DESIGN OF METADATA SCHEMA FOR COLLECTING, UTILIZING AND PRESERVING BIOMECHANICAL DATA OF WINTER SPORTS EXPERIMENT

Hye-young Kim and Yu-hyun Ra
Korea National Sport University (KNSU), Seoul, Republic of Korea

The purpose of this study was to present metadata elements for managing biomechanical data of winter sports experiments and to establish an Integrated Data Management System with the presented metadata elements. To sort and select metadata elements, metadata crosswalk was done with four internationally used metadata schema; Dublin Core, OECD, DataCite, MODS. And a new metadata schema was designed with common elements from the crosswalk and the metadata elements drawn from general experimental designs and R&D portal. The schema was applied to a research data platform *Datanest* and the Integrated Management System for experimental data of winter sports was established.

KEY WORDS: winter sports, database, metadata, schema

INTRODUCTION: Usage of data in the field of sports is expanding nowadays and researches on sports data are vitalized. In Korea, especially, winter sports analytics are encouraged by PyeongChang 2018 Winter Olympics. Accordingly, amount of sports data are generated and collected. For instance, data collected by KNSU for winter sports analytics are motion capture data of 3-dimensional motion analysis system (60Hz), EMG signal data (3000Hz), foot pressure data (100Hz), motion data of IMU sensor (200Hz), Bio-signal data of wearable devices (20Hz), video records (30Hz), physical information and the questionnaire data of the experimental target and more. Thus, to manage collected data, this study presented a metadata schema for the winter sports experimental data and applied the schema to a data platform which can be easily accessed by both researchers and sports practitioners.

In Korea, research on the integrated management of data collection was raised in the field of science and extended to the field of sports. Kim, S. et al. (2010) emphasized the need of a study regarding the development of the integrated metadata schema, which can ensure interoperability for heterogeneous scientific data, in order to manage and utilize scientific data. Afterwards, in the field of sports, Cho, J. (2012) pointed out that yet there are no practical Sports Big Data infrastructure or application examples that can provide sufficient evidence for the effectiveness and the strategic value of Big Data although the utilization of Big Data is being highlighted in sports and health sector. Park, S. and Lee, J. (2013) insisted that sports public data should also strengthen its role in offering a chance to accumulate new intellectual property. Jang, J and Kim, S. (2014) built multiple entity model by developing metadata elements for four entities which are sports match, match contributors, moving records, and record management business.

METHODS: 1) Among globally used metadata standards, Dublin Core, schema of OECD, schema of DataCite, and MODS were selected for metadata crosswalk. Metadata elements of those standards were translated from one standard to other standards. Then common elements were extracted from the crosswalk. 2) Metadata elements for experiments and researches were drawn from general experimental designs and online portal of National Science & Technology Information Service (NTIS). NTIS is the national R&D information portal providing government-funded R&D information on topics since 2002. 3) A new metadata schema was designed with common elements and the metadata elements drawn from general experimental designs and NTIS. 4) Schema designed in this study was applied to a research data platform *Datanest*. *Datanest* is a scientific data sharing platform developed by Korea Institute of Science and Technology Information (KISTI)**.** This platform was selected and used for this research since its usage was consistent with the final purpose of designing metadata schema for winter sports experimental data, which is establishing a data platform of winter sports data that can be easily accessed by both researchers and sports practitioners.

1. Dublin Core: A set of metadata elements standardized by ISO15836 which features

simplicity, scalability, and syntax independence.
2. Metadata Set of OECD: A set of metadata developed by Organization for Economic Cooperation and Development to quote data sets and tables by essential metadata elements Dublin Core.
3. Schema of DataCite: A metadata schema used in DataCite, an international consortium for the description of the scientific data.
4. MODS: Metadata object description schema which can describe variety of information, including public data recorded on websites and electronic media.

RESULTS:

Table 1
Metadata Crosswalks

Dublin Core	OECD	DataCite	MODS
Title	Main title, Subtitle	Title	title
Creator	Author(s)	Creator	namePart
Subject	Classification	Subject	subject, classification
Description	Keyword(s), abstract	Description	abstract, note
Publisher	Is copyrighted by	Publisher	publisher
Contributor	Is edited by	Contributor	namepart
Date	Publication date	Date	dateCreated/Other
Type	-	ResourceType	typeOfResource
Format	Size, physical form	Size, Format	internetMediaType
Identifier	Identifier	Identifier	identifier
Source	-	-	Original relatedItem
Language	Language(s)	Language	language
Relation	Link(s)	relationType	relatedItem
Coverage	Period, Countries	Start/EndDate	temporal, geographic
Rights	-	Rights	accessCondition
-	Periodicity	-	-
-	Update method	-	-
-	Variable index	-	-
-	-	nameIdentifier	-
-	-	Version	edition
-	-	-	location

As presented in Table 1, common elements of all four metadata sets are Title, Creator, Subject, Description, Publisher, Contributor, Date, Format, Identifier, Language, Relation, and Coverage. Common elements are fundamental for identifying the resource data. Thus, all the elements except the Creator and Format, which are automatically collected in *Datanest*, were included in metadata schema for winter sports biomechanical data. Among partly common elements, elements of Dublin Core (Type, Source, Rights) were included in the schema design but obligation level was fixed in optional level since their influence on the experiment and collected data is insignificant.

In Table 2, metadata elements needed for utilizing experimental data such as sports event, experimental purpose, prime factor measured in the experiment, experimental time, weather, location are added in the schema design based on general experimental plan research. In addition, metadata elements for experimental field, experimental target, device operator which affects the raw data are added in the design. Lastly, metadata about device and government department and supporting institution drawn from NTIS are added.

In Table 2, 'Occ.' stands for occurrence, cardinality and quantity constraints for the elements. Occurrence level '0-n' means optional and repeatable, '0-1' means optional, but not repeatable, '1-n' means required and repeatable, and '1' means required, but not repeatable. 'Obl.' stands for obligation. Obligation level 'M' means mandatory, 'R' means strongly recommended, 'O' means optional.

Table 2
Metadata Schema for Experimental Data of Winter Sports

ID	Element	Occ.	Obl.	ID	Element	Occ.	Obl.
1	title	1	M	8.4	devCompany	1	M
1.1	mainTitle	1	M	8.5	devPossAgency	0-1	O
1.2	eqvTitle	0-1	R	8.6	nfec	0-1	O
2	description	0-1	O	9	devOperator	1-n	M
3	sportsEvent	1	M	9.1	operatorName	1	M
3.1	evntName	1	M	9.2	operatorInst	1	M
3.2	evntType	1	M	9.3	operatorRole	1	M
4	exptBasic	1	M	9.4	operatorMbox	0-1	O
4.1	exptDate	1	M	10	exptContributor	0-n	R
4.2	exptTime	1-n	M	10.1	contributorName	0-1	R
4.3	exptPurpose	0-1	R	10.2	contributorInst	0-1	R
4.4	exptFactor	1	M	10.3	contributorRole	0-1	R
4.5	exptMethod	1	M	10.4	contributorMbox	0-1	O
5	weatherInfo	1	R	11	dataType	0-n	O
5.1	exptTemp	1	R	12	coverage	0-1	O
5.2	exptHumid	1	R	12.1	temporal	0-1	O
6	locationInfo	1	M	12.2	geographical	0-1	O
6.1	exptLocation	1	M	13	language	0-1	R
6.2	exptLatitude	0-1	R	14	identifier	0-1	R
6.3	exptLongitude	0-1	R	15	publisher	0-1	O
6.4	exptField	0-1	R	16	rights	0-1	O
7	exptTarget	1	M	17	source	0-1	O
7.1	targetClass	1	M	18	relatedItem	0-1	R
7.2	targetName	1	M	19	govDept	0-1	R
7.3	targetGender	0-1	R	20	support	0-1	R
7.4	targetBirth	0-1	R	21	projectInfo	1	M
8	exptDevice	1-n	M	21.1	projectName	1	M
8.1	devClass	0-1	R	21.2	projectDetail	1	M
8.2	devName	1	M	21.3	responsibleP	1	M
8.3	devModel	0-1	R	21.4	projectPeriod	1	M

Finally, the schema shown in Table 2 was encoded in data platform, and the data management system for winter sports experimental data was established. Extensible Markup Language was used for encoding. Encoding process of schema is partly shown in Figure 1.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="WinterSports_expt">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="title" maxOccurs="1" minOccurs="0">
          <xs:complexType>
            <xs:sequence>
              <xs:element name="mainTitle" maxOccurs="1" minOccurs="1" type="xs:normalizedString"/>
              <xs:element name="eqvTitle" maxOccurs="1" minOccurs="0" type="xs:normalizedString"/>
            </xs:sequence>
          </xs:complexType>
        </xs:element>
                .
                .
                .
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Figure 1: Applying designed metadata schema

DISCUSSION: The metadata schema for collecting, utilizing and preserving winter sports biomechanical data contributes to more precise research analytics and coaching plan. Firstly, Information of experimental environment are collected by 'exptBasic', 'weatherInfo', 'locationInfo' set in designed schema. These element sets let researchers and coaches to retrieve basic information about the biomechanical experiment. Secondly, metadata about experimental target, devices, and device operator, which could influence the raw data, are assured to be preserved at the data platform in order to help utilization of the collected data.

Also, this schema make data searching process easier for researchers and coaches involved in the winter sports experiment. The searcher can use any keyword to find certain data set. For instance, by searching experimental device's name, the searcher can find all the data measured by the searched device. This applies to sports event, location, target's name, contributor, and so on. In addition, 'relatedItem' helps the searcher to access to related data sets when they find one necessary data set for their research.

A clear definition of the details of the experiment makes the desired statistical analyses possible, and almost always improves the usefulness of the results, SAS (2005). Accordingly, designing metadata schema and establishing the Integrated Management System for biomechanical data of winter sports experiments could result in promotion of winter sports analytics. However, development of data management system for sports is yet in early stages. Therefore, validity test targeting researchers and coaches will be held before expending the usage of the data management system developed in this study.

CONCLUSION: Throughout this study, metadata schema for management of the experimental data generated and collected by the winter sports biomechanical research was presented and applied to data platform. The metadata elements suggested in this study could assist researchers and coaches involved in the winter sports experiment by helping them retrieve basic information about the biomechanical experiment data with easier data searching process. Furthermore, the establishment of the efficient database assures systematic management of research results and collaborative research through data sharing. Thus brings out progress in analytical studies and coaching skills. Moreover, reasonable outcome of sports data analysis is directly linked to the performance improvement of sports practice and the development of sports industry. Analytics can make a difference in scoring touchdowns, improving wearable technology, or preventing injuries. Therefore, further research about sports data integration management and expansion of data collection is required for the development of the overall sports industry.

REFERENCES:
Bongard, V., McDermott, A.Y., Dallal, G.E. Schaefer, E.J. (2007), Effects of Age and Gender on Physical Performance, *AGE*, 29, 77-85.
Cho, J. (2012), Utilization and Prospect of Sport Big Data, *Journal of Measurement and Evaluation in Physical Education Studies*, 14, 1-11.
Green, T. (2009), We Need Publishing Standards for Datasets and Data Tables, OECD Publishing White Paper, 3-29.
Jang, J., Kim, S. (2014), A Study on the Development of a Metadata Schema for Sports Moving Records, *Journal of Korean Biblio Society for Library and Information Science*, 25, 29-57.
Kim, S., Han, S., Lee, T., Kim, Y. (2010), A Study on a Model for Using and Preserving Scientific Data, *Journal of Korean Biblio Society for Library and Information Science*, 21, 81-93.
Park, S., Lee, J. (2013), The Future Utility of Sports Public Big Data and the Plan for Its Further Development, *Journal of Sport and Leisure Studies*, 54, 539-546.
SAS (2005), Concepts of Experimental Design, A SAS White Paper, 1-28.
Starr, J., Ammann, N., Ashton, J., Barton, A., Elliott, J., et al. (2015), DataCite Metadata Schema for the Publication and Citation of Research Data, *DataCite Metadata Kernel*, v3.1. 1-38.