



HAL
open science

Development of IPSJ Data Science Curriculum Standard

Tetsuro Kakeshita, Kazuo Ishii, Yoshiharu Ishikawa, Hitoshi Matsubara,
Yutaka Matsuo, Tsuyoshi Murata, Miyuki Nakano, Takako Nakatani,
Haruhiko Okumura, Naoko Takahashi, et al.

► **To cite this version:**

Tetsuro Kakeshita, Kazuo Ishii, Yoshiharu Ishikawa, Hitoshi Matsubara, Yutaka Matsuo, et al.. Development of IPSJ Data Science Curriculum Standard. Open Conference on Computers in Education (OCCE), Aug 2021, Tampere, Finland. pp.156-167, 10.1007/978-3-030-97986-7_13 . hal-04161448

HAL Id: hal-04161448

<https://inria.hal.science/hal-04161448v1>

Submitted on 1 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Development of IPSJ Data Science Curriculum Standard

Tetsuro Kakeshita¹, Kazuo Ishii², Yoshiharu Ishikawa³, Hitoshi Matsubara⁴, Yutaka Matsuo⁴, Tsuyoshi Murata⁵, Miyuki Nakano⁶, Takako Nakatani⁷, Haruhiko Okumura⁸, Naoko Takahashi⁹, Norimitsu Takahashi¹⁰, Gyo Uchida⁴, Eriko Uematsu¹¹, Satoshi Saeiki¹², and Hiroshi Kato⁷

¹ Saga University, Saga, Japan

² Suwa University of Science, Nagano, Japan

³ Nagoya University, Aichi, Japan

⁴ University of Tokyo, Tokyo, Japan

⁵ Tokyo Institute of Technology, Tokyo, Japan

⁶ Tsuda University, Tokyo, Japan

⁷ The Open University of Japan, Chiba, Japan

⁸ Mie University, Mie, Japan

⁹ Kokugakuin University, Tokyo, Japan

¹⁰ Digital Growth Academia, Inc., Tokyo, Japan

¹¹ Musashino Gakuin University, Saitama, Japan

¹² The Japan Data Scientist Society, Tokyo, Japan

kake@is.saga-u.ac.jp

Abstract. The Information Processing Society of Japan (IPSJ) published a curriculum standard for university-level education majoring in data science (DS)¹. In this paper, we shall report strategy and development of the DS curriculum standards. Data science education and the development of data scientists are recognised to be quite important in both social and business contexts. The IPSJ DS curriculum standard is developed by integrating various related initiatives and has the following unique features. (1) The DS curriculum standard covers a wide range of related fields to ensure international compatibility through mapping to the ACM Data Science curriculum and the European EDISON Data Science Framework. (2) It collaborates with the IPSJ Data Scientist certification (under development) by referring to the Data Scientist skill checklist (assistant level) developed by the Data Scientist Society of Japan. (3) It clarifies the knowledge and skills required of students majoring in data science. (4) It assigns a time to each educational content so that the curriculum size becomes approximately 675 class hours (60 credits in the Japanese credit system). (5) It collaborates with the Model Curriculum for Mathematics, Data Science and AI Education (literacy level) developed by the Japan Inter-University Consortium for Mathematics and Data Science, supported by the Japanese government.

¹ An English translation of the IPSJ Data Science Curriculum Standard is available at: https://www.ipthree.org/wp-content/uploads/IPSJ-DS-Curriculum_202104_en.pdf

Keywords: Data science, curriculum development, statistics, artificial intelligence, computer science, data management, software engineering, data scientist certification

1 Introduction

Various initiatives, such as [1] in Japan, are promoted around the world to develop curriculum standards and professional certification schemes in the field of data science (DS). In this context, the Information Processing Society of Japan (IPSJ) intends to contribute to the development of systematic data science education from the standpoint of both the academic and professional community of computing discipline in Japan.

IPSJ has developed curriculum standards for computing education at the university level for nearly 30 years and has promoted a variety of activities related to computing education at the primary, secondary, high school, and university levels. The J17 curriculum standard [2] was supported by the Ministry of Education of Japan (MEXT). But as of FY2017, IPSJ could not publish a curriculum standard for the DS domain.

However, as of the end of FY2019, various efforts are in progress on curriculum standards in the DS domain in the United States, Japan, and Europe so that standardisation to ensure international compatibility is coming into view. Furthermore, IPSJ is working on the development of a data scientist (DS) certification scheme. IPSJ recently published its strategy for data science and intends to create collaboration between the DS certification and the DS curriculum [3].

Observing the above situation, IPSJ developed a curriculum standard for college students majoring in data science from the viewpoint of computing discipline. Although the Japan Inter-University Consortium for Mathematics & Data Science [4] published two model curriculums [5] in the data science field, the model curriculums do not cover expert level. Compared with other efforts in DS curriculum development, there are significant differences between the IPSJ DS curriculum and other DS curriculums. The major differences are listed below.

- (1) IPSJ DS curriculum covers a wide range of related fields to ensure international compatibility by integrating the ACM Data Science curriculum [6] and the European EDISON Data Science Framework [7].
- (2) IPSJ DS curriculum is in collaboration with the IPSJ Data Scientist certification (under development) by referring to the Data Scientist (DS) skill checklist [8] (assistant level) developed by the Japan Data Scientist Society.
- (3) IPSJ DS curriculum assigns time to the educational contents so that the curriculum size becomes approximately 675 class hours (60 credits in the Japanese credit system).
- (4) IPSJ DS curriculum regards the model curriculum as a prerequisite since the model curriculum is designed to teach data science literacy to all college students and its size is much smaller than the IPSJ DS curriculum.

IPSJ DS curriculum is intended to be used as a reference by universities and undergraduate level professional education programmes majoring in data science.

This paper is organised as follows. We shall overview the representative work related to our DS curriculum standard in Section 2. EDISON Data Science Framework, ACM Data Science Curricula, and Data Scientist Skill Checklist will be introduced. The development policy of the IPSJ DS curriculum will be explained in Section 3. We

have three major policies. The first one is the integration of classification criteria of the related BOK and skill set. The second one is the assignment of time to each educational topic. The third one is the adjustment of the contents considering the allocated time restriction. Our effort related to these policies will be explained in Sections 4 to 6 respectively. In the last section, we shall conclude the paper with an explanation of our vision to implement systematic development and evaluation system of the data scientists of various levels.

2 Related Works

Many efforts are in progress in the DS field. Many DS programmes and seminars are provided by many universities and companies. Target students/engineers and levels are quite diverse among these programmes and seminars. Development of curriculum and body of knowledge (BOK) is essential in the DS field to provide systematic education programmes and seminars for recurrent education. We shall introduce three major works which we refer to develop the IPSJ DS curriculum standard.

2.1 EDISON Data Science Framework [7]

The EDISON Data Science Framework defines DS-BoK as a set of DS knowledge areas (KA) through the analysis of various BOKs related to data science. The EDISON DS Framework also contains DS Model Curriculum and assigns ECTS (European Credit Transfer and Accumulation System) to each KA as represented in Table 1.

Although the EDISON Framework does not provide further detail of the topics at each KA, it covers a wide area of the related fields. Thus, we shall utilise the KAs and the assigned ECTS in developing the IPSJ DS curriculum since DS has an interdisciplinary nature.

Table 1. Organisation of EDISON DS-BoK and Assigned ECTS.

KA Groups	KA#	DS Knowledge Areas	ECTS
Data Science Analysis	KA01.01	Statistical methods for data analysis	13.2
	KA01.02	Machine Learning	19.4
	KA01.03	Data Mining	9.7
	KA01.04	Text Data Mining	7.9
	KA01.05	Predictive Analysis	14.6
	KA01.06	Computational modeling, simulation and optimisation	9.4
Data Science Engineering	KA02.01	Big Data Infrastructure and Technologies	8.4
	KA02.02	Infrastructure and platforms for Data Science applications	14.0
	KA02.03	Cloud Computing technologies for Big Data and Data Analytics	12.5
	KA02.04	Data and Applications security	4.6
	KA02.05	Big Data systems organisation and engineering	15.5

KA Groups	KA#	DS Knowledge Areas	ECTS
	KA02.06	Data Science (Big Data) applications design	15.5
	KA02.07	Information systems (to support data driven decision making)	11.3

2.2 ACM Data Science Curriculum [6]

ACM Data Science Task Force announced the final report of “Computing Competencies for Undergraduate Data Science Curricula” in January 2021. Different from the EDISON DS-BoK, the educational contents are described in detail. Furthermore, the curriculum defines the priority of each content using T1, T2, and E. T1 stands for mandatory content, while T2 represents semi-mandatory content with 80% of the enrollment. E stands for elective content with 50% of the expected enrollment. This information is utilised to develop the IPSJ DS curriculum standard.

On the other hand, the ACM DS curriculum does not contain certain important KAs such as mathematics, statistics, modeling and simulation, and business fundamentals. Thus, we shall add these KAs to our DS curriculum.

We assign a sequence number (SEQ#) to each topic to define the mapping between the ACM DS curriculum and our DS curriculum. The prefix of the SEQ# corresponds to each KA defined in the ACM DS curriculum.

2.3 Data Scientist Skill Checklist [8]

The Japan Data Scientists Society first published the Data Scientist Skill Checklist (DS Skill Checklist) in November 2015, which defines the set of skills required for three levels of the data scientists. In June 2017, the DS Skill Checklist was fully adopted in the ITSS+ (DS domain) developed by the Information-Technology Promotion Agency (IPA), which is a governmental agency governed by the Ministry of Trade and Industry (METI) of Japan. The DS skill checklist is revised every two years and the latest version was published in October 2019.

The DS skill checklist classifies data scientists into three levels: Assistant, Full, and Senior Data Scientist Levels. Each level roughly corresponds to ITSS (IT Skill Standards [9]) levels 3 to 5, which corresponds to SFIA levels 4 to 6 [10] respectively.

The DS skill checklist defines essential skills and other skills at each level. A data scientist is evaluated at a certain level by satisfying all essential skills and a specific percentage of all skills (assistant level 70%, full level 60%, senior level 50%). Table 2 illustrates the organisation of the checklist.

Table 2. Organisation of the Data Scientist Skill Checklist.

Category	Subcategories
Data Engineering	IT Security, Data Processing, Data Sharing, Data Structure, Data collection, Data storage, Programming, Environment construction
Data Science	Graphical Models, Grouping, Sampling, Simulation, Data assimilation, Data understanding/validation, Data processing, Data visualisation, Pattern discovery, Connotation extraction & insight, Speech/music processing, Image/video processing, Basic mathematics, Machine learning techniques, Testing/decision making, Language Processing,

	Optimisation, Time series analysis, Understanding properties/relationships, Analysis process, Prediction(<-Forecasting)
Business Fundamentals	Data Acquisition, Understanding data from a business perspective, Problem definition, Activity Management, Contracts and Rights Protection, Code of Conduct, Business Implementation, Conception and Design, Analysis and evaluation, Logical thinking

The DS Skill Checklist is developed from the viewpoint of business and covers detailed skills of the KA's which are missing in the ACM DS curriculum. Thus, we shall integrate the DS Skill Checklist and the ACM DS curriculum to fully cover the EDISON DS-BoK.

3 Curriculum Development Policy

The development policy of the IPSJ DS curriculum is defined in this section. Each of the previous works introduced in Section 2 has its limitations explained in the section. We shall overcome these problems by integrating these three works. We shall mainly discuss the intention of each policy in this section. Implementation of the major policies will be explained in the succeeding sections.

A) Covers a Wide Area of Data Science to Ensure International Compatibility

DS is considered multidisciplinary since various existing areas are related to DS. Some of them are statistics, artificial intelligence, machine learning, data engineering, business analysis, and security and privacy as can be observed in EDISON DS-BoK. ACM DS curriculum and DS Skill checklist cover only a part of DS-BoK so that both are required to cover DS-BoK.

Furthermore, different KAs are used in the ACM DS curriculum and DS Skill checklist due to their different viewpoints so that integration of the classification criteria is necessary. Thus, IPSJ utilises the organisation of DS-BoK with some modification to reduce the hierarchy of the KAs. The detail will be explained in Section 4.

B) Collaboration with IPSJ Data Scientist Certification (under Development)

IPSJ is currently developing a certification scheme to recognise the ability of data scientists. The certification scheme is based on the DS Skill Checklist since it is approved by a governmental agency (IPA) and is widely accepted among Japanese IT companies. The DS Skill Checklist defines three levels of the data scientist. IPSJ is planning to provide two, assistant and full, levels of the data scientist among them.

When we developed our DS curriculum, we selected the DS Skill Checklist of the assistant level since the DS curriculum is prepared for undergraduate students. There are also many DS seminars in the Japanese industry that reference the DS Skill Checklist of the assistant or the full levels. Utilisation of the topics at the assistant level from the DS Skills Checklist will be useful to develop a unified system to develop and evaluate data scientists in academia and industry in Japan.

C) Specify the Knowledge and Skills Required of the Students Majoring in DS

Although EDISON DS-BoK covers a wide range of data science areas, it only specifies the required knowledge and skills abstractly. Thus, we shall utilise the ACM Data Science curriculum and the DS Skills Checklist to specify them concretely. The topics derived from the ACM DS curriculum are assigned serial numbers beginning

with "ACM" in the IPSJ DS curriculum standard, while the topics originated from the DS Skills Checklist are numbered with a serial number beginning with "DS". All other topics are assigned serial numbers beginning with "IPSJ".

D) The Size of the IPSJ DS Curriculum is Approximately 675 Class Hours (60 Credits in the Japanese Credit System)

More than 124 credits are required to graduate from university in Japan. A student can typically earn 2 credits by taking 90-minutes class 15 times (22.5 class hours). IPSJ commonly assigns 60 credits to develop a curriculum standard for undergraduate majored in some computing disciplines such as CS, CE, SE, IS, and IT. We shall also adopt this policy in developing our DS curriculum standard.

At a typical Japanese university, 20-30 credits are assigned to general education for all undergraduate students regardless of their major subject. Approximately 10 credits are assigned for the graduation research project. Thus 85-95 credits can be assigned to the education for the major subject. Considering the diversity of the education programme, it is realistic to leave a certain amount of room to utilise the characteristics of each educational programme.

We shall explain the details of the time assigned to the topics contained in the IPSJ DS curriculum in Section 5. When we assign a specific time to each topic, we need to adjust the contents of the topic so that we can teach the topic within the assigned time constraints. The detail of the adjustment will be described in Section 6.

E) Students are Assumed to Have Studied the Model Curriculum in Mathematics, Data Science, and AI (Literacy Level)

The Model Curriculum for Mathematics, Data Science and AI Education (Literacy) [5] is developed by the Japan Consortium for Strengthening Bases for Mathematics and Data Science education and is authorised by the Ministry of Education (MEXT) of Japan. Since all university students are expected to learn the model curriculum, the model curriculum is considered a prerequisite of the IPSJ DS curriculum standard.

Thus, no time will be assigned for the topics included in the model curriculum.

4 Integrating Classification Criteria

We shall integrate classification criteria of the ACM DS curriculum and DS skill checklist utilising EDISON DS-BoK in this section since EDISON DS-BoK covers the broadest areas.

4.1 General Principle

The integrated KA has two levels represented in Table 3. The integrated KA also utilises the classification of the ACM DS curriculum as well as EDISON DS-BoK to balance the number of topics at each subcategory.

Table 3. Organisation of the IPSJ Data Science Knowledge Areas.

Top Level Category	Subcategories
A: Basic Mathematics and Mathematical Statistics	1. Basic Linear Algebra, 2. Differential and Integral Calculus, 3. Mathematical Statistics

Top Level Category	Subcategories
B1: Data Mining	1. Similarity Measures, 2. Data Preparation, 3. Information Extraction, 4. Cluster Analysis, 5. Classification and Regression, 6. Pattern Mining, 7. Outlier Detection, 8. Time Series Data, 9. Web Data Mining
B2: Artificial Intelligence	1. General, 2. Knowledge Representation and Reasoning (Logic-Based Models), 3. Knowledge Representation and Reasoning (Probability-Based Models), 4. Planning and Search Strategy
B3: Machine Learning	1. General, 2. Supervised Learning, 3. Unsupervised Learning, 4. Mixed Methods, 5. Deep Learning
C: Modeling and Simulation	1. Sampling, 2. Data Visualisation, 3. Analysis Process, 4. Data Understanding and Verification, 5. Language Processing, 6. Image and Video Processing
D1: Fundamental Computer Science	1. Basic Computer Architecture, 2. Storage System Basics, 3. Operating System Basics, 4. File System, 5. Network, 6. Web and Web Programming, 7. Compilers and Interpreters, 8. Algorithmic Thinking and Problem Solving, 9. Algorithms, 10. Data Structures, 11. Programming, 12. Numerical Computation
D2: Big Data System	1. Challenges Associated with Scale, 2. Computer Architecture to Handle Big Data, 3. Parallel Computing Framework, 4. Distributed Data Storage, 5. Parallel Programming, 6. Big Data Application Development Technology, 7. Cloud Computing, 8. Complexity Theory, 9. Software Support for Big Data Applications
E1: Security and Privacy	1. Security Basics, 2. Privacy Protection, 3. Data Security, 4. Data Consistency, 5. Security Analysis
E2: Human-Computer Interaction	1. General
F: Software Engineering	1. Software Design and Development, 2. Software Testing, 3. Collaboration with Related Domains
G1: Data Collection, Management, and Governance	1. Database Basics, 2. Data Collection, 3. Data Processing, 4. Data Sharing, 5. Data Storage, 6. System Operation, 7. Data Management, 8. Information Retrieval
G2: Professionalism	1. Continuing Professional Development, 2. Communication, 3. Teamwork, 4. Economic Issues, 5. Privacy and Confidentiality, 6. Ethical Issues, 7. Legal Issues, 8. Intellectual Property, 9. Automation
G3: Business Fundamentals	1. Logical Thinking, 2. Project Management, 3. Business Implementation

We define a mapping between the IPSJ DS KA's and the three KA's defined in Section 2. We also assign SEQ# to each topic of the IPSJ DS curriculum standard. A SEQ# beginning "ACM" and "DS" respectively represents a topic originated from the ACM DS curriculum or DS Skill Checklist, while a SEQ# beginning "IPSJ" represents a topic added by IPSJ. We also confirm that the IPSJ DS KA system covers the ACM DS curriculum and DS Skill Checklist by assigning the SEQ#.

4.2 Individual Considerations

38 topics are assigned to category A, which are related to the mathematical basis for data science and AI. It is assumed that students have taken Mathematics I, which is a compulsory course in the Japanese high school curriculum guidelines, and the model curriculum (literacy level) established by the Japan Inter-University Consortium for Mathematics and Data Science Education, but is not assumed that students have taken Mathematics II, III, A, B or more. In other words, students are not supposed to have learned exponential and logarithmic functions, differential and integral calculus, vectors, and matrices. Therefore, each school needs to discard related topics according to its admission policy.

Regarding Category B2, Artificial Intelligence, we put topics in that category based on the ACM Data Science Curricula, though some may suggest Artificial Intelligence should entail a wider/narrower range of topics.

Category C contains 44 topics, including two DS Knowledge Areas of EDISON DS-BoK: "Predictive Analytics (KA01.05: 11 topics)" and "Computational modeling, simulation and optimisation (KA01.06: 33 topics)". And it contains subcategories of the "Data Science" category in the Data Scientist Skill Checklist, such as Prediction, Understanding properties/relationships, Sampling, Data visualisation, Analysis process, Data understanding/validation, Connotation extraction & insight, Time series analysis, Language Processing, and Image/video processing. However, there are no corresponding topics in the ACM DS curriculum.

Categories G1 to G3 cover the following topics:

1. Identification of business challenges; gathering, storing, processing, and sharing data; understanding data from a business perspective
2. Building operating environments, planning systems
3. Codes of conduct (data ethics, compliance), activity management, contracts, protection of rights
4. Logical thinking, application in business

The ACM curriculum focuses mainly on teaching technical aspects, whereas the Japan Data Scientist Society's skills checklist focuses more on displaying the skills required in business. The differences between the two are particularly conspicuous.

By integrating the two curriculums into our curriculum, we have been able to show how the knowledge, skills, and dispositions in the ACM curriculum are applied in business. If students can learn, through education, to bridge the gap between theory and practice, we will have succeeded in achieving the goals of this category. In terms of educational methods, we recommend providing business case studies alongside theory, using practical exercises as opportunities to discuss ways of applying knowledge, and evaluating the ethical aspects of handling data from a business perspective. Such education should also motivate students to study DS.

5 Assigning Education Time

We assign education time to each topic in the IPSJ DS curriculum standard. The general principle and the individual effort are explained in this section.

5.1 General Principle

The EDISON Data Science Model Curriculum (MC-DS) contains the assignment of ECTS credit points for the BSc programme for major DS-BoK knowledge areas as represented in Table 1. We assign the education time to each IPSJ KA so that the assigned time is proportional to the ECTS points to the corresponding KA of the DS-BoK. The assigned education time is represented in Table 4. Since MC-DS does not provide assignments to the three categories G1, G2 and G3, we assign 112.5 hours to these categories considering the balance to the other categories.

The hours for each KA are divided into three categories: T1, T2, and E. The definitions of T1, T2 and E are the same as the ACM DS Curriculum.

The IPSJ DS curriculum is designed so that the effective teaching time is approximately 675 hours (60 credits x 11.25 hours/credit), where the effective teaching time is the sum of the number of hours allocated to each teaching topic multiplied by the completion rate defined for the three categories.

Table 4. Top-Level Categories and Assigned Time.

Top Level Category	Assigned Time (h)				Effective Time (h)
	T1	T2	E	Total	
A: Mathematical Fundamentals and Statistics	18.25		19.00	37.25	27.75
B1: Data Mining	28.00	36.00	78.00	142.00	95.80
B2: Artificial Intelligence	3.25	17.00	12.50	32.75	23.10
B3: Machine Learning	13.75	19.50	13.50	46.75	36.10
C: Modeling and Simulation	11.80	18.38	47.00	77.18	50.00
D1: Fundamental Computer Science	81.16	22.50	26.25	129.91	112.28
D2: Big Data System	7.00	59.50	47.00	113.50	78.10
E1: Security and Privacy	41.93		4.00	45.93	43.93
E2: Human-Computer Interaction	16.83			16.83	16.83
F: Software Engineering	50.50	25.00	38.00	113.50	89.50
G1: Data Collection, Management and Governance	27.01	25.67	17.00	69.68	56.04
G2: Professionalism	16.14	12.01	7.99	36.14	29.74
G3: Business Fundamentals	10.50		8.33	18.83	14.67
Total	326.12	235.55	318.57	880.24	673.85

The Japan Data Scientist Society assigns the education time for the topics belonging to the DS Skill Checklist. They assigned a time based on their experience at data science seminars for business persons. A total of 180 hours is assigned to the topics derived from the DS Skill Checklist, which is 26.7% of the education time of the IPSJ DS curriculum.

When allocating education time for ACM education topics, the time allocated for skill topics should be approximately two to three times the time allocated for knowledge topics. The time allocated for attitude topics will be decided by the team in charge after considering the time allocated for knowledge and skills.

5.2 Individual Considerations

The three categories B1, B2 and B3 are assigned 135, 72, and 117 topics respectively. As the numbers of topics are big, we simply divided the allocated time of each top-level category by the number of topics of the corresponding category. Since the allocated time for some of the topics is very small for its content, we recommend that any potential lecturer or course organiser may want to integrate multiple topics so that they can have a reasonable amount of time to cover them, as it would be usually the case that covering multiple relevant topics would make the lecture more efficient.

Category C was assigned 50 hours of effective time (Assigned Time (h) T1: 11.8, T2: 18.38, E: 47.0, Total: 77.18; as shown in Table. 4). It was assigned based on “Hours required for education” of corresponding topics in the Data Scientist Skill Checklist. Because, indeed, “Hours required for education” in the Data Scientist Skill Checklist of corresponding topics in Category C was quite shorter than the assigned time proportional to the ECTS points, the assigning education time was adjusted and elongated based on the ECTS points.

The topics initially assigned to Security and Privacy (category E1) come from the ACM DS curriculum and are rather rich and of high level, while the time assigned to information security is limited to 4.6 ECTS in DS-BoK. Thus, we decided to assign 45 hours to this area, which corresponds to two classes at a Japanese university.

Although the number of topics belonging to Software Engineering (category F) is not many, the granularity of the topics varies. Some of the topics are simple knowledge, while some topics require extensive software development projects. Thus, we assign 0.5 to 6 hours to each topic, taking into account the educational content.

6 Adjusting Educational Contents

The educational contents of each topic depend on the time assigned to the topic so that there is a strong relationship between the decision made in Sections 5 and 6. We shall summarise the decision to adjust educational contents in this section. We shall mainly explain the differences by focusing on the changes made to the educational content of the original curriculum.

Due to the limitation of assigned time for Security and Privacy (category E1), we assign time only to the topics with T1 priority. Although a data scientist should have sufficient knowledge of security and privacy, a data scientist is not an expert on computer security. The Japanese government provides a certification scheme [11] for computer security professionals based on ISO/IEC 27000 series. A data scientist is expected to collaborate with such security professionals.

There are many descriptions such as “can practice something in all fields” in Software Engineering (category F) of the ACM DS curriculum. We have changed the wording to “can practise something in multiple fields” because it is considered unrealistic considering the assigned time constraints.

We also added the following topics based on the revision plan of SWEBOK 2021: Design patterns for AI and IoT applications, Value creation and proposition, Agile software development practice, and DevOps. However, since the contents of the revision of SWEBOK have not yet been finalised, changes may be made in the future.

7 Conclusion and Future Direction

IPJSJ has made the draft DS Curriculum Standards available for public comment and published the DS Curriculum Standard on April 2021 after a minor revision considering the submitted comments and the updates of the ACM DS curriculum.

The educational content of this data science curriculum standard is quite rich, and not many universities can cover all these contents by themselves. However, with the Corona disaster, online education has become popular in universities. The Japanese Standards for the Establishment of Universities have recently been revised to allow multiple universities to collaborate to run a single educational programme. We believe that a combination of online education and university collaboration is quite effective to realise quality assurance of specialised education in data science.

IPJSJ is also working to develop a collaboration of college-level education majoring in data science and certification scheme for data scientists in industries. A seamless development and evaluation system are essential for the effective development of data scientists expected to contribute to many aspects of society and business.

Acknowledgement. This research is supported by JSPS Kakenhi #20K03232.

References

1. Integrated Innovation Strategy Promotion Council of Japan: AI Strategy. https://www.kantei.go.jp/jp/singi/ai_senryaku/pdf/aistratagy2019en.pdf (2019).
2. Information Processing Society of Japan: J17 Computing Curriculum Standard. https://www.ipsj.or.jp/annai/committee/education/j07/curriculum_j17.html (in Japanese) (2017/)
3. Information Processing Society of Japan: Data Scientist Strategy. https://www.ipsj.or.jp/release/20210413_DSstrategy.html (in Japanese) (2021).
4. Japan Inter-University Consortium for Mathematics & Data Science. <http://www.mi.u-tokyo.ac.jp/consortium/en/>.
5. Japan Inter-University Consortium for Mathematics & Data Science: Model Curriculum for Mathematics, Data Science and AI Education (Literacy) - Learning Data-Thinking (in Japanese). http://www.mi.u-tokyo.ac.jp/consortium/model_literacy.html (2020).
6. ACM Data Science Task Force: Computing Competencies for Undergraduate Data Science Curricula. <http://dstf.acm.org/> (2021).
7. EDISON Data Science Framework. <https://edisoncommunity.github.io/EDSF/>.
8. The Japan Data Scientist Society: Data Scientist Skill Checklist (in Japanese). <https://www.datascientist.or.jp/common/docs/skillcheck.xlsx> (2019).
9. IPA: IT Skill Standard. https://www.ipa.go.jp/english/humandev/forth_download.html.

10. SFIA Foundation: Skills Framework for the Information Age, Version 8. <https://sfia-online.org/en>.
11. IPA: Registered Information Security Specialist Examination. <https://www.ipa.go.jp/files/000009645.pdf>.