A Comparative Study of Descriptive Metadata in Audio Archives of IRIB Using the PBCore Metadata Standard

Zohre MIRHOSSIENI[1,*], Masoume KAZEMLOU[2]

[1]*Associate Professor of knowledge and Information Science Department, Tehran North Branch, Islamic Azad University, Tehran, Iran*

[2]*MSc in Knowlege and Information Science, Tehran North Branch, Islamic Azad University, Tehran, Iran*

Abstract. Objective: The purpose of the present research was to compare the descriptive metadata of the audio archives of the Islamic Republic of Iran Broadcasting (IRIB) to the content classes of the PBCore metadata standard (i.e. Intellectual Content, Intellectual Property, Instantiation, and Extension).

Methods: Using comparative survey, the tags of the PBCore metadata standard were extracted and translated. A checklist was also developed of fields classified under Intellectual Content, Intellectual Property, Instantiation, and Extension. Using these tags and fields the applications of each audio archive were compared to the PBCore metadata standard.

Findings: The results indicated that the descriptive metadata of IRIB's audio archives conforms to the PBCore metadata standard by 63 percent. The results also showed that audio archives conformed to the PBCore standard in terms of intellectual content, intellectual property, and instantiation, but not in terms of extension. This could be because storage and retrieval experts at IRIB focus primarily on domestic demands and ignore international standards. The results also suggest that the PBCore metadata standard can be used to describe the archives of IRIB for storage and retrieval.

Conclusion: Overall, the results indicated the conformity of the audio archives of IRIB to the PBCore metadata standard. However, extensions are needed in the web, intranet, and internet to provide a repository for metadata from external systems. This will also allow for integration and standardization in converting and creating metadata for transfer between IRIB centers and foreign archives.

Keywords: Metadata, descriptive metadata, IRIB, audio archives, music libraries, PBCore metadata standard.

1. INTRODUCTION

The advent of computers and information technology created a new environment and new media. It also created new challenging standards, methods, and procedures for organizing information. The ever-expanding knowledge organization systems that have always followed their basic principle i.e. easy and fast access to knowledge resources have opted for one of the following approaches in response to recent developments: (1) matching the existing systems with the new environment, and (2) developing modern standards, tools, and techniques to exploit the potentials of the new environment. Development of different metadata to identify, discover, describe, and utilize accessible knowledge resources in the newest information context, i.e. the World Wide Web, has been based on the second approach (Fatahi and Taheri, 2009).

The music archives of the Islamic Republic of Iran Broadcasting (IRIB) contain a large repository of audio files that are constantly used in research, programming, production, broadcast, and sales. In the shift toward digitization, these files require metadata that provide information about their content, quality, and accessibility. Due to the importance of metadata in

---

* Corresponding author. *Email address: z_mirhoseini@iau-tnb.ac.ir*

Special Issue: The Second National Conference on Applied Research in Science and Technology

digital archives, there have been various attempts to standardize them. This research focuses on the PBCore metadata standard (Public Broadcasting Metadata Dictionary). Also PBCore 2.0 is discussed as one of the major standards for broadcast and media (Cox et al., 2006) and it is applied to the audio archives of IRIB. Finally a model is proposed for using metadata in the digital music archives of this organization.

1.1. Problem Statement

An unsolved problem in the information community is whether digital resources of information centers will be fully accessible in the long term, since cataloging musical records is a challenging task that faces a huge volume of recordings, fast pace of new releases, constant mixture of information, and integration of music styles (Ghadimi, 2009).

IRIB has the largest collection of audiovisual resources in Iran. Digitization of these resources has become essential due to the increasing expansion of audio resources and the increasing variety of storage tools. Metadata are a great tool for developing digital archives. Metadata facilitates the discovery of relevant information, helps organize electronic resources, provides digital identification, and helps support archiving and preservation of the resource. Moreover, metadata allows resources to be found by relevant criteria, identifies resources, brings similar resources together, distinguishes dissimilar resources, and gives location information. There is no specific metadata for Persian music in IRIB. Thus the present research compares the meta tags in the music archives of IRIB to the PBCore metadata standard and to provide an appropriate model for describing Persian music.

1.2. Research Question

1. How well does the metadata of IRIB's music archives conform to the Intellectual Content class of the PBCore metadata standard?

2. How well does the metadata of IRIB's music archives conform to the Intellectual Property class of the PBCore metadata standard?

3. How well does the metadata of IRIB's music archives conform to the Extension class of the PBCore metadata standard?

4. How well does the metadata of IRIB's music archives conform to the Instantiation class of the PBCore metadata standard?

1.3. Hypothesis

The descriptive music metadata of the audio archives of IRIB conform to the PBCore metadata standard.

2. REVIEW OF THE LITERATURE

Hariri and Hosseini (2012) examined the metadata schemas of audiovisual resources. Schemas were evaluated using three checklists: Meta elements based on different metadata specifications, core meta elements for digital resources, and digital resource storage specifications. The results showed that Metadata Encoding and Transmission Standard (METS) and Preservation Metadata Implementation Strategies (PREMIS) conformed the most to the specifications defined for digital audiovisual resources. In addition, the results indicated that there is no single metadata schema to meet the technical specifications of all digital resources. However, Metadata Authorization Description Schema (Marc), Dublin Core Metadata Element Set (DC), and Metadata Object Description Schema (MODS) can be used for general resources, while PBCore and Visual Resources Association Core Categories (VRA Core) can be used for visual art works. Moreover, to store technical metadata, Metadata for Images in XML Schema (MIX) can

be used for visual resources, Audio-MD for audio resources, and Video-MD for video resources.

Rastgu (2012) provided a metadata model for organizing digital archives in IRIB by examining different international metadata standards in a sample of storage and retrieval worksheets for audiovisual resources. The results showed that a mixture of MODS, MIX, PBCore, TV-Anytime, MPEG-8, and MPEG-21 can be effectively used for the archives of IRIB. PBCore can be used to describe resources, MODS can be used to organize digital objects, and TV-Anytime can be used for data transfer. Moreover, MODS could be used to standardize metadata, MIX could be used to organize still images, PREMIS could be used to preserve metadata and content, and MPEG-8 and MPEG-21 could be used to observe copyrights.

According to Rubin (2011), broadcasting has been completely transformed from an analog media to a production and distribution environment that is digital from end-to-end. Due to the need for a single, unified metadata schema able to deliver digital content across multiple platforms, PBCore v1.0, based in large part on Dublin Core, was released in 2005, and has been widely adopted not only by broadcasters, but also by media archives as the preferred scheme for descriptive and technical metadata. Based on an expanding community of user institutions, PBCore is proving to be very valuable in meeting its original goal as a standard for rich media.

3. METHODOLOGY

The present research was a comparative survey. PBCore tags were extracted from the organization's website (www.pbcore.org). The tags were translated into a checklist which contained fields categorized under Intellectual Content, Intellectual Property, Extension, and Instantiation.

Checklist. Specifications of the PBCore metadata standard

Resource Type	Content Class	Measure		
Music	Intellectual Content	Asset Type[1]		☐
		Asset Date		☐
		Identifier		☐
		Title		☐
		Subject		☐
		Description		☐
		Genre		☐
		Relation	Relation Type	☐
			Relation Identifier	☐
		Coverage	Coverage	☐
			Coverage Type	☐
		Audience	Audience Level	☐
			Audience Rating	☐
	Intellectual Property	Annotation		☐
		Creator	Creator	☐
			Creator Role	☐
			Contributor	☐
			Contributor Role	☐
		Publisher	Publisher	☐
			Publisher Role	☐
		Copyright	Copyright Summary	☐
			Rights Link	☐
			Rights Embedded	☐

---

[1] Asset contains the essence, metadata, and copyright of the work. In the language of FRBR, AssetType can be used to describe an asset at the "work" level (PBCore website, September 25, 2014).

Extension	Extension		Extension Wrap	□
			Extension Value	□
			Extension Authority Used	□
			Extension Embedded	□
		Part		□
	Instantiation	Instantiation Identifier		□
		Instantiation Identifier		□
		Instantiation Date		□
		Instantiation Dimensions		□
		Instantiation Physical		□
		Instantiation Digital		□
		Instantiation Standard		□
		Instantiation Location		□
		Instantiation Media Type		□
		Instantiation Generation		□
		Instantiation File Size		□
		Instantiation Time Start		□
		Instantiation Duration		□
		Instantiation Data Rate		□
		Instantiation Colors		□
		Instantiation Tracks		□
		Instantiation Channel Configuration		□
		Instantiation Language		□
		Instantiation Alternative Modes		□
		Instantiation Essence Track	Essence Track Type	□
			Essence Track Identifier	□
			Essence Track Standard	□
			Essence Track Encoding	□
			Essence Track Date Rate	□
			Essence Track Frame Rate	□
			Essence Track Playback Speed	□
			Essence Track Sampling Rate	□
			Essence Track Bit Depth	□
			Essence Track Frame Size	□
			Essence Track Aspect Ratio	□
			Essence Track Time Start	□
			Essence Track Duration	□
			Essence Track Language	□
			Essence Track Annotation	□
			Essence Track Extension	□
		Instantiation Relation	Instantiation Relation Type	□
			Instantiation Relation Identifier	□
		Rights Instantiation	Rights Summary	□
			Rights Link	□
			Rights Embedded	□
		Instantiation Annotation		□
		Instantiation Part		□
		Instantiation Extension		□

The fields or meta tags of the audio archives of IRIB were examined through the software of each archive and were compared to the PBCore checklist.

3.1. Population

The population of this research consisted of eight audio archives of IRIB located in Tehran: Sound Department, IRIB World Service, Center for Music and Songs, Channel 1, Channel 2,

Channel 3, Channel 4, and Channel 5. Metadata elements were extracted from archives with considerable audio resources, and only music and songs were examined. The application of each archive was used to extract meta elements.

4. FINDINGS

Tables 1-7 provide the results of comparing the meta tags of IRIB's audio archives to the PBCore checklist.

Table 1. Conformity of the meta elements of IRIB Sound Department by PBCore content classes.

Content Classes	Frequency		Percentage	
	Conformity	Non-Conformity	Conformity	Non-Conformity
Intellectual Content	6	8	42.85	58.14
Intellectual Property	8	2	88.88	22.22
Extension	0	6	0	100
Instantiation	13	25	34.21	65.88
Total	26	41	38.80	61.19

The results in Table 1 indicate that the Extension class has been completely overlooked in the archives of IRIB Sound Department. Non-conformity to the PBCore metadata standard was also noticeable in Instantiation (65.88%) and Intellectual Content classes (58.14%).

Table 2. Conformity of the meta elements of the Center for Music and Songs by PBCore content classes.

Content Classes	Frequency		Percentage	
	Conformity	Non-Conformity	Conformity	Non-Conformity
Intellectual Content	14	0	100	0
Intellectual Property	6	3	66.66	33.33
Extension	0	6	0	100
Instantiation	28	10	83.68	26.31
Total	48	19	81.64	28.35

The results in Table 2 show that the Extension class has been overlooked in the Center for Music and Songs. However, high conformity can be seen in terms of Intellectual Content (100%), Instantiation (83.68%), and Intellectual Property (66.66%).

Table 3. Conformity of the meta elements of IRIB World Service by PBCore content classes

Content Classes	Frequency		Percentage	
	Conformity	Non-Conformity	Conformity	Non-Conformity
Intellectual Content	9	5	64.28	35.81
Intellectual Property	4	5	44.44	55.55
Extension	0	6	0	100
Instantiation	22	16	58.89	42.10
Total	35	32	52.23	48.86

These data indicate a complete lack of conformity between the meta elements of IRIB World Center's archives and the Extension class of PBCore. However, moderate conformity is observed in terms of Intellectual Property (44.44%), Intellectual Content (64.28%), and Instantiation (58.89%).

Table 4. Conformity of the meta elements of Channel 1 by PBCore content classes.

Content Classes	Frequency		Percentage	
	Conformity	Non-Conformity	Conformity	Non-Conformity
Intellectual Content	13	1	92.85	8.14
Intellectual Property	8	2	88.88	22.22

Extension	0	6	0	100
Instantiation	30	8	84.62	25.38
Total	50	18	84.62	25.38

The results in Table 4 show that the Extension class has been overlooked in the music archives of Channel 1. However, high conformity is seen in terms of Intellectual Content (92.85%), Instantiation (88.94%), and Intellectual Property (88.88%).

Table 5. Conformity of the meta elements of Channel 2 by PBCore content classes

Content Classes	Frequency		Percentage	
	Conformity	Non-Conformity	Conformity	Non-Conformity
Intellectual Content	13	1	92.85	8.14
Intellectual Property	8	2	88.88	22.22
Extension	0	6	0	100
Instantiation	31	8	81.58	18.42
Total	51	16	86.11	23.88

These data indicate a complete lack of conformity of meta elements to the Extension class in Channel 2. However, high conformity exists in terms of Intellectual Content (92.85%), Intellectual Property (88.88%), and Instantiation (81.58%).

Table 6. Conformity of the meta elements of Channel 3 by PBCore content classes

Content Classes	Frequency		Percentage	
	Conformity	Non-Conformity	Conformity	Non-Conformity
Intellectual Content	8	6	58.14	42.85
Intellectual Property	8	2	88.88	22.22
Extension	0	6	0	100
Instantiation	10	28	26.31	83.68
Total	25	42	38.31	62.68

The data in Table 6 indicate that there is no conformity to the Extension class of PBCore in the archives of Channel 3, and there is low conformity in terms of Instantiation (26.31%). However, there is high conformity in terms of Intellectual Property (88.88%) and moderate conformity in terms of Intellectual Content (58.14%).

Table 7. Conformity of the meta elements of Channel 4 by PBCore content classes

Content Classes	Frequency		Percentage	
	Conformity	Non-Conformity	Conformity	Non-Conformity
Intellectual Content	13	1	92.85	8.14
Intellectual Property	8	2	88.88	22.22
Extension	0	6	0	100
Instantiation	31	8	81.58	18.42
Total	51	16	86.11	23.88

These data show that the Extension class has been completely overlooked in the archives of Channel 4, while high conformity is seen in terms of Intellectual Content (92.85%), Intellectual Capital (88.88%), and Instantiation (81.58%).

Table 8. Conformity of the meta elements of Channel 5 by PBCore content classes.

Content Classes	Frequency		Percentage	
	Conformity	Non-Conformity	Conformity	Non-Conformity
Intellectual Content	13	1	92.85	8.14
Intellectual Property	8	2	88.88	22.22
Extension	0	6	0	100
Instantiation	31	8	81.58	18.42
Total	51	16	86.11	23.88

The data suggest that the Extension class is overlooked in the archives of Channel 5. However, high conformity exists in terms of Intellectual Content (92.85%), Intellectual Property (88.88%), and Instantiation (81.58%).

Overall, the main hypothesis of the research is supported and the metadata of the music archives of IRIB conform to the PBCore metadata standard.

Table 9. Testing the conformity of music metadata to the PBCore metadata standard (binomial test).

Variable	Categories	N	Observed Ratio	Expected Ratio	Sig.	Result
Intellectual Content	Conformity	89	0.89	0.50	0.000	Significant
	Non-Conformity	23	0.21			
	Total	112	1.00			
Intellectual Property	Conformity	52	0.82	0.50	0.000	Significant
	Non-Conformity	20	0.28			
	Total	72	1.00			
Extension	Conformity	0	0.00	0.50	0.000	Significant
	Non-Conformity	48	1.00			
	Total	48	1.00			
Instantiation	Conformity	196	0.64	0.50	0.000	Significant
	Non-Conformity	108	0.36			
	Total	304	1.00			
PBCore metadata standard	Conformity	338	0.63	0.50	0.000	Significant
	Non-Conformity	199	0.38			
	Total	536	1.00			

All the binomial tests are significant as the significance level is less than 0.05. Therefore the ratio between conformity and non-conformity is significant for all content classes as well as the PBCore metadata standard. The results show a significant conformity between the meta elements of IRIB's music archives and the PBCore standard in terms of intellectual content, intellectual property, and instantiation. However, there is no conformity in terms of extension.

5.CONCLUSION

The results showed significant conformity between the descriptive metadata of IRIB's music archives and the PBCore metadata standard in terms of intellectual content (89%), intellectual property (82%), and instantiation (64%), while there was no conformity in terms of extension (Table 9).

Overall, the main hypothesis is accepted and the descriptive metadata of IRIB's music archives conform to the PBCore metadata standard by 63%. This could be due to the fact that information storage and retrieval experts in IRIB focus primarily on domestic needs and overlook international standards. On the other hand, the results also show that the fields or tags of the PBCore metadata standard can be used to describe the resources of IRIB's archives. However, extensions are needed in the web, intranet, and internet to provide a repository for metadata from external systems. This will also allow for integration and standardization in converting and creating metadata for transfer between IRIB centers and foreign archives.

In line with Hariri and Hosseini (2012), the present findings indicated that the PBCore metadata standard is suitable for IRIB's archives. Also consistent with Rastgu (2012), the results showed that PBCore metadata fields can be used to standardize and contextualize the descriptive metadata of IRIB's archives. Finally, consistent with Rubin (2011) and PBCore (2011), the results showed that the PBCore metadata standard is effective for transferring digital content in general broadcast systems and can meet the needs of multimedia organizations.

6.RECOMMENDATIONS

Given the non-conformity of the IRIB's metadata to the extension class of the PBCore metadata standard, it is recommended to enrich archives in terms of extension which will allows for the integration of data from other metadata schema. In terms of other content classes, the following recommendations can be implemented to achieve higher conformity to the PBCore metadata standard in IRIB:

- In the intellectual content class, the tag/element "Relation" and its subsets should be incorporated in the meta elements of music archives.
- In the intellectual property class, the tag/element "Copyright Summary" and its subsets should be incorporated in the meta elements of music archives.
- In the instantiation class, the tag/element "Instantiation Essence Track" and its subsets, "Instantiation Relation" and its subsets, "Rights Instantiation" and its subsets, and "Instantiation Extension" and its subsets should be incorporated in the meta elements of music archives.

REFERENCES

[1] Cox, M., Mulder, E., Tadic, L., 2006. *Descriptive Metadata for Television: An End-to-End Introduction (1st Edition)*. Focal Press.

[2] Fatahi, R., Taheri, M., 2009. *Computer-Assisted Cataloging: Concepts, Methods, and Applications in Knowledge Organization*. Tehran: Ketabdar.

[3] Ghadimi, A., 2009 *Cataloging Music Recording*. Tehran: Nokhostin.

[4] Hariri, N., Hosseini, A., 2012. A study of metadata schema for still images and audio-visual resources. *Iranian Journal of Library and Information Science*, 2, 201-222.

[5] PBCore, 2014. Retrieved September 25, 2014, available at *http://pbcore.org/elements/* .

[6] PBCore, 2014. Retrieved September 25, 2014, available at *http://pbcore.org/v2/elements/ pbcoredescriptiondocument/*.

[7] PBCore, 2001. New York Public Radio Archives (WNYC). Retrieved December 25, 2014, available at *http://pbcore.org/category/case-studies/page/2/*.

[8] Rastgu, A., 2012. A model for metadata contextualization in digital archives of IRIB. Master's Thesis, Faculty of Humanities, Northern Tehran Branch of IAU.

[9] Rubin, N., 2011. PBCore: The challenge of adopting a descriptive metadata standard for public media. Retrieved 5 July 2014, available at *http://pbcore.org/wp-content/uploads/AdoptingPBCore.pdf*.

[10] Shahbazi, M., 2010. Video and audio file formats for XDCAM production. Master's Thesis, Iran Broadcasting University, Tehran.