Localization of Schema.org for Manuscript Description in the Iranian-Islamic Information Context

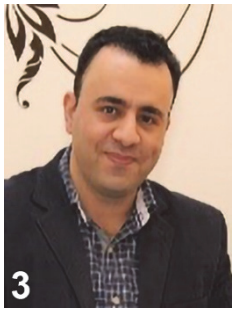
Morteza Mohammadi Ostani*, Mozaffar Cheshmeh Sohrabi**,
Seyed Mahdi Taheri***, Ahmad Shabani****, Asefeh Asemi*****

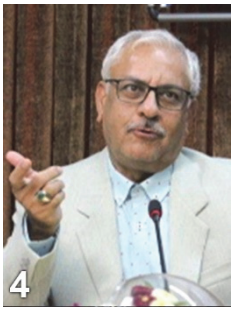
Department of Knowledge and Information Science, University of Isfahan, Azadi Ave., Isfahan, Iran.
* <mmohamadiostani@gmail.com>, **<mo.sohrabi@edu.ui.ac.ir>,
****, <shabania@edu.ui.ac.ir>, *****<asemi@edu.ui.ac.ir>
Department of Knowledge and Information Science, Allameh Tabataba'i University,
Dehkadeh-ye-Olympic, Tehran, Iran, ***<taherismster@gmail.com>

Morteza Mohammadi Ostani is a teacher, researcher and librarian. He is currently works as a visiting professor at the Department of Knowledge and Information Science, University of Qom, Iran. He studied Knowledge and Information Science at the University of Isfahan, where he defended his PhD thesis in the field of knowledge organization and retrieval. His research interests include Knowledge Organization, Ontology, Metadata, Semantic Web and Knowledge Management.

Mozaffar Cheshmeh Sohrabi, PhD (Corresponding author), is currently Associate Professor at the Department of Knowledge and Information Science at the University of Isfahan. In 2006, he completed his PhD in Information and Communication Sciences from Stendhal University, France. His research topic is the Semantic Web, Semantic Retrieval, Ontology, Data Mining, Quantitative Science Studies and Research Ethics.

Seyed Mahdi Taheri is Associate professor at Department of Knowledge and Information Science at the Allameh Tabataba'i University, Tehran. He is currently head of the Central Library and Document Center of Allameh Tabataba'i University. He got his PhD in Knowledge and Information Science from the Islamic Azad University Science and Research Branch, Tehran. His research interests currently focus on number of areas in Knowledge Organization including Interoperability of the Search Engines, Metadata Standards, Semantic Web, Digital Library and Semantic Retrieval.





Ahmad Shabani is a Professor and researcher with more than thirty years of experience in Knowledge and Information Science. He is Professor at Department of Knowledge and Information Science at the University of Isfahan. His main research topics include Information Management, Knowledge Management, Research Method and Theoretical Issue of Information Science.

Asefeh Asemi is Associate professor at Department of Knowledge and Information Science, University of Isfahan, Iran. She is currently works as a researcher at the Corvinus University, Budapest. She completed her PhD in Information Technology at University of Pune, India. Her key research interests are Library Systems, Business Informatics, Information Systems, Big Data and Data Mining.

Mohammadi Ostani, Morteza, Mozaffar Cheshmeh Sohrabi, Seyed Mahdi Taheri, Ahmad Shabani and Asefeh Asemi. 2021. "Localization of Schema.Org for Describing Iranian-Islamic Information Context's Manuscripts." *Knowledge Organization* 48(5): 345-356. 36 references. DOI:10.5771/0943-7444-2021-5-345.

Abstract: This study aims to assess the localization of Schema.org for manuscript description in the Iranian-Islamic information context using documentary and qualitative content analysis. The schema.org introduces schemas for different Web content objects so as to generate structured data. Given that the structure of Schema.org is ontological, the inheritance of the manuscript types from the properties of their parent

types, as well as the localization and description of the specific properties of the manuscripts in the Iranian-Islamic information context were investigated in order to improve their indexability and semantic visibility in the Web search engines. The proposed properties specific to the manuscript type and the six proposed properties to be added to the "CreativeWork" type are found to be consistent with other schema properties. In turn, these properties lead to the localization of the existing schema for the manuscript type compatibility with the Iranian-Islamic information context. This schema is also applicable to centers with published records on the Web, and if markup with these properties, their indexability and semantic visibility in Web search engines increases accordingly. The generation of structured data in the Web environment through this schema is deemed to promote the concept of the Semantic Web, and make data and knowledge retrieval easier.

Received: 20 January 2020; Revised: 22 April 2020, 24 May 2020, 8 July 2020, 24 August 2020; Accepted 27 August 2020

Keywords: properties, context, text, schema.org, creativework, description

1.0 Introduction

Tim Berners Lee, the web creator, has proposed a five-star rating system for data so as to realize the Semantic Web and the idea of Linked Data. This ranking system which focuses on data structure is of a cumulative style. Dourado (2014) reported that the more structured the Web data, the faster the realization of the Linked Data and Semantic Web ideas. Unlike the unstructured data, structured data allow a range of automated tasks and are processed much better than unstructured data. The existing Web is rapidly developing, and in its new architecture, i.e., the Semantic Web, content delivery should be structured. The idea of structured data has accelerated since 2011 with the introduction of Schema.org together with the Linked Data method. Yu (2014, 447) argues that Schema.org is developed as a vocabulary for the structured data markup (metadata) on the Web. Likewise, Guha et al. (2015) claim that Schema.org applies the same data model, language, syntax, and goal similar to the Linked Data. In this approach, the creation of the structured data (metadata) to improve the interoperability of information systems with the Web search engines is indeed introduced for different types of Web content object schemas.

Moreover, Schema.org allows the description of Web documents (content objects) by human or computer programs (technological agents) to make their management and retrieval easier (Schmea.org, 2020). According to Hjørland (2008 and 2016) "Knowledge Organization (KO) is about activities such as document description, indexing and classification performed in libraries, databases, archives etc." or "KO is about describing, representing, filing and organizing documents and document representations as well as subjects and concepts both by humans and by computer programs". So, the document description activity is considered as KO process, and WWW is considered a new medium, though its principles are part of the field of KO.

Soergel (1999, 1119) stated that "classification has long been used in library and information systems to provide guidance to the user in clarifying her information need and to structure search results for browsing. Recently, other fields, such as [artificial intelligence] (AI) ... have discovered the need for classification, leading to the rise of what these fields call ontologies". Following this, Nogales et al. (2016) stated that the Schema.org can be used to develop metadata ontology. Hjørland (2016) believed that ontologies are "a new kind of KOS that became important from the 1990s". Expanding on Soergel's idea, Hjørland maintained that "the ontologies are basically classification systems and that they represent a reinvention of classification". Consequently, based on the new conception of ontology, Schema.org has gained importance and is viewed as a core subject in KO.

As a metadata standard, the Schema.org vocabularies, including the collection of properties (attributes and relationships) for describing Web documents and developing metadata ontologies, are categorized as Knowledge Organization Systems (KOS) based on the definition of KOS by Hodge (2000), Hjørland (2016) and Mazzocchi (2018).

Providing access to content objects such as manuscripts through Web search engines and developing appropriate mechanisms to improve the interoperability of the manuscript databases with them is a new issue which is dealt with in this study. The organization and access to manuscripts in their structural sense is indeed realized by Semantic Web standards and technologies.

Metadata records of the manuscripts in the Iranian-Islamic information context require semantic indexability and visibility by Web search engines, thereby increasing users' access. However, due to the lack of access to the original copy and its precise description through metadata, indexability and high access points are essential in this context.

Most manuscripts are described through different data elements and specific standards. Factors such the lack of documentation tools for name and subjects (Matlabi 2013; Azimi and Nazi 2011), the lack of unit rules and standards (Nazi and Ghasempour 2011), and the existence of different dates in the manuscripts such as *Shahanshahi*, *Jalali*, *Yazdgerdi*, *Babylonian*, animal, lunar, coded chronometry systems (Nabavi et al. 2014), hinder the semantic understanding of the manuscripts in the Iranian-Islamic information context in the realm of Web search engines. Indeed, in the dynamic and ever-changing Web environment, the manuscript metadata records encounter different problems in indexability and visibility in different search engines.

Realizing the Semantic Web followed by semantic indexability of data through computers has recently promoted the accessibility of the manuscripts and made data and knowledge retrieval easier. However, there is no standard metadata for the specific and local properties of the manuscripts in the Iranian-Islamic information context. The cataloging rules and principles of manuscripts like Anglo-American Cataloging Rules (AACR), Ancient, Medieval, Renaissance and Early Modern Manuscripts (AMREMM), Descriptive Cataloging of Rare Books (DCRB), Electronic Access to Medieval Manuscripts (EAMMS) are not indeed appropriate to describe and organize manuscripts in the Iranian-Islamic information context for retrieval because their focus is mainly on the description of western and medieval manuscripts in their appropriate context. Given its cultural and religious aspects, the Iranian-Islamic information context has local and native features that are not suggested in the aforementioned standards. Indeed, the precise description of the decorations, arts, corrections and script types, ownership issues, the different structure of preface, etc. constituted these local and native features.

As for the cultural heritage context (including galleries, libraries, archives, and museums) in Schema.org, schemas such books, articles, periodicals, movies, creative works, music, websites etc. are developed for each manuscript type. Given the richness and high volume of subjects in the Islamic countries and the data repositories of the Iranian libraries, the localization of the existing schema for the description and accessibility of such content objects (documents) such as manuscripts are necessary. Considering their particular physical and content properties, the manuscripts indeed require more description than other printed documents.

Therefore, the main objective in this study was how to improve the semantic indexability and visibility of the manuscripts in the Iranian-Islamic information context. The Schema.org has proposed schemas for different documents to resolve this problem. To localize the manuscript schema in Schema.org compatible with the Iranian-Islamic information context, the properties[1] required for describing and organizing the knowledge stored in the manuscripts in the Iranian-Islamic information context should be thus identified and localized with respect to the Schema.org structure and context.

2.0 Background

There exist some studies on Schema.org mainly dating back to the last few years, indicating that this field of study is still new and worth further investigation. These studies are divided into two general groups:

1. The studies developing and designing schemas for the two levels of new vocabulary and properties, and the ones using vocabulary and properties of Schema.org in specific information contexts so as to improve the interoperability of the metadata records. The first group included a handful of studies conducted by Aldaej (2015), Aldaej and Krause (2014), Aghadeh (2018), and Fardehosseini (2019) on the design and development of schema.

Aldaej (2015) and Aldaej and Krause (2014) marked up and measured the resources and contents available in the Virtual Learning Environments (VLE) through the vocabulary related to learning resources in Schema.org, by adopting an experimental method. According to the results, Schema.org did not have sufficient semantic capability to describe the learning contexts and the VLE, in particular. This is attributable to the lack of appropriate and supportive vocabulary in describing the virtual learning environment(s). Thus, a metadata schema for learning resources was formally added to the Schema.org in April 2013.

Similarly, Aghadeh (2018) designed extensions for authority metadata records in Schema.org based on Microdata[1] using the empirical and content analysis methods. He proposed adding variant forms of authority metadata records to increase and improve access points. His results revealed an improvement and increase in access points of authority records as well as the indexability and semantic visibility of variant forms. In another study, Fardehosseini (2019) discussed the impact of standards and conceptual models for the cultural heritage context on the Schema.org. He focused on Resource Description and Access (RDA) and the effect of conceptual model of Library Reference Model (LRM) on improving the performance of the types of "CreativeWork" in the Schema.org, by adding the cultural heritage context (including galleries, libraries, archives, and museums) properties to the types of "CreativeWork". Comparing the properties of the types of the "CreativeWork" with the properties of the Library Reference Model (LRM), Fardehosseini et al. (imprint) also reported consistency in types like "book", "thing", and "organization". It is believed that in Schema.org many relations between types (data entities) do not fit the target context. As such, Schema.org should consider the properties for the description of the types that cover relationships appropriate to each context. Many cultural heritage context properties should be thus developed and added to Schema .org to improve its performance.

2. These studies are based on Schema.org and its applications, that is, the vocabulary and properties of Schema.org information contexts with specific focus are used to improve the interoperability of the metadata records. Some examples are: Mixter et al. (2014) describing thesis and dissertations with the Schema.org vocabulary, Howsky et al. (2013) discussing the positive impact of Schema.org metadata on Web information resources, Balci et al. (2018) and Simsek et al. (2017) describing the tourism domain

with the Schema.org vocabulary. The results of these studies mainly indicate an improvement in accessibility and semantic visibility in real world Web. Some other studies have focused on specific applications of the Schema.org in: markup Web pages with the Schema.org vocabulary (e.g., Rosati and Mayernik 2013), Web Observatory (e.g., DiFranzo et al. 2014), entity search on the Web data (e.g., Homoceanu 2015), and machine readable Web Application Programming Interface (API) (e.g., Simsek et al. 2018).

Aghadeh (2018) and Fardehosseini (2019), in addition to designing and developing schema, assessed the interoperability of metadata records based on Schema.org vocabulary. Their findings showed improved accessibility and semantic visibility in Web search engines. Several other studies including Wallis et al. (2017), Taheri et al. (2018), and Taheri et al. (2017) only examined the interoperability with Schema.org vocabulary. For instance, Wallis et al. (2017) applied Schema.org as a common data format to collect data in the Europeana network database so as to increase the visibility of the cultural objects of its member centers in the Web search engines. As the results showed, the Schema.org was successful in increasing the visibility of the cultural objects. Wallis et al. (2017) also provided some recommendations for the implementation of Schema.org in the Europeana database. By adopting the empirical method, Taheri et al. (2017) assessed the reaction of Google and Bing search engines to the metadata records based on Schema.org in syntactic context of Microdata and their representation in the form of resource description framework (RDF). The findings showed that the property values of Schema.org were indexed in the syntactic context Microdata and RDF, thereby lacking semantic visibility, such properties were not recognized by Google and Bing search engines. Likewise, Taheri et al. (2018) used an empirical method to compare the reaction of Google, Yahoo and Bing search engines to indexing and visibility of metadata records based on Schema.org in the two syntactic contexts of HTML and Microdata. The findings showed that all three search engines were able to index metadata records, however it was only Google that found metadata records semantically.

Yet, no study has been ever conducted on the localization of Schema.org for the manuscript description in the Iranian-Islamic information context. Some studies conducted by Arabgari (2017), Nabavi et al. (2014), Fakhari Saadat (2014), Arabgari et al. (2013), and Omrani (2006) only compared the manuscript records with the metadata standards of the cultural heritage contexts like: Anglo American Cataloging Rules (AACR), MARC, Dublin Core (DC), Metadata Object Description Schema (MODS), Metadata Authority Description Schema (MADS), Text Encoding Initiative (TEI), and Resource Description and Access (RDA).

However, such studies were mainly focused on the application of Schema.org for business purposes. Besides, most of these studies used Schema.org to describe and improve indexability and semantic visibility of the objective contents. Some other studies developed properties and vocabularies compatible with their information context (i.e., localization), while some other only focused on interoperability studies that confirmed the applicability of the Schema.org therein. Overall, a review of the literature revealed that no study has been ever conducted on the localization of the manuscripts in the Iranian-Islamic information context. Given that such precious and valuable resources are essential in promoting the Islamic-Iranian cultural values of the country, this study attempted to localize Schema.org for manuscript description in the Iranian-Islamic information context using documentary and qualitative content analysis.

3.0 Methodology

The research method used in this study is of applied-developmental, given that the required properties for describing the manuscripts in the Iranian-Islamic information context is applicable to generating the structured data on the Web as well as mapping the knowledge graph. Localization and recommendation of specific new properties for the manuscripts in the Iranian-Islamic information context, which were absent in this schema, is a kind of development of the existing knowledge in this context. According to the research objectives, this study was thus conducted in two parts: 1) document analysis method and 2) the qualitative content analysis method.

In the first part, the literature and sources related to the manuscripts in the Iranian-Islamic information context and the existing standards were reviewed systematically by applying the documentary analysis method. The required properties (attributes and relationships) for describing and organizing the manuscripts were extracted from the related literature, and were then categorized and completed through Delphi technique[2] in two rounds. These properties consisting of three hierarchical levels including 19 main and 224 sub-properties[3] (Mohammadi Ostani et al. 2018) were considered as the research statistical population.

In the second part, the content analysis method was used to analyze, identify and match the properties of the research population with the properties of the Schema.org. Then, localization of the specific properties of the manuscripts was performed in accordance with the Schema.org context.

In the content analysis method, it is important to determine the "content units" and "semantic units". The content units are of referential and thematic type, while the semantic units are manuscript properties. The recording units are in the form of words and phrases of elements' names which include the required properties for the description of the manuscripts in the Iranian-Islamic information context. The context units, in turn, include the properties of the

No.	Properties	Expected Type	No.	Properties	Expected Type
1	additionalType	URL	7	mainEntityofPage	CreativeWork or URL
2	alternativeName	Text	8	name	Text
3	description	Text	9	potentialAction	Action
4	disambiguationDescription	Text	10	sameAs	URL
5	Identifier	PropertyValue or Text or URL	11	subjectOf	CreativeWork or Event
6	Image	ImageObject or URL	12	URL	URL

Table 1. "Thing" type properties.

No.	Properties	Expected Type	No.	Properties	Expected Type
1	about	Thing	9	hasPart	CreativeWork or Trip
2	abstract	Text	10	isPartOf	CreativeWork or Trip
3	accessMode	Text	11	producer	Organization or Person
4	author	Organization or Person	12	publisher	Organization or Person
5	Creator	Organization or Person	13	schemaVersion	Text or URL
6	dateCreated	Date or DateTime	14	timeRequired	Duration
7	exampleOfWork	CreativeWork	15	workExample	CreativeWork
8	fileFormat	Text or URL	16	workTranslation	CreativeWork

Table 2. Some of "CreativeWork" type properties.

"CreativeWork" and "Thing" types as parents in the Schema.org context.

To localize the manuscript schema in schema.org according to the Iranian-Islamic information context and also to determine the required properties, the properties identified in the first part were analyzed and matched with the properties of "Thing" and "CreativeWork" parent types through a checklist so as to identify similar properties. To complete the manuscript schema, the value of each property, its description and data type were then determined according to the Schema.org structure and context.

4.0 Results

Given that the manuscripts of other content objects including the "book", "article", "blog", "website", etc. are considered as creative works, according to the ontological approach of the Schema.org, "Manuscript" type are categorized in the "CreativeWork" type. According to the Schema.org data model, all the properties of parent types, the "CreativeWork" and "Thing", inherited all of its subtypes like "book", "article", "blog", "map", "webpage", etc. The inherited properties of the "Thing" type for "Manuscript" type as a parent type are tabulated in Table 1.

According to Table 1, twelve properties of the "Thing" type are applicable to the "Manuscript" type and can be applied to describe the manuscripts in the Iranian-Islamic information context. Some of the properties inherited from the "CreativeWork" type for the "Manuscript" type as a parent type are tabulated in Table 2.[4]

Similarly, ninety-two properties of the "CreativeWork" type are applicable to the "Manuscript" type and can be applied to describe the manuscripts in the Iranian-Islamic information context.

To localize the schema.org for the manuscripts in the Iranian-Islamic information context, the properties of statistical population consisting of 19 main and 224 sub properties were matched to the inherited properties of the parent types ("Thing" and "CreativeWork"), see Table 3 below.

As shown in Table 3, sixty-three main and sub-properties of the manuscripts in the Iranian-Islamic information context were matched to "Thing" and "CreativeWork" type properties. In turn, twenty-eight main and sub-properties of the manuscripts in the Iranian-Islamic information context matched the "description" property of the "Thing" type.

In the next step, each remaining main and sub-properties of the research population was assessed according to the context and structure of the Schema.org. As the findings showed, twenty-one specific properties of the related manuscripts were identified to be specifically localized to the Iranian-Islamic information context. The specific properties of the type of the manuscripts in the Iranian-Islamic information context along with their expected type and description are tabulated in Table 4.

No.	Main and sub main properties	Equivalent in Schema.org	Inherited from	Expected type
1	Author	Author	CreativeWork	Organization or Person
2	Title	Name	Thing	Text
3	Alternative title	alternativeName	Thing	Text
4	Type of manuscript	Genre	CreativeWork	Text or URL
5	Classification	Identifier	Thing	Product or Text or URL
6	Identifier	Identifier	Thing	Text or URL
7	Access condition	conditionsOfAccess	CreativeWork	Text
8	Date of scribal	dateCreated	CreativeWork	Date or Date-Time
9	Place of scribal	locationCreated	CreativeWork	Place
10	Date of re-scribal	dateModified	CreativeWork	Date or Date-Time
11	subject	about	CreativeWork	Thing
12	Correction	correction	CreativeWork	CorrectionComment or Text or URL
13	Registration number	Identifier	Thing	Text or URL
14	Pervious registration number	Identifier	Thing	Text or URL
15	Rights	copyrightHolder	CreativeWork	Organization or Person
16	Right holder	copyrightHolder	CreativeWork	Organization or Person
17	Patronage (*muhdi-al-ilayh*)	sponsor	CreativeWork	Person
18	Language	inLanguage	CreativeWork	Language or Text
19	Related formats	workExample	CreativeWork	CreativeWork
20	Translator	translator	CreativeWork	Organization or Person
21	Certificates of transmission (*Ijāzah*)	description	Thing	Text
22	*Mojiz*	description	Thing	Text
23	*Mojaz*	description	Thing	Text
24	Date of *Ijāzah*	description	Thing	Text
25	Form	fileFormat	CreativeWork	Text or URL
26	Contents	description	Thing	Text
27	Summary	abstract	CreativeWork	Text
28	*ʿarḍah did*	description	Thing	Text
29	Material	material	CreativeWork	Product or Text or URL
30	Paper	material	CreativeWork	Product or Text or URL
31	Papermaker	contributor	CreativeWork	Person
32	Cover	material	CreativeWork	Product or Text or URL
33	In Cover	material	CreativeWork	Product or Text or URL
34	Endpaper	material	CreativeWork	Product or Text or URL
35	Dusk jacket	material	CreativeWork	Product or Text or URL
36	Spine	material	CreativeWork	Product or Text or URL
37	Cover maker	contributor	CreativeWork	Person
38	Miniaturist	contributor	CreativeWork	Person
39	Illuminator	contributor	CreativeWork	Person
40	Binding	description	Thing	Text
41	Type	description	Thing	Text

Table 3. Matching manuscript properties to the properties of "CreativeWork" and "Thing" types.
(To be continued on next page)

No.	Main and sub main properties	Equivalent in Schema.org	Inherited from	Expected type
42	Place of binding	description	Thing	Text
43	Date of binding	description	Thing	Text
44	Binder	contributor	CreativeWork	Person
45	Treatment	description	Thing	Text
46	Place of repair	description	Thing	Text
47	Date of repair	description	Thing	Text
48	Restoration maker	contributor	CreativeWork	Person
49	Damage and pests	description	Thing	Text
50	Cover	description	Thing	Text
51	Paper	description	Thing	Text
52	Binding	description	Thing	Text
53	Symbol	description	Thing	Text
54	Symbol note	description	Thing	Text
55	Sign	description	Thing	Text
56	Music symbol	description	Thing	Text
57	*Barfaraz*	description	Thing	Text
58	Stamp	description	Thing	Text
59	Stamp mark	description	Thing	Text
60	Stamp shape	description	Thing	Text
61	Stamp description	description	Thing	Text
62	Catchwords	description	Thing	Text
63	Accompanying material	description	Thing	Text

Table 3. (Continued)

No.	Property	Description
1	contentStructure	Description of the beginning and end of the text body or manuscript including pre-incipit, incipit, final incipit, post-final incipit, rubric, final rubric, and colophon, where a defective, figure, and language is considered for each. Values are mentioned after (:) (e.g. "incipit: *besmele* ...")
2	copyrightComment	In this note, items like the ownership sign, inheritance, and waqf (*wāqif*, *mawqūf ʿalayhi*, *mawqūf*, place and sign of Waqf) are recorded before their value.
3	correctionComment	A note about manuscript correction that includes collator, place, date and sign (*balagha*) of collation, auditor (*musmiʿ*), auditors (*sāmiʿūn*), and place and date of *samāʿ*. Values are mentioned after (:), (e.g. "place of collation: Astan Qudes Razavi Library" or "collator: Hussein al-Abdi").
4	correctionType	Determine the type of correction in the manuscript. Expected values include "collation (*muqābalah*)", "*samāʿ*", "*qirāʾaht*".
5	custody	Activities to preserve and maintain the manuscript. Acceptable values are "photographing", "transmitting", or "displaying".
6	decorationComment	By applying decorations in paper and cover (dusk Jacket, endpaper, and spine), a reference is made to the theme and style of miniature. The expected value for paper and cover include Watermark, Marbled paper, Rule-borders and frame, Roundel, Headpiece, Mandorla, Onlays, Arabesque, Chrysography, Vignette, etc.

Table 4. Local and specific properties of the manuscripts in the Iranian-Islamic information context in the Schema.org.
(To be continued on next page)

No.	Property	Description
7	decorationType	Determine the type of decorations in the manuscript. (e.g., "Miniature", "Illumination", "Painting", etc.)
8	encodedDateCreated	In some cases, the date of writing the manuscript is expressed in code(s) as meaningful or meaningless term(s), showing the date mostly by adding up the numerical values of the alphabet letters.
9	encodedDateModified	Encoded date of re-scribal
10	layout	Layout type of the manuscript in terms of column, number of lines, tabs, and tables. Value are mentioned after (:) e.g. "line number: 15".
11	locationModified	A place of re-scribal
12	nameLocation	Location of name (title) in the manuscript
13	nameProvider	In some cases, the name (title) may be provided by someone else. Expected value include: "cataloger", "scribe", "none".
14	numberOfPages	Manuscript number of pages or leaves
15	originality	The originality and authenticity of the manuscript. Expected value include: "original manuscript" (*Musawwadah*, *Mubayyaḍah*), "copies" (transcript) and "model (exemplar) manuscript".
16	pagination	A description of the manuscript pagination. The expected value can be written in "letters", "numeric", "*abjad*", or "code".
17	physicalDescription	Describe the article dimensions, text and cover; the paper color, cover color (dusk Jacket, endpaper, spine), and ink color (text, margin, incipit, colophon, *Ijazah*, entailment). Expected values for dimension: "octavo", "crown octavo", "super octavo", "folio", "quarto", "12*15 cm", expected values for paper color : "*hanaei*", "*nukhudi*", expected values for ink color: "*dhahab*", "cinnabar", "ahmar", "vermilion (*zunjufr*)", "lapis lazuli (*lāzuward*, *lāziward*)".
18	re-scribalScribe	The person who completes the unfulfilled manuscript of the scribe.
19	scribe	The person who writes a manuscript by hand.
20	script	The script used in writing of the manuscript. Items like type, size and quality of script are addressed in this property. Expected values for script type: "*Kufi*", "*Naskh*", "*Nastaliq*", "*Taliq*", "*Siyaq*", "*Ruq'ah*", "*Ghubar*", "*Toqrah*", etc; expected values for script size : *"Ghubar"* (very small), *"Khafi"* (small), "*Jali"* (big); expected values for script quality: "excellent", "*Khosh*", "readable", "unreadable".
21	uniformName	A title (name) with which the work is better known and shows the works related to it.

Table 4. (Continued)

The most suggested properties are specific to the manuscripts and are compatible with the Iranian-Islamic information context. These properties are of special status in the standards of cultural heritage context, but are not mentioned in the Schema.org.

Six properties are suggested to be added to the "CreativeWork" properties' type, as these properties, apart from the manuscript type in the Iranian-Islamic information context, are also applicable to other types of "CreativeWork" like "book", "article", "blog", "map", "webpage", etc. These six properties are tabulated in Table 5.

As observed in Table 5, the last four properties are in the relational form existing between entities of the "CreativeWork" type in the bibliographic world. These properties can be, in turn, defined and applied to the Schema.org to describe and organize the manuscript type in the Iranian-Islamic information context. It should be noted that the relations in the bibliographic world are indeed much greater between entities or types.

5.0 Discussion and conclusion

To localize the manuscript type in the Iranian-Islamic information context, the manuscript properties (identified in the research population) were matched to the properties of the parent type manuscripts. The results indicated that sixty-three properties corresponded to the "Thing" and "CreativeWork" properties, as shown in Table 3. The "description" property was, in turn, consistent with twenty-eight properties of the manuscript type in the Iranian-Islamic information context. This property describes one data entity in general and, given its generality, it can accommodate many other cases. This property is indeed similar to the note area in the International Standard for Bibliographic De-

No.	Property	Expected type	Description
1	accrual Method	Text	The method by which items are added to the collection. The best recommended is applied the controlled vocabulary. Expected value include: "purchase", "donate", "transfer", "order", "loan", "entailment", "*waqf*".
2	descriptor	Person	Explanation or description of creative works
3	hasAddition	CreativeWork or Text or URL	Applying this property, when creative works are unrelated to the body (in margin like along translation), like a creative works that is added to it. Inverse property: isAdditionOf.
4	hasDescription	Text or URL	Apply this property, when creative works are related to body (in margin with translation). A description of this creative works. Related items like extent of margin, place and date of margin, date of translation, and style of translation is referred. Inverse property: isDescriptionOf.
5	isAdditionOf	CreativeWork or Text or URL	A creative works that this creative works is an added of Inverse property: hasAddition.
6	isDescriptionOf	Text or URL	A creative works that this work is a description of. Inverse property: hasDescription.

Table 5. Suggested properties to add for the "CreativeWork".

scription (ISBD) for the description and organization of the content objects.

Twenty-one specific properties were accordingly suggested for the manuscript type in the Iranian-Islamic information context, as shown in Table 4. These proposed properties are mostly addressed in the metadata standards of the cultural heritage context, as well. However, the lack of Schema.org attention to the standards of cultural heritage is much felt in the schemas of other entities. As such, localizing and suggesting such properties like "contentStructure", "copyrightComment", "correctionType", "decorationType" and "decorationComment" are essential in describing and organizing the manuscripts in the Iranian-Islamic information context.

Six properties were suggested to be added to the properties of "CreativeWork" type, as depicted in Table 5. These properties in addition to being of manuscript type, are applicable and inherited to other types of "CreativeWork" like "book", "article", "blog", "map", "webpage", etc. The four proposed properties were in relational form, applicable to the bibliographic world, and were commensurate with the Schema.org context in all subtypes of "CreativeWork". The proposed property of "accrualMethod" complied with the Dublin Core Metadata Standard, which fitted well into the Schema.org context in all subtypes of "CreativeWork".

However, there exists no study in the literature investigating the localization of the manuscript type in the Iranian-Islamic information context. The present study is indeed different from other studies conducted in the field of Schema.org. For instance, some studies including Aldaej (2015) and Aldaej and Krause (2014) that have developed the schema or properties, suggested the proposed schema for virtual learning context resources. Also, Aghadeh (2018) designed added properties for authority records in Schema.org, and Fardehosseini et al. (imprint) developed properties of "book" type by applying metadata elements of cultural heritage context standards (like LRM). In general, metadata standards of the cultural heritage context have more properties than Schema.org for describing content objects, indicating its maturity and focus for the organization of content objects.

This study thus intended to localize the properties of manuscript type in accordance with the Iranian-Islamic information context. The proposed schema is applicable to centers that have published or tend to have metadata records related to the manuscripts in the Iranian-Islamic information context on the web. By using the Schema.org vocabulary for markup, it is indeed possible for researchers and other users to better retrieve and access manuscripts in formats like Microdata, RDF, or JSON-LD. With Microdata markup, the structured data is integrated within the main HTML of the page, whereas JSON-LD uses a Javascript object to insert all of markup into the head of the page, which is often a simpler implementation from a development perspective (Jaiswal, 2018). The use of such a schema will also lead to more structured data in the Web environment, the realization of Semantic Web, and retrieval of related knowledge. A record of the manuscripts in the Iranian-Islamic in-

```
<div itemscope itemtype="http://schema.org/Manuscript">
        <span itemprop="about">علوم اسلامی -- قرن ۱۲ق</span>
        <span itemprop="about">اسلام - پرسشها و پاسخها</span>
        <div itemprop="author" itemscope itemtype="http://schema.org/Person">
        <span itemprop="name">مهدی ابن ابی‌ذر</span>
        </div>
        <div itemprop="scribe" itemscope itemtype="http://schema.org/Person">
        <span itemprop="name">جعفر بن محمدحسین الکاشانی</span>
        </div>
        <div itemprop="copyrightHolder" itemscope itemtype="http://schema.org/Organization">
        <span itemprop="name">کتابخانه ملی جمهوری اسلامی ایران</span>
        </div>
        <span itemprop="identifier">1085855</span>
        <span itemprop="inLanguage">عربی</span>
        <span itemprop="name">مشکلات العلوم</span>
</div>
```

Figure 1. A markup record based on the designed schema in Microdata syntax.

formation context marked up with the designed schema in the Microdata syntax is illustrated in Figure 1.

Moreover, localization and application of this schema along with its suggested properties make it possible to represent manuscripts metadata records in the Google Knowledge Graph (KG) and display structured data in other Web search engines, which in turn lead to more indexability and semantic visibility of the metadata records followed by easier knowledge retrieval. Online information providers and Web designers are thus eager to use this schema to make their content, especially the content of cultural heritage context, more indexable and visible. Indeed, many content owners spend a lot of money and consult with companies active in SEO, seeking high rankings in Web search engine results, while adhering to and applying this schema and markup its content objects can lead to high ranking in search engine results, and thus higher semantic visibility in Google's Knowledge Graph.

The implications of the results obtained are to increase the end-user satisfaction resulting from the relevance of the search results, and also to display the semantic metadata records of the manuscripts in the Iranian-Islamic information context through this schema, as are the ultimate goal of any search engine. As such, the properties identified for the manuscripts in the Iranian-Islamic information context can be applied to create metadata application profiles for describing and organizing manuscripts, especially in the electronic environment. These properties can also increase access points and improve accessibility to the manuscripts, which consequently can increase manuscriptology studies. Moreover, the properties of the manuscripts in the Iranian-Islamic information context have different applications, including enhancing or completing Text Encoding Initiative (TEI), incorporating other content standards, and implementing them in library software or Content Management Systems (CMS). Finally, conducting further studies to improve the semantic of Schema.org, and extending its vocabularies on the basis of the reference conceptual models for the cultural heritage context and content standards, like LRM, RiC CRM, CIDOC CRM, and RDA would be contributive.

Notes

1. Microdata was introduced in 2009 as a part of the HTML5 specifications (Homoceanu 2015, 22), and the initial syntactic context was recommended by the Schema.org to markup web contents.
2. A method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem. In other words, Delphi technique is a series of intensive questionnaires combined with controlled opinion feedback (Habibi et al. 2014, 8).
3. Properties, fit to the structure of the Schema.org, included the attributes and relationships that were termed the "property" in this study.
4. For a look at all the properties of the "CreativeWork" type, see http://schema.org/Manuscript.

References

Aghadeh, Samira. 2018. "Designing Authority Data Schema Based on Microdata Method and Study of Search Engines Reactions to it." Master's thesis, Allameh Tabataba'i University.

Aldaej, Abdulaziz A. 2015. "An Enhanced Semantic VLE Based on Schema.org and Social Media." PhD diss., University of Surrey.

Aldaej, Abdulaziz A. and Paul Krause. 2014. *An Enhanced Approach to Semanitic Markup of VLEs Content Based on Schema.org*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.666.2801&rep=rep1&type=pdf

Arabgary, Leila. 2017. "A Comparison of Metadata Standards Entries in Persian Manuscripts Databases with Metadata Standard Entries in Manuscripts Databases Outside of Iran." *Research on Information Science & Public Libraries* 23: 467-86.

Arabgary, Leila, Masoumeh KarbalaAgha Kamran and Saied Rezaei Sharifabadi. 2013. "Metadata Standards Employed in Persian Manuscript Databases." *National Studies on Librarianship and Information Organization* 24(3): 132-48.

Azimi, Habibollah and Ayoub Nazi. 2011. "Analysis of Cataloging Manuscripts in Iran to Respect the Presence or Absence of a Pattern in it and Providing Improvement Solutions to Create a National Standard Template." *Information and Library Science* 14(2): 263-82.

Balci, Boran Taylan, Umutcan Simsek, Elias Kärle and Dieter Fensel. 2018. "Analysis of Schema.org Usage in the Tourism Domain." *ArXiv* abs/1802.05948: n. pag.

DiFranzo, Dominic, John S. Erickson, Marie Joan Kristine T. Gloria, Joanne S. Luciano, Deborah L. McGuinness and James Hendler. 2014. "The Web Observatory Extension: Facilitating Web Science Collaboration through Semantic Markup." In *WWW '14 Companion: Proceedings of the 23rd International Conference on World Wide Web.* New York: Association for Computing Machinery (ACM), 475-80.

Dourado, Antônio Miguel Torres. 2014. "An Approach to Publish a Data Warehouse Content as Linked Data." Master's thesis, Instituto Superior de Engenharia do Porto (isep), 2019, http://recipp.ipp.pt/bitstream/10400.22/5607/1/DM_AntonioDourado_2014_MEI.pdf

Fakhri Sa'adat, Adeleh. 2014. "The Study of Compliance Rate of Astan Quds Razavi's Bibliographical Records Manuscripts Database with Resource and Descriptive Access (RDA)." Master's thesis, Imam Reza International University.

Fardehosseini, Mahsa. 2019. "Investigating the Impact of Library Context of Standards and Concepts Models on Improving the Usability of Rich Snippets." PhD diss., Islamic Azad University, Science and Research Branch (Tehran).

Fardehosseini, Mahsa, Seyed Madi Taheri, Najdla Hariri, Fahimeh Babalhavaeji and Fatemeh Nooshinfard. 2020. "Representing Properties and Relationships Between Entities of Creative Works in Schema.org Based on Library Reference Model (LRM)." *Iranian Journal of Information Processing and Management* 36(2): 533-62.

Guha, R.V., Dan Brickley and Steve Macbeth. 2015. "Schema.org: Evolution of Structured Data on the Web: Big Data Makes Common Schemas Even More Necessary." *Acmqueue* 13(9). http://queue.acm.org/detail.cfm?id=2857276

Habibi, Arash, Azam Sarafrazi and Sedigheh Izadyar. 2014. "Delphi Technique Theoretical Framework in Qualitative Research." *International Journal of Engineering Science (IJES)* 3(4): 8-13.

Hawskey, Martin, Phil Barker and Lorna M. Compbell. 2013. "New Approaches to Describing and Discovering Open Educational Resources." In *Proceedings of OER 13: Creating a Virtuous Circle.* Nottingham, England. http://publications.cetis.org.uk/wp-content/uploads/2013/04/OER13_resourcediscovery.pdf

Hjørland, Birger. 2016. "Knowledge Organization." *Knowledge Organization* 43(6): 475-84. Also available in *ISKO Encyclopedia of Knowledge Organization*, edited by Birger Hjorland and Claudio Gnoli. http://www.isko.org/cyclo/knowledge_organization

Hjørland, Birger. 2008. "What is Knowledge Organization (KO)?" *Knowledge Organization* 35(2/3): 86-101. https://www.ergon-verlag.de/isko_ko/downloads/ko35200823c.pdf

Hodge, Gail. 2000. *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files.* Washington, DC: Council on Library and Information Resources. http://www.clir.org/pubs/reports/pub91/contents.html

Homoceanu, Silviu. 2015. "What Search Engines Can't Do: Holistic Entity Search on Web Data." PhD diss., Technische Universitat Branunschweig.

Jaiswal, Sandhya. 2018. "What Is the Difference Between Microdata, RDFa & JSON-LD?" https://www.quora.com/What-is-the-difference-between-Microdata-RDFa-JSON-LD

Matlabi, Daryooush. 2013. "Cataloging Manuscripts." *Koliyat* 186: 2-3.

Mazzocchi, Fulvio. 2018. "Knowledge Organization System (KOS)." *Knowledge Organization* 45(1): 54-78. Also available in *ISKO Encyclopedia of Knowledge Organization*, edited by Birger Hjørland and Claudio Gnoli, http://www.isko.org/cyclo/kos

Mixter, Jeffrey K., Patrick O'Brien and Kenning Arlitsch. 2014. "Describing Theses and Dissertations Using Schema.org." In *DC-2014: The Austin Proceedings,* edited by William Moen and Amy Rushing. Texas, USA: Dublin Core Metadata Initiative, 138-46.

Mohammadi Ostani, Morteza, Mozaffar Cheshmeh Sohrabi, Ahamd Shabani, Asefeh Asemi and Seyed Mahdi Taheri. 2018. "To Identify and Determine the At-

tributes of the Iranian-Islamic Information Contexts' Manuscripts as to Information and Knowledge Organization." *Library Philosophy and Practice* 2176. https://digitalcommons.unl.edu/libphilprac/2176/

Nabavi, Majid, Gholamreza Fadaei and Nader Naghshineh. 2014. "Standards Utilized for Manuscript Description by Online Islamic Databases." *National Studies on Librarianship and Information Organization* 25(2): 22-33.

Nogales, Alberto, Miguel-Angel Sicilia, Salvador Sánchez-Alonso and Elena Garcia-Barriocanal. 2016. "Linking from Schema.org Microdata to the Web of Linked Data: An Empirical Assessment." *Computer Standards & Interfaces* 45: 90–99. doi:10.1016/j.csi.2015.12.003

Nazi, Ayoub and Sakineh Ghasempour. 2011. "Manuscript Cataloging in the National Library of Iran: Trends and Methods." *National Studies on Librarianship and Information Organization* 22(3): 104-14.

Omrani, Seyed Ebrahim. 2006. "Suggested Bibliographic Elements for the National Catalog of Manuscripts and Lith-
ographic Printing Worksheet." *Koliyat* 109-111: 9-18.

Rosati, Antonia and Matthew Mayernik. 2013. *Facilitating Data Discovery by Connecting Related Resources*. http://www.semantic-web-journal.net/system/files/swj484.pdf

Simsek, Umutcan, Elias Karle and Dieter Fensel. 2018. "Machine Readable Web APIs with Schema.org Action Annotations." In *Proceedings of the 14th International Confer-
ence on Semantic Systems. 10-13 September 2018, Vienna, Austria*, edited by Anna Fensel, Victor de Boer, Tassilo Pellegrini, Elmer Kiesling, Bernhard Haslhofer, Laura Hollink and Alexander Schindler. Procedia Computer Science 137: 255-61. https://arxiv.org/abs/1805.05479

Simsek, Umutcan, Elias Kärle, Omar Holzknecht and Dieter Fensel. 2017. "Domain Specific Semantic Validation of Schema.org Annotations." In *Perspectives of System Informatics: 11th International Andrei P. Ershov Informatics Conference, PSI 2017 Moscow, Russia, June 27–29, 2017*, edited by Alexander K. Petrenko and Andrei Voronkov. Lecture Notes in Computer Science 10742. Switzerland: Springer, 417-29. https://arxiv.org/pdf/1706.06384.pdf

Soergel, Dagobert. 1999. "The Rise of Ontologies or the Reinvention of Classification." *Journal of the American Society for Information Science* 50: 1119–20.

Taheri, Seyed Mahdi, Sara Zolghadr and Nadjla Hariri. 2018. "Comparing the Function of Web Search Engines in Indexing and Making Visible of the Metadata Records Based on Microdata Method." *Knowledge Studies* 16: 83-101. doi:10.22054/JKS.2019.38110.1210

Taheri, Seyed Mahdi, Reza Nikzad Bahle and Mitra Samiee. 2017. "Study on Search Engines' Reaction to the Metadata Records Created Based on Combined Method of Rich Snippets and Linked Data." *Iranian Journal of Information Processing and Management* 33: 639-58.

Wallis, Richard, Antoine Isaac, Valentine Charles and Hugo Manguinhas. 2017. "Recommendations for the Application of Schema.org to Aggregated Cultural Heritage Metadata to Increase Relevance and Visibility to Search Engines: The Case of Europeana." *Code4Lib Journal 36*. http://journal.code4lib.org/articles/12330#appendix1

Yu, Liyang. 2014. "Schema.org and Semantic Markup". In *A Developer's Guide to the Semantic Web*. New York: Springer, 475-516. doi:10.1007/978-3-662-43796-4_10