importance of linked data for the semantics of cultural and scientific heritage and it provides some reasons on using JSON-LD as a standard. From a conceptual point of view, cultural and scientific heritage should widen for broad human and computer access. Web semantics support the transformation of information to knowledge on a large and open scale and cultural and scientific heritage should benefit from that, especially if we see the latter as a pillar in a lean system, such as E-Democracy that seeks for permanent improvement. Cultural and scientific heritage has a primary role in education of individuals but also in the optimization process of knowledge society through the humane collective approach of trial and error. In order to build a society of inclusion using vertical (temporal) and horizontal (spatial) inquiries in social values, religions, beliefs and rituals we must facilitate broaden and less restrictive access. One way to achieve this task is by using linked data for connectivity and meaning of information and the standard of JSON-LD for extensive and simple digital access.

Keywords: cultural and scientific heritage, linked data, JSON-LD, web semantics, E-Democracy

1 Introduction

Cultural and scientific heritage (CSH) is important for knowledge society, which is a pre-requisite for E-Democracy that is an instrument and a goal itself for the digital era. In this paper, we briefly introduce the knowledge society in the context of digital democracy (i.e. Electronic Democracy or E-Democracy) and we discuss the role of CSH for both of them. We also try to prove that a web-oriented architecture (WOA) based on web semantics (WS) and linked data (LD) is most suitable for knowledge dissemination in a democratic society. From the conceptual level, we go down to the logical and physical level and we propose the standard of JSON-LD as the formatted data that endorses the implementation of LD, subject to a simpler human and computer manipulation of resources. We compare JSON-LD to other standards and we try to identify pros and cons of using either of them.

2 CSH in the Context of E-Democracy and Knowledge Society

While in the 1970s CSH was seen as rather tangible, material (buildings and sites having universal value from the perspective of art, history etc.), the 2000s have included in CSH the intangible practices, knowledge, skills etc. that belong to individuals, groups or communities [3]. The benefits from tangible CSH are well known: economical (e.g. tourism), social (e.g. symbolic representation of unity and belonging), psychological (e.g. enjoyment and delectation) and cultural itself. The intangible CSH and particularly historical knowledge are very important in a society based on continuous optimization. Although there are studies that attempt to evaluate quantitatively CSH, tangible [2] and even intangible [8], the qualitative quota is priceless.

In this paper we focus on intangible CSH and especially on its role in building a knowledge society, an open network of networks society cross-trained with technology [4]. In digital era, data and information are abundant, but processing and applying them in order to obtain knowledge are the keys for better results of individuals and communities. Intangible CSH should be viewed from at least twofold perspective: temporal and spatial.

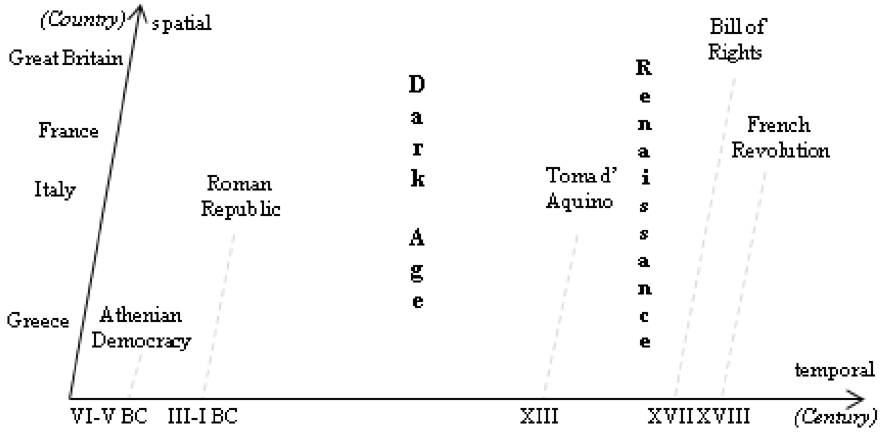

Fig. 1. Twofold perspective of intangible CSH

Fig. 1 illustrates the intangible CSH from democracy's point of view and we may acknowledge the importance of it. Without the rediscovery of antiquity wisdom by Thomas of Aquinas who knows how far the dark age would prevail!? We nowadays know that the Aristotelian logics and his cultural, social and political legacy had an enormous contribution to revitalization of human life in the dawn of the spiritual rebirth: Renaissance. Thenceforth, the liberalism of John Locke (on which the Bill of Rights is framed) and the ideological crusade of Jean-Jacques Rousseau for democracy (the platform of French Revolution) led to modern civilization.

Yet, the democracy's antique CSH was very tangible from the perspective of its storage equipment, i.e. manuscripts that penetrated the middle age through the Arabic-Latin Medieval Iberia's translators channel. Therefore, without the preservation of

CSH for more than a millennium, chances for democracy's reactivation would have been seriously diminished. This may be another empirical solid reason for the endeavor of nowadays scholars that advocate for digital preservation of CSH. One may never know when units of CSH become an input in the process of transforming human knowledge in outputs of a better society.

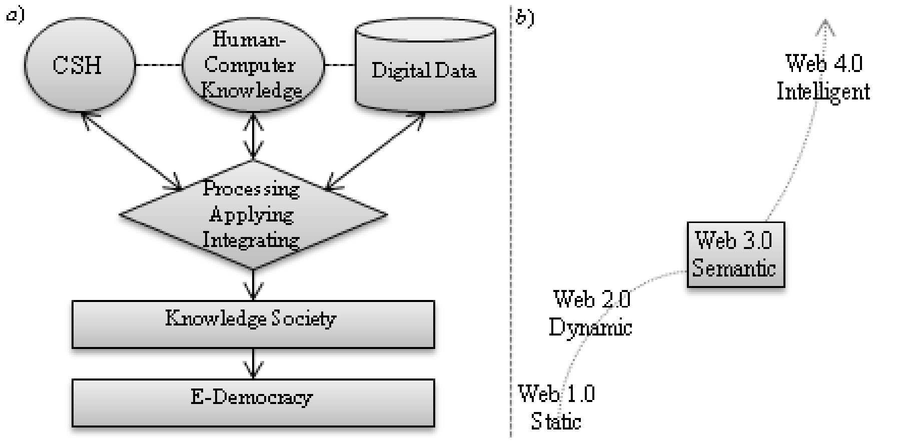

Fig. 2. Semantic digital world: *a)* Knowledge society model; *b)* Web evolution

A more elaborated description of E-Democracy was already presented in a previous article [10]. We just want to discuss here the importance of knowledge society as a system that is based on participation, deliberation and inclusion. From a computational point of view, it may be seen as a self-teaching entity that is subject to permanent optimization through trial and error processes. Knowledge society and E-Democracy overlap each other from many points of view, but the former does not imply the political process and it may develop less democratic approaches. Fig. 2 illustrates a model of knowledge society that relies on semantics, on meaning of the human-digital relationship and the data that describes this connection. Each input of the model reshapes itself and each other through a conceptual middleware framework developed under three characteristics: processing, applying and integrating (PAI). CSH, knowledge and data are subjects to continuous transformation and improvement, and from a digital perspective they should be part of an open and supple architecture, like Web 3.0. Cloud computing and services oriented architecture are definitely tools for mediating PAI for ubiquitous devices and all categories of online end-users. However, a more pervasive and wide approach is needed in a paradigm that is open (source) and accessible to most if not all users like WOA, which requires few resources (hardware, financial, knowledge etc.). More, WS adds new value to any digital architecture through metadata, giving meaning and integrating data in an ultra-network of networks. Practically, web data is dubbed into information when LD and WS are employed, and in the near future semantics will probably level up to intelligent digital systems in the paradigm of Web 4.0.

3 LD and JSON-LD as Support of WOA and WS

WOA is built up on and it is an extension of Services Oriented Architecture (SOA), a modular software architecture based on open standards and interoperability of separately designed services but with greater openness towards end-users. More, the former keeps the principle of the latter: reusability, contract, loose coupling, abstraction, composability, autonomy, statelessness and discoverability. Finally, WOA is a global emergent SOA defined neither by a vendor nor by a standards body. It is based on Representational Stateless Transfer (REST), Hyper-Text Transfer Protocol (HTTP), Uniform Resource Identifier (URI) or more generic IRI (Internationalized Resource Identifier) and communication dependent on client state [7]. Designed as a style of using HTTP, REST is not a standard but a guide on what is to be delivered, providing a sort of semantic for communication using four verbs (PUT, POST, GET, DELETE). The principles of REST [5] reside on using: a) IRI, b) resources through representations, c) self-descriptive messages and d) hypermedia as the engine of application state (HATEOAS). With WOA and REST there are certain advantages: easy access via IRI, simple produce-consume process, supple interface through browsers, no traditional troublesome communication contracts, easy implementation in programming languages and scripts and high performance (also relying on cache mechanisms).

WS is not a concurrent for Web 2.0, but rather a natural evolution with a key impact on knowledge management by trying to bring into play content that is machine-accessible through some meaning. While in Web 2.0 searching, extracting, maintaining and viewing information require human involvement, WS enables knowledge management through [1]: knowledge organized in conceptual spaces, automated tools for maintenance and knowledge discovery, semantic query answering, query answering over several documents and defining available parts of documents. The logic behind WS is based on processing metadata (data about data) that captures part of the meaning of data. In order to provide semantic interoperability by overcoming differences in terminology, WS uses ontology, an explicit and formal specification of a conceptualization, which relies on two main components: terms (i.e. concepts) and relationships (e.g. hierarchies) between terms. The process of retrieving information is the logic that stands behind software agents that make decisions and select courses of action. This process may be difficult and one way to make it easier is to begin from the moment of data structuring and implementation.

3.1 LD

LD supports the openness of WOA, the principle of HATEOAS and the meaning of WS, being a basis for a new kind of web by implementing four concepts [6]:

- Use IRIs/URIs as names for things. Not only documents and web content, but also real objects (e.g. terms) and abstract concepts (e.g. relationships) should be referenced with IRI because: a) globally unique names are created in a decentralized manner and b) names are means of accessing information describing identified entities.

- Use HTTP IRIs, so that people can look up those names. It is also important for data coherence not to confuse the objects with their describing web documents.
- When someone looks up an IRI, provide useful information, using the standards. (We herein advocate for JSON-LD and JavaScript opposed to RDF and SPARQL)
- Include links to other IRIs, so that they can discover more things. Using LD, hyper-links have types that describe relationships between things in a global space.

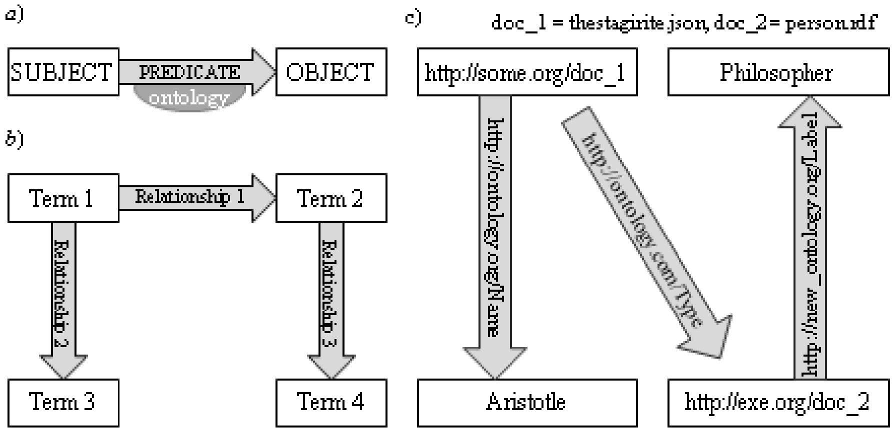

Fig. 3. LD with example

One item of LD contains three elements (i.e. a triple): two terms (i.e. subject and object identified by IRIs) and the relationship between them (i.e. predicate, defined by ontology and identified by IRI). There are three categories of links in LD [6]: a) relational (point at related things in other data sources), b) identity (IRI aliases that enable different views on the same thing) and c) vocabulary (self-descriptive data).

LD does not only link concepts through some meaning but creates a global database that has a supple architecture and not a structured one like relational databases or even data warehouses. While this architecture is flexible, querying data from LD files is more complicated than using Structured Query Language in a relational database system, for example. Query depends on data implementation and the most common approach is to use SPARQL (SPARQL Protocol And RDF Query Language) to access RDF (Resource Description Framework) files. Not only that RDF and SPARQL have a less human-readable language, but also they require a database system that is not easy to be implemented. The solution to overpass this drawback is to use a standard that is validated by millions of web applications and which abides the principles of WOA and WS: JavaScript Object Notation (JSON) for LD, i.e. JSON-LD. JSON-LD facilitates interoperability between different systems and applications, as JSON is a more simple way of supporting web services than XML (eXtensible Markup Language) standards.

3.2 JSON-LD

Starting with 2014, JSON-LD is a standard of World Wide Web Consortium (W3C) that preserves all gimmicks of JSON and it offers a smooth upgrade to LD paradigm without disturbing already deployed systems by introducing [11]:

- IRIs to identify JSON objects,
- context for disambiguation of shared IRI keys between documents,
- web references in JSON documents,
- annotation of strings with their language,
- data types such as date and time,
- graphs or networks in a single document.

Table 1. JSON-LD main terminology with examples

Keywords	Description	Example

@context	short-hand names that are used throughout a JSON-LD document	{
		"@context":
		 {"name":"http://schema.org/name"
		},
		 "name": "Aristotle
		}

@id	uniquely identifying things that are being described in the document with IRIs or blank node identifiers	{
		"@context":
		
		{"name":"http://schema.org/name"},
		"@id":"http://ex.Aristotle.com/",
		"name": "Aristotle
		}

@value	data that is associated with a particular property	{
		"@value": "2014-06-30",
		"@type": "http://www.ex.org/date"
		}

@type	type of value utilizing: @type, value object or JSON type (number, true, false)	
@graph	grouping a set of nodes in one conceptual entity	"@graph": [{"@id": "http://ex.org/Aristotle", "@type": "Philosopher"}, {"@id": "http://ex.org/Aquino", "@type": "Priest"}]

More, JSON-LD has some important factors that make it a design choice: simplicity (i.e. no extra processors or software libraries), compatibility (i.e. zero edits with JSON files or usable as RDF), expressiveness (e.g. serializing almost all real data model) and terseness (e.g. easily human readable). The minimum terminology to deploy JSON-LD is presented in Table 1, while Table 2 presents additional keywords.

Table 2. JSON-LD additional keywords with examples

Keywords	Description	Example

@language	language for a particular string value or the default language of a JSON-LD document	{
		"@context": {"@language": "bg"},
		"name": "Велико Търново"
		}

container	default container type for a term	{
		"@context":{"child":
		{"@container": "@set"}},
		"child": ["Dan"", "Mary", "Dana"]
		}

@list	ordered set of data	
@set	unordered set of data	
@reverse	serializing in the reverse direction	"children":{"@reverse": "http://ex.org/Parent"}

@index	information that processing should continue deeper into a JSON data structure	{
		"@context":{"paper":
		{"@container": "@index"}},
		"paper":
		{"en": "CSH,
		"ro": "Moştenirea Culturală}
		}

@base	setting the base IRI against which relative IRIs are resolved	{
		"@context":
		{"@base": "http://ex.com/doc"},
		"@id": ""
		}

@vocab	setting a common prefix to be used for all properties and types that do not match a term and are not an IRI	{
		"@context":{
		"@vocab": "http://schema.org/",
		"databaseId": null},
		"name": "Aristotle",
		"databaseId": "23987520"
		}

JSON-LD is superior to RDF (we consider in this paper that RDF stands for RDF/XML) especially if we take into consideration some elements presented in Table 2 (e.g. set and list), which helps providing an intuitive schema that is adapted for Web

APIs. The legitimacy of JSON-LD is given by the fact that its kernel, i.e. JSON, was used as an alternative to XML, the kernel of RDF, and it became a standard imposed by the market and not by a vendor or an organization. XML is also a good choice for data interchange, hierarchies and even graphs but it unwillingly maps the variables in different programming languages. XML is generic and a markup language, but not a data serialization one, while JSON is more limited, but design for data serialization. The gap between RDF and JSON-LD is much narrower thanks to efforts of W3C (and RDF serializations like N3, Turtle), but there are some points where the latter may be fitter for LD than the former even if it was not primarily designed for WS [9].

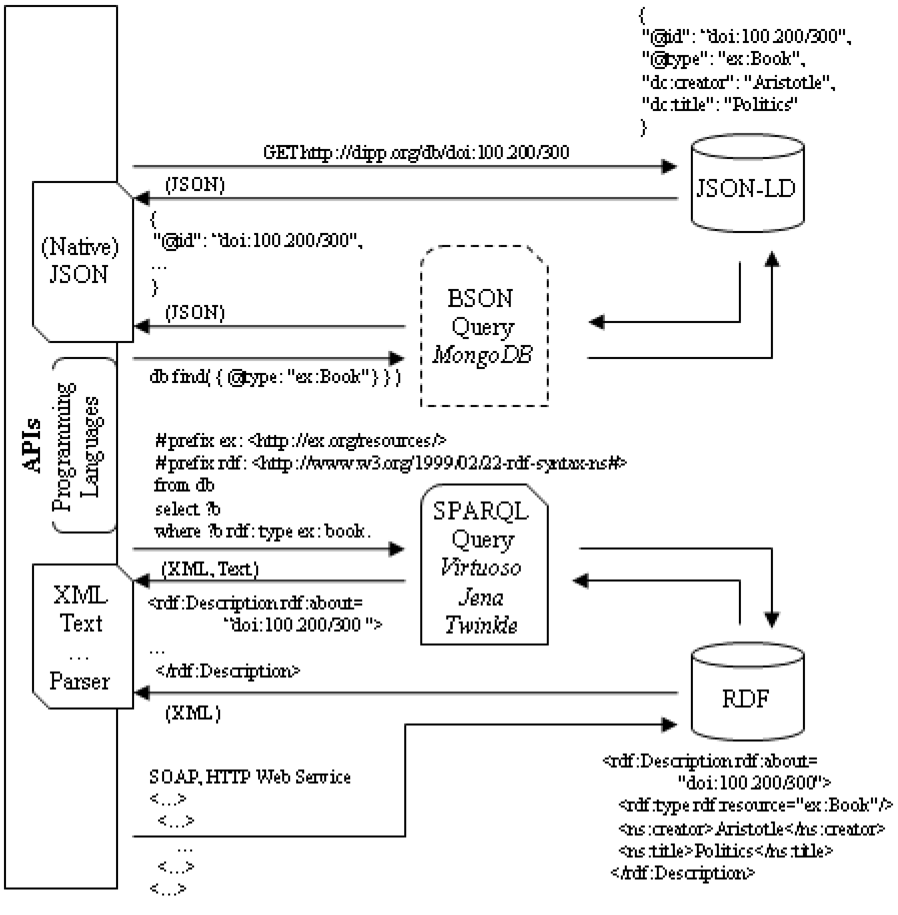

Fig. 4. LD Queries

Their database system configurations and, especially, their query languages also give the superiority of JSON-LD over RDF. Firstly, their core formats (i.e. JSON and XML) demand different approaches for accessing an item from data transaction point of view. JSON (data-oriented) is text-based and position independent and stores data in objects (records, associative arrays) and lists (arrays, vectors); being easily parsed and generated. XML (document-oriented) is more extensible, storing data in trees, but from data-interchange point of view has two important drawbacks: it carries a large envelope and it hardly complies with data formats of most programming languages. Secondly, JSON-LD and RDF inherit the above characteristics from their originators and each one's query language depends on each one's structure.

Fig. 4 illustrates the format of JSON-LD and RDF/XML along with examples of simple queries for each of them. Both of them undertake middleware approaches, but the JSON-LD's one (exemplified with Binary JSON DB system MongoDB) is lighter from both human and machine point of view. RDF direct query presumes using XML over SOAP or HTTP WS with a response filtered through some parser. JSON-LD queries demand only simple (REST/HTTP) GET requests with responses easily and/or directly read by most programming languages.

Table 3 presents the author's opinion on the pros and cons of a comparison between JSON-LD and RDF, based on their conceptualization [11, 12].

Table 3. JSON-LD versus RDF. (Alphabetical order) Pros and Cons (P/C)

Criterion	JSON-LD		RDF	
	Impact	**P/C**	**Impact**	**P/C**
Conversion	RDF to JSON-LD, easy	+	JSON-LD to RDF, hard	−
Fit for REST	High	+	High	+
Fit for WOA	Very high	+	Low	−
Fit for WS	High	+	High	+
Implementation	Easy	+	Difficult	−
LD age	Young	−	From beginning	+
LD usage	Reduced	−	Intensive	+
Query language	JavaScript, easy	+	SPARQL, elaborate	−
Syntax	Easy, intuitive	+	Elaborate	−
Web age	Old (extended JSON)	+	Old (extended XML)	+

While Web age and LD age from Table 3 are not very important, although they show some sort of stability, syntax and query language are important. More, fitness for WOA is very important if we believe in high accessibility and openness of data interchange and sharing. JSON has been proving to be a highly appreciated approach for Web APIs and JSON-LD inherits its properties and it adds a new (LD) flavor.

WOA is an architecture that supports integration with few resources, WS gives machine meaning to documents and, in addition, LD offers possibility for creating large databases and JSON-LD physically implements the data-information level of the previous three. Seeing CSH also as a permanent inter-connection of different nodes of patrimony, we advocate for the four approaches (i.e. WOA, WS, LD and JSON-LD) in order to implement a network of networks and communities that share and transform data, information, knowledge and even wisdom.

Fig. 5 illustrates how CSH should evolve from physical (LD) to logical (WS) and then to conceptual (WOA) level and back, so that it becomes available to a large public. Querying and transforming (e.g. PAI) data and information from the storage level of LD to the meaning level of WS, given by the model of term-concept-reference (TCR), is a necessary step so that machine readability of data could enlarge and enable wider dissemination of information and knowledge at WOA level.

While PAI are mostly seen as sequential methods of transformation in data-information-knowledge-wisdom chain (e.g. from data to information processing is employed), we herein want to suggest that the physical, logical and conceptual levels are somehow overlaid (e.g. there are logical and even conceptual aspects in LD). More, the paradigms of LD, WS and WOA influence each other, and a connection between LD and WOA should not be impossible (transcending meaning).

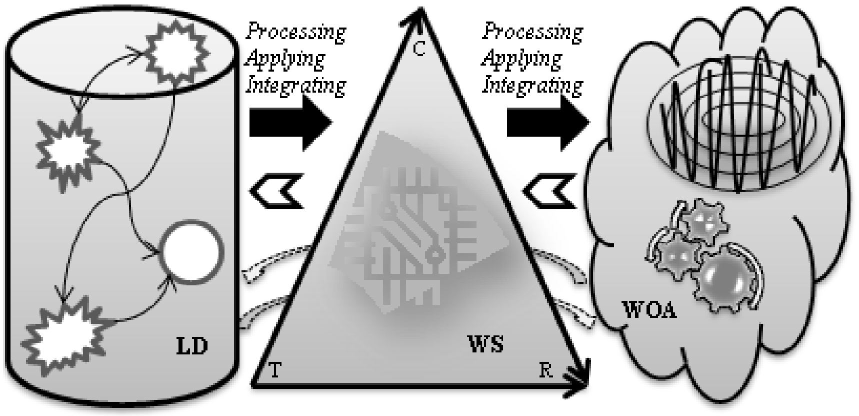

Fig. 5. CSH chain of PAI

4 Conclusions

Cultural and scientific heritage has an important role in a society that seeks for permanent improvement under constraint of epistemic uncertainty. We permanently need to learn from our legacy and compare it with the new society outputs. The accelerating dynamics of nowadays era demand fast preservation of testimonial cultural and scientific outputs, which may eventually become inputs themselves. While preservation of heritage is very important, so it is its proximate and ulterior accessibility and discoverability. These two desiderata may be better accomplished through a paradigm based on WOA, REST, WS and LD (with JSON-LD). In such a paradigm, described herein, an organization, a community or a specific group does not arrest CSH, which is rather exposed to humans (directly through Web APIs based on WOA and REST) and machines (through additional WS and LD).

A large database of CSH is permanently developing using a network or a graph of links between nodes of knowledge capital, i.e. LD. This database should be accessed by digital devices through semantics, which facilitate and mediate data discoverability and interchange, i.e. WS. Finally, CSH must be delivered to all people interested in it with minimal costs in an integrating and accessible environment, i.e. WOA.

While digital preservation of CSH will enable broad online access to knowledge, efforts must be made to eliminate exclusion of non-digital categories (defined by age or territory) or to integrate them in a more open environment. Further research should explore means to achieve a higher penetration of digital CSH in non-digital spaces.

Acknowledgement

This work was cofinanced from the European Social Fund through Sectoral Operational Programme Human Resources Development 2007-2013, project number POSDRU/159/1.5/S/134197 „Performance and excellence in doctoral and postdoctoral research in Romanian economics science domain".

References

1. Antoniou, G., & Harmelen, F. v. (2008). A Semantic Web Primer (2nd ed.). London, UK: The MIT Press.
2. Báez, A., & Herrero, L. C. (2012). Using contingent valuation and cost-benefit analysis to design a policy for restoring cultural heritage. Journal of Cultural Heritage , 13 (3), 235-245.
3. Baker, K. (2013). Information literacy and cultural heritage: developing a model for lifelong learning. Oxford, UK: Chandos Publishing.
4. Castells, M. (2005). The Network Society: from Knowledge to Policy . In M. Castells, & G. Cardoso (Eds.), The Network Society: From Knowledge to Policy (pp. 3-22). Washington, DC: Center for Transatlantic Relations.
5. Fielding, R. T. (2000). Architectural Styles and the Design of Network-based Software Architectures. University of California, Information and Computer Science. Irvine: https://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation.pdf.
6. Heath, T., & Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space. (J. Hendler, Ed.) Ebook: Morgan & Claypool Publishers.
7. Hinchcliffe, D. (2009, September 7). Web-Oriented Architecture (WOA). Retrieved May 1, 2014, from InfoQ: http://www.infoq.com/presentations/Web-Oriented-Architecture-Dion-Hinchcliffe
8. Lee, J.-S. (2014). Measuring the benefits of the Intangible Cultural Heritage Hall in Jeonju Korea: Results of a Contingent Valuation Survey. Journal of Cultural Heritage, http://dx.doi.org/10.1016/j.culher.2014.05.001.
9. Sporny, M. (2014, January). Retrieved from JSON-LD and Why I Hate the Semantic Web: http://manu.sporny.org/2014/json-ld-origins-2/
10. Turcoane, O. (2013). E-Democracy's Representatives, A Proposal And A Case Study. Econ-omy Informatics , 13 (1), 73-90.
11. World Wide Web Consortium. (2014, January). Retrieved from JSON-LD 1.0: http://www.w3.org/TR/json-ld/
12. World Wide Web Consortium. (2014, February). Retrieved from RDF 1.1 Concepts and Ab-stract Syntax: http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/