



Utilizing Real-World Data for Sports Data Collection with CDISC Standards

Hyeon-moon, Kim
PD(Program Director)
Ministry of Culture, Sports and Tourism
Korea Creative Contents Agency



Meet the Speaker

Hyun-Moon, Kim

Title: PD(R&D Program Director)

Organization: MCST-KOCCA

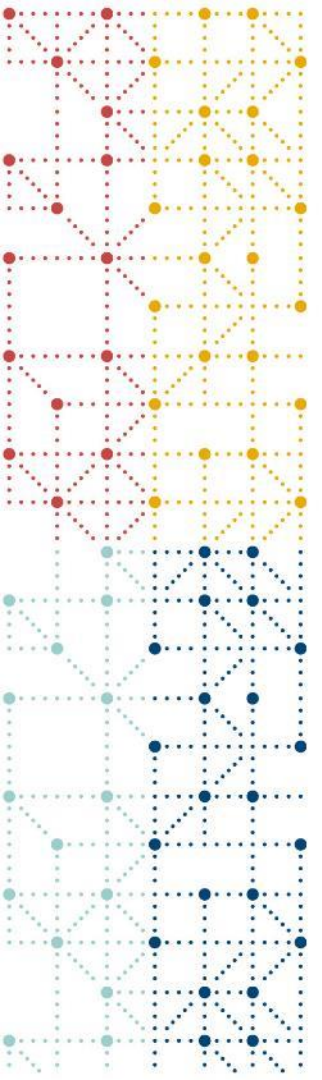
Kim Hyun-moon is the R&D program director of the Ministry of Culture, Sports and Tourism. In particular, he is in charge of sports technology and is also in charge of planning medical technologies related to digital therapeutics in connection with content.

As a culture technology PD, he is planning and working on various projects related to the intelligence and precision of sports R&D based on data. In addition, the digitization of sports mileage is being promoted through the analysis of exercise amount and exercise effectiveness based on standard data such as CDISC and Meta Data..



Agenda

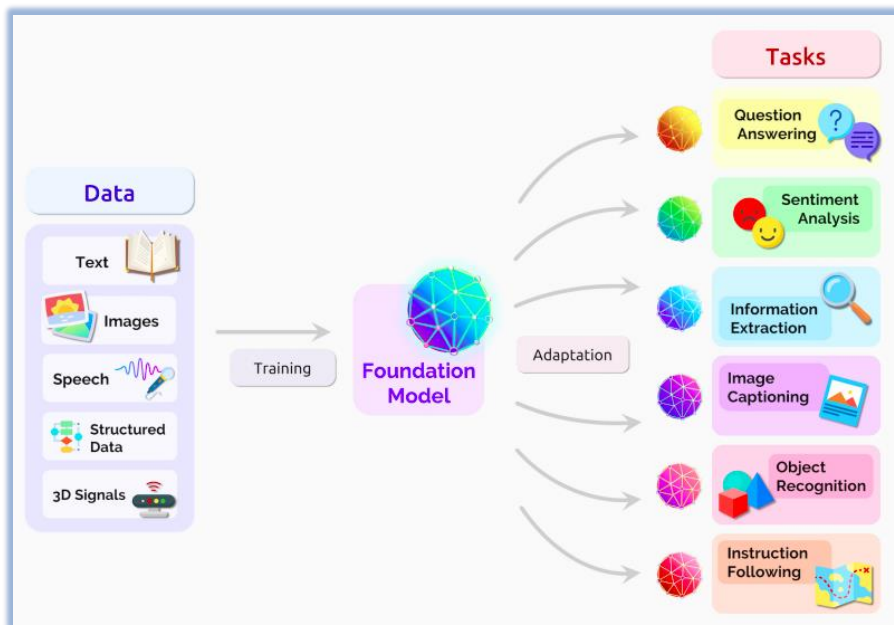
1. LM(Language Model) Background
2. CDISC & Sports big data Strategy of the MCST for Promoting Sports health care & data industry



LM(Language Model) Background

LLM(Large Language Model)

Foundation Model (ex; ChatGPT)-based generation service AI



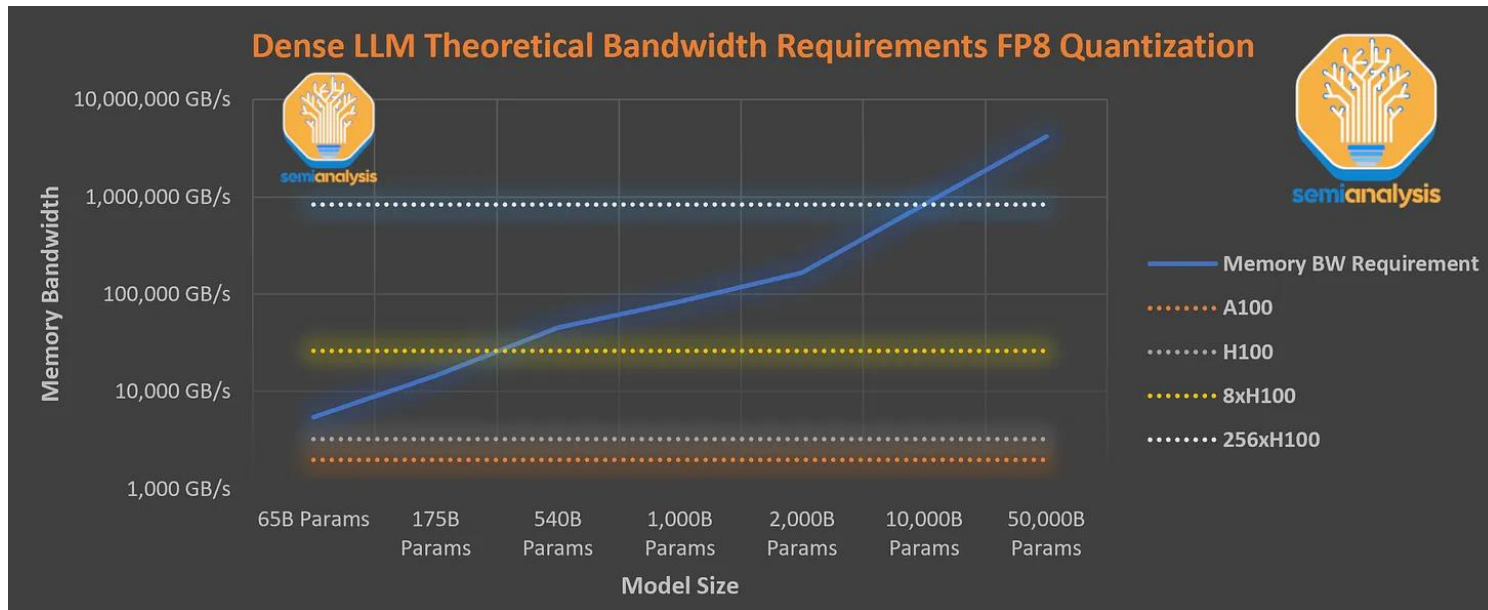
Generative AI Wave

- How to use AI in the industry
- For field-specific purpose data
- Efficiency and individual data-driven learning

LLM(Large Language Model)

Why ChatGPT rely on the Mixture of Experts architecture

→ frontier models require trillions of parameters and require GPUs for inference



LM(Language Model)

Specialized Generative AI

Horizontal AI

Accounting 


Customer service 


Marketing 

Sales development 

Sales 
Sales GPT

Vertical AI


Entertainment localization 


Entertainment VFX 

Healthcare 

Pharma 

Common AI

Automation & integration 

Enterprise search 

Notes: Inclusion criteria: Public claim to have added generative AI features in this specific domain. Not comprehensive.

AIMultiple⁴

VFLM(Vertical Function Language Model)

Hippocratic AI vs. GPT-4 Performance on Major Healthcare Categories

	Number of Tests	Existing LLMS			Hippocratic AI	
		Commercial LLM #1	Commercial LLM #2	GPT-4	Ours	Improvement vs Best Competitor
Pharmacist	4	52.5%	32.3%	62.6%	76.5%	13.9%
Dentist	3	73.6%	59.9%	78.8%	91.9%	13.1%
Nurse	18	54.6%	36.1%	73.6%	80.9%	7.3%
Physician	23	47.8%	33.4%	76.4%	83.1%	6.7%
Medical Coder	6	49.9%	31.5%	57.3%	68.2%	10.9%
Health Equity	6	77.8%	73.5%	84.9%	87.0%	2.1%
Ancillary	22	59.8%	39.3%	75.8%	85.0%	9.2%
Compliance	10	59.3%	46.1%	81.6%	91.2%	9.6%
Administrative	14	64.3%	44.2%	76.0%	86.0%	10.0%
	106	60.0%	44.1%	74.1%	83.3%	



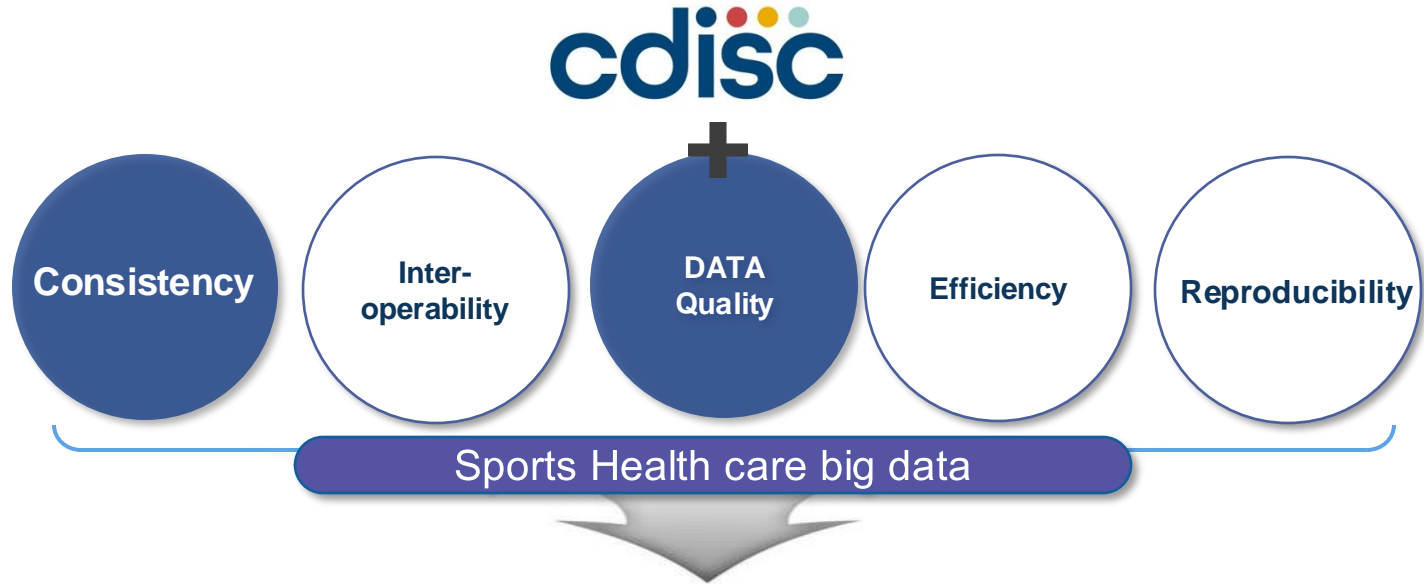
Yes!! VFLM(Vertical Function Language Model) No!!! LLM(Large Language Model)

VFLM, not LLM, for explainable AI services
Standard Data model is the foundation of the VFLM AI service

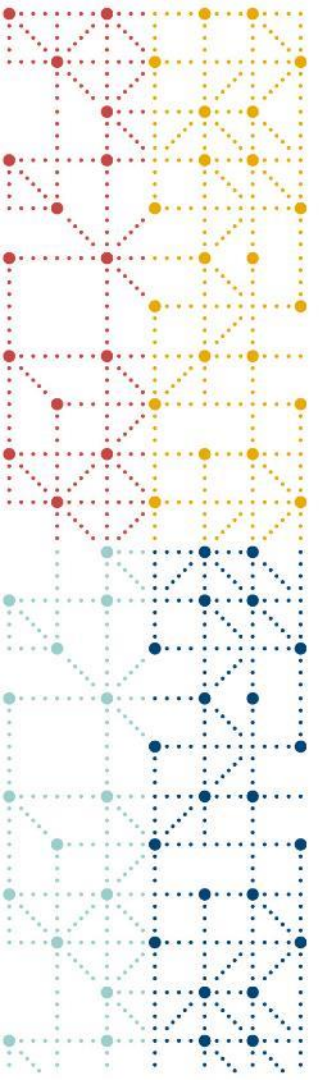
- **Consistency:** Standards ensure that the data used to train the model is consistent. This is crucial because inconsistent data can lead to a model that performs poorly or is biased.
- **Interoperability:** Standards allow different datasets to be combined and used together. This is particularly important for large language models, which often require vast amounts of data from various sources.
- **Quality:** Standards help maintain the quality of the data. High-quality data is essential for training a model that is accurate and reliable.
- **Efficiency:** Standards can make the process of collecting and preparing data more efficient. This can save time and resources, which is especially important when dealing with large amounts of data.
- **Reproducibility:** Standards make it easier for others to reproduce the results of the model. This is important for verifying the model's performance and for further research and development.

Sports big data for sports industry 4.0

CDISC standards are needed to build sports health big data and LLM



Sports Industry 4.0



CDISC & Sports big data Strategy: Promoting Sports health care industry



Sports health care

Sports Health Care, associated with sports medicine, is a field that focuses on maintaining and improving health through sports and physical activity.

- **Prevention and Treatment of Injuries:** Sports Health Care professionals work to prevent and treat sports-related injuries. This includes assessment, diagnosis, treatment planning and implementation, surgery, and post-operative care.
- **Physical Fitness:** Sports medicine deals with physical fitness and the treatment and prevention of injuries related to sports and exercise. It's a distinct field of health care that emerged in the late 20th century.
- **Health Promotion:** The World Health Organization (WHO) has a Sport for Health Programme that promotes participation in sports and works with the sports community to advance health for all. The program aims to raise awareness and stimulate sports environments to promote health and well-being.
- **Physical Activity:** Regular physical activity helps prevent and treat noncommunicable diseases (NCDs) such as heart disease, stroke, diabetes, and breast and colon cancer. It also helps prevent hypertension, overweight, obesity, and can improve mental health, quality of life, and well-being.

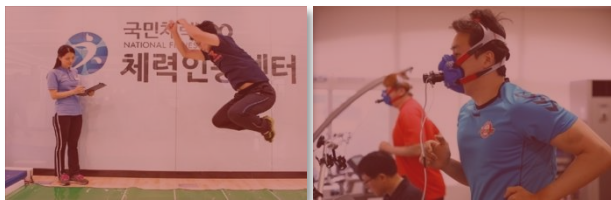
Sports health care & data resources

Sports data is being collected through major institutions..

analog collection → useless data

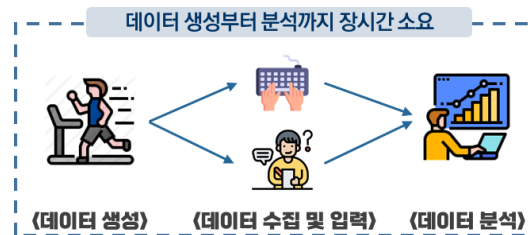
collecting and building a physical fitness DB for athletes and the public

수집 대상	수집 데이터	수집 방식	진 단
산 업	✓ 기초 정보 (기업현황, 경기 정보 등)	리서치 조사	✓ 단순 통계 데이터
국 민	✓ 기초체력 정보	수기 수집	✓ 데이터 수집 취약 ✓ 디지털화 수준 매우 낮음
엘 리 트 (+ 지역 체육인)	✓ 전문체력 정보 (경기력 관련)	아날로그 (수기 방식)	✓ 데이터 기술 외산 의존 ↑ ✓ 데이터 활용방안 제한적



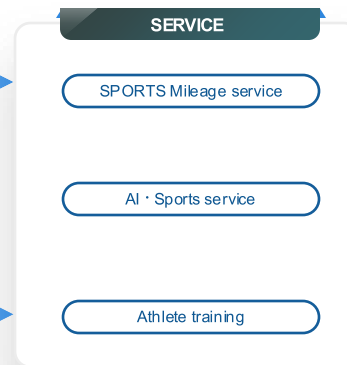
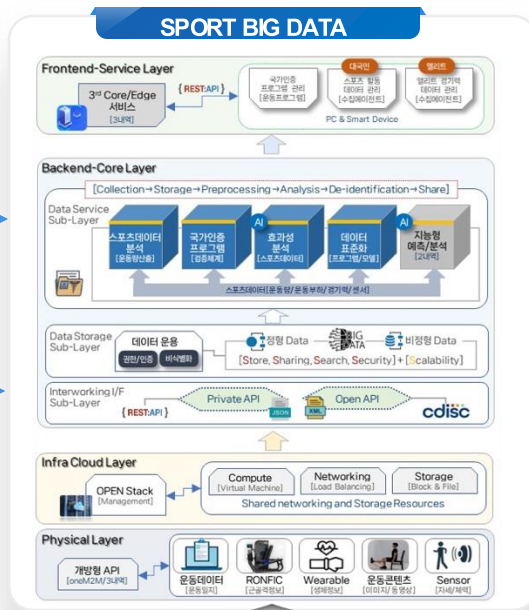
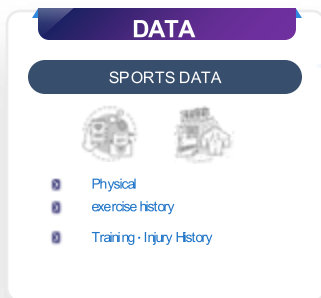
Low level of digitization, Closed data

▶ Collect data by handwriting



▶ Low data usage

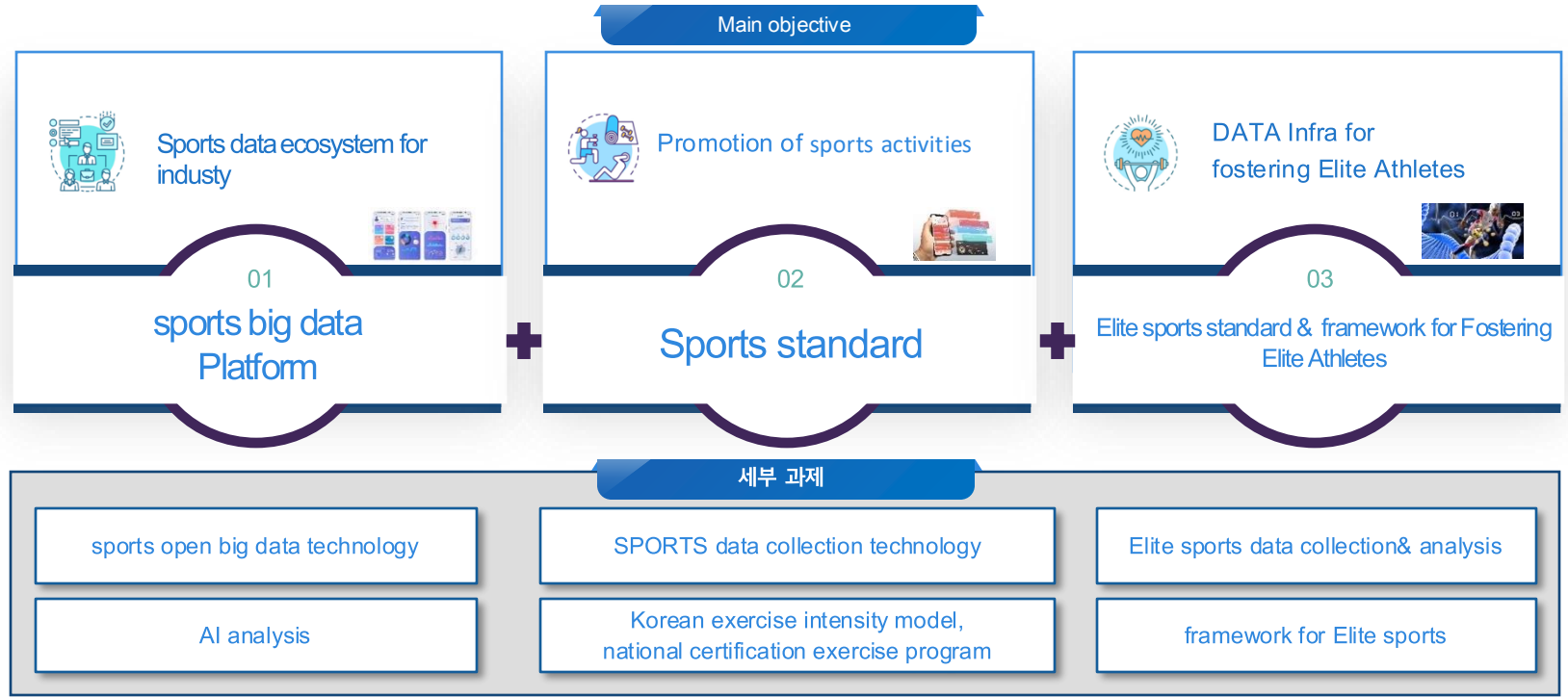
Sports big data concept



data standardization

evaluation, testing, assembly, and use of the equipments, facilities, and protective gears utilized in sports and recreational activities.

Sports big data concept



Sports data standard concept

SPORTS
BIG DATA

data standardization



Therapeutic Area Data Standards for Type 1 Diabetes - Exercise and Nutrition Modules Version v1.0 (Provisional)

Prepared by the
CDISC Type 1 Diabetes Standards Development Team

Notes to Readers

- This is the Provisional Version 1.0 of the Therapeutic Area Data Standards for Type 1 Diabetes - Exercise and Nutrition Modules.
- This document is based on CDASHIG v2.1, CDASH Model v1.1, SDTM v1.7, SDTM Implementation Guides (SDTMIG v3.3 and SDTMIG-MD v1.1), and Define-XML v2.1.

CDISC SDTM Controlled Terminology

CDISC SDTM Controlled Terminology, 2023-05-25

ions

CDISC Submission Name	Test Code	CDISC Synonym	CDISC Definition
15M01001 - 15M01001 Functional Test Code	C141007	15-Minute Walk/Run Functional Test	15-Minute Walk/Run test code
15M01001 - 15M01001 Functional Test Code	C141008	15-Minute Walk/Run Functional Test	15-Minute Walk/Run - Walk the 15-meter walk/run path
15M01001 - 15M01001 Functional Test Code	C141009	15-Minute Walk/Run Functional Test	15-Minute Walk/Run - If you, time to walk or run 15 meters
15M01001 - 15M01001 Functional Test Code	C141010	15-Minute Walk/Run Functional Test	15-Minute Walk/Run - If you, did subject wear glasses?
15M01001 - 15M01001 Functional Test Code	C141011	15-Minute Walk/Run Functional Test	15-Minute Walk/Run - Test grade
15M01001 - 15M01001 Functional Test Code	C141012	15-Minute Walk/Run Functional Test	15-Minute Walk/Run test name
15M01001 - 15M01001 Functional Test Code	C141013	15-Minute Walk/Run Functional Test	15-Minute Walk/Run - Test grade
15M01001 - 15M01001 Functional Test Code	C141014	15-Minute Walk/Run Functional Test	15-Minute Walk/Run - If you, time to walk or run 15 meters
15M01001 - 15M01001 Functional Test Code	C141015	15-Minute Walk/Run Functional Test	15-Minute Walk/Run - Walk the 15-meter walk/run path
15M01001 - 15M01001 Functional Test Code	C141016	15-Minute Walk/Run Functional Test	15-Minute Walk/Run - If you, did subject wear glasses?
15M01001 - 15M01001 Functional Test Code	C141017	15-Minute Walk/Run Functional Test	15-Minute Walk/Run - Test grade
15M01001 - 15M01001 Functional Test Code	C141018	15-Minute Walk/Run Functional Test	15-Minute Walk/Run test name



ACSM'S

Guidelines for Exercise Testing and Prescription

Eleventh Edition



CDISC Exercise module



Therapeutic Area Data Standards for Type 1 Diabetes - Exercise and Nutrition Modules Version v1.0 (Provisional)

Prepared by the
CDISC Type 1 Diabetes Standards Development Team

Notes to Readers

- This is the Provisional Version 1.0 of the Therapeutic Area Data Standards for Type 1 Diabetes - Exercise and Nutrition Modules.
- This document is based on CDASH v2.1, CDASH Model v1.1, SDTM v1.7, SDTM Implementation Guides (SDTMIG v1.3 and SDTMIG-MD v1.1), and Defining-AMT v2.1.

Revision History

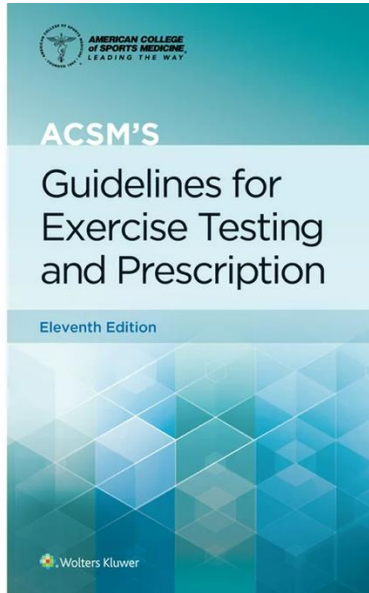
Date	Version
2021-06-10	1.0 Provisional

See [Appendix B](#) for Representations and Warranties, Limitations of Liability, and Disclaimers.

- CDISC Therapeutic Area Data Standards for Type 1 Diabetes - Exercise and Nutrition Modules include information on the level of exercise performed by the subject, such as the type of exercise (e.g., cardio, interval, strength), frequency of exercise, and duration of a single exercise session
- In addition, the guide provides information on how to represent data obtained from wearable devices measuring exercise parameters. This may include data on the intensity of exercise, which can be expressed in absolute or relative terms



ACSM Guideline



- The American College of Sports Medicine (ACSM) is a professional organization that promotes and integrates scientific research, education, and practical applications of sports medicine and exercise science to maintain and enhance physical performance, fitness, health, and quality of life.
- ACSM's Guidelines for Exercise Testing and Prescription delivers scientifically based standards on exercise testing and prescription to the certification candidate, the professional, and the student. This manual gives succinct summaries of recommended procedures for exercise testing and exercise prescription in healthy and diseased patient



CDISC & ACSM for sports data standard

- ACSM provides guidelines for exercise prescription and testing, CDISC provides a standardized framework for collecting and managing data related to exercise and nutrition in clinical trials.
- ACSM & CDISC help to improve the efficiency and accuracy of Sports data collection & tabulation.
- ACSM & CDISC is important roles in promoting and advancing the **field of sports medicine and exercise data science.**

- We ensure accurate and reliable sports data collection in National R&D Programs: Community-Based Rehabilitation

ACSM & CDISC

ex) vo2max

	검사 RE-RETESTCD	폐기능_1
VO2MAX		
FVC		
FEV1		
IC		
FICO2		
FI02		
AT		
OXYPULSE		

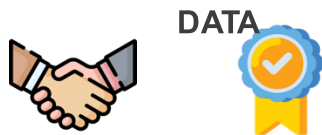
VO2 Max Direct	
VO2 Max Test Name PRTRT <small>Hidden/pre-populated</small>	GRADED EXERCISE TEST
Indicate whether the VO2 max test was performed. REPERF <small>If "No" then RESTAT = "NOT DONE" where RETESTCD = "VO2MAX" if "Yes" then Not Submitted.</small>	<input type="radio"/> Yes <input type="radio"/> No <small><From NY codelist></small>
Record date of measurements using this format (DD-MON-YYYY). REDAT REDTC PRSTDTCD	<input type="text"/>
Record time of measurement (as complete as possible). RETIM REDTC PRSTDTCD	<input type="text"/>
Record the kind of test. RETSTCND NSRE.RETSTCND	<input type="radio"/> Exercise to submaximal effort <input type="radio"/> Exercise to exhaustion
Record the device used for the VO2 max exercise test. EXER_SPDEVID PR.SPDEVID <small>DI.SPDEVID and DI.DIVAL where DIPARMCD = "DEVTYPE"</small>	<input type="radio"/> Treadmill <input type="radio"/> Cycle ergometer <input type="radio"/> Rowing machine
Record the device used for the maximal oxygen consumption measurement/exchange. VO2EXC_SPDEVID RE.SPDEVID <small>DI.SPDEVID and DI.DIVAL where DIPARMCD = "DEVTYPE"</small>	<input type="radio"/> Ergospirometer <input type="radio"/> Douglas bag <input type="radio"/> Tissot tank
Record test result. VO2MAX_REORRESU <small>REORRESU where RETESTCD = "VO2MAX"</small> <small>Hidden/pre-populated</small>	<input type="text"/> Unit mL/kg/min <small><From UNIT codelist></small>
Standardized Test Criteria Name VO2MAX_RESTCRNM NSRE.RESTCRNM <small>Hidden/pre-populated</small>	ACSM'S GUIDELINES FOR EXERCISE TESTING AND PRESCRIPTION (10TH EDITION)
Indicate whether the standardized test criteria were met for the VO2 max test. VO2MAX_RESCMIND NSRE.RESCMIND	<input type="radio"/> Yes <input type="radio"/> No <small><From NY codelist></small>

Strategies of Sports big data & standard R&D

Pre R&D

Cross-team Collaboration & planning

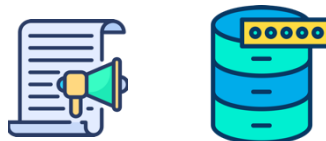
- ✓ Create data standards requires collaboration across different teams (sports, medical & data science)
- ✓ define the fields to collect, their relationships, data types, validation rules, and how these map to other system



R&D

Documentation Development & Implementation

- ✓ create technical documents such as schemas, specifications, models, and vocabularies
- ✓ implemented, ensuring data integrity and interoperability
- ✓ consider the standards of the systems and tech you integrate with

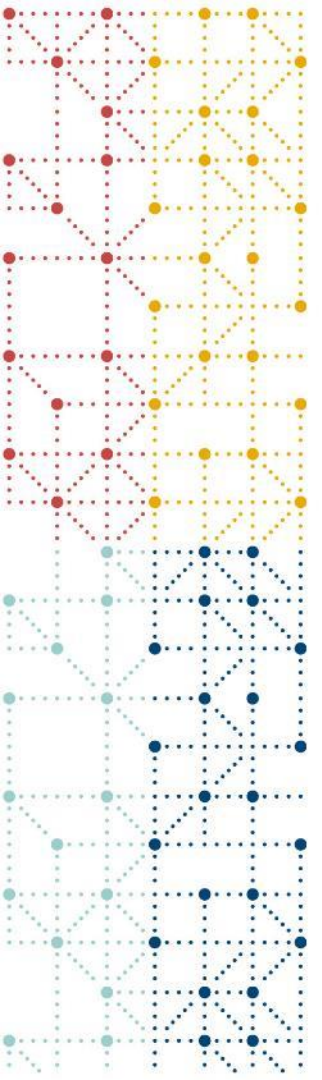


SERVICE R&D

Review & service test

- ✓ review and update to ensure they remain relevant and effective
- ✓ Service on test bed (such as AI, data API service)





Thank you!

cdisc

Clear data. Clear impact.