

## Образование в информационном обществе

# ДАТА-ГРАМОТНОСТЬ КАК НОВАЯ ЦИФРОВАЯ КОМПЕТЕНЦИЯ

Статья рекомендована к публикации главным редактором Т.В. Ершовой 09.10.2020 г.

### **Дерябин Андрей Александрович**

*MSc Social Psychology*

*Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации, Федеральный институт развития образования, Научно-исследовательский центр социализации и персонализации образования детей, научный сотрудник  
Москва, Россия  
deryabin-aa@ranepa.ru*

### **Попов Александр Анатольевич**

*Доктор философских наук, доцент*

*Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации, Федеральный институт развития образования, заведующий научно-исследовательским сектором «Открытое образование»  
Московский городской педагогический университет, Институт системных проектов, заведующий лабораторией компетентностных практик образования  
Новосибирский государственный технический университет, гуманитарный факультет, профессор кафедры социологии и массовых коммуникаций  
Москва, Россия  
popov-aa@ranepa.ru*

### **Аннотация**

Рассматривается понятие «дата-грамотности» как одной из базовых цифровых компетенций и науки о данных как направлении научно-инженерного образования и социокультурной практики. Приводятся примеры реализации образовательных модулей в области науки о данных для подросткового и юношеского возраста. Развитие предметных навыков учащихся соотносится с требованиями стандартной индустриальной модели исследования данных. Наряду с развитием технических компетенций в профессиональном образовании, рассматриваются аспекты общей дата-грамотности, связанные с критическим мышлением. Материалом для анализа послужили публикации преимущественно 2014-2020 годов в научных изданиях, индексируемых в базе данных Scopus. Статья подготовлена в рамках выполнения научно-исследовательской работы государственного задания РАНХиГС.

### **Ключевые слова**

*Наука о данных; машинное обучение; образовательные программы; дата-грамотность; анализ данных; data science; data literacy; искусственный интеллект; профориентация; информатика; обществознание*

### **Введение**

Наука о данных (data science) затрагивает все больший круг сфер обыденной жизни, но общее образование в России пока еще не отреагировало на новые вызовы, связанные с этим аспектом цифровизации. Изменения в технологиях, экономике и обществе за последние два десятилетия сформировали новую повестку как для преподавания математики и информационных технологий, так и для общественных наук и медиа-образования. Образование во всех этих областях требует переосмысления содержания и методов преподавания в связи с растущим значением науки о данных и искусственного интеллекта в контексте фундаментальных изменений в экономике и на рынке труда.

---

© Дерябин А.А., Попов А.А., 2020. Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

УДК 371.64/69

В то же время попытки внедрить образование в сфере науки о данных на уровне общего образования достаточно редки даже в тех странах, которые лидируют в использовании цифровых технологий для повышения своей конкурентоспособности и благосостояния.<sup>1</sup>

Термин «наука о данных» (data science) широко используется, когда речь идет об использовании компьютерных методов для сбора, обработки, анализа и визуализации больших объемов данных. При этом далеко не всегда речь идет о так называемых «больших данных» (big data), под которым понимаются данные, чей объем, многообразие и скорость прироста превосходят возможности традиционных систем хранения и обработки информации, существовавших на рубеже 2000-х годов [1]. Методы науки о данных могут применяться для работы как с «большими», так и с тривиальными объемами данных.

Основные компетенции в data science и машинном обучении на уровне профессионального образования преимущественно задаются соответствующими элементами стандартного процесса по исследованию данных (CRISP-DM)<sup>2</sup>, который включает в себя следующие фазы: *понимание целей – начальное изучение данных – подготовка данных – моделирование – оценка – внедрение* [2]. На более низком уровне эта модель предусматривает наличие компетенций по поиску, получению данных, их курированию, манипулированию ими, обработке, анализу, визуализации и представлению результатов исследования в разных формах.

## 1 Дата-грамотность и ее социокультурные измерения

Учитывая то, что данные сегодня являются повсеместными и затрагивают все аспекты существования индивида в обществе, в академическом сообществе обсуждается не только преподавание науки о данных как учебной дисциплины для будущих специалистов в этой области, но и понятие «дата-грамотности» (data literacy), применимой в более широком контексте общественной жизни и этической проблематики. Дата-грамотность как способность понимать, что такое данные и на что они способны, рассматривается как важный элемент формирования личности, способной принимать рациональные решения, осуществлять этический выбор и вырабатывать политические мнения в современном цифровом мире [3].

В этом отношении понятие дата-грамотности близко к ранее вошедшим в обиход «медиа-грамотности» или «информационной грамотности», формируя наряду с ними букет современных цифровых компетенций.

Вне зависимости от того, нацелено ли образование в области науки о данных на подготовку к профессиональной деятельности или просто является сопутствующим техническим навыком, оно имплицитно формирует, с одной стороны, *рационального индивида*, понимающего, как и с какими последствиями он использует или предоставляет данные, когда использует свой телефон, читает новости, прокладывает маршрут в навигаторе, или взаимодействует с онлайн-магазином. С другой стороны, навыки анализа данных позволяют индивиду интерпретировать статистику и непредвзято судить об общественных, технических и политических процессах и повестках, формируя, таким образом *информированного гражданина* [3, С.29]. Примером может быть сложность интерпретации неполных и противоречивых статистических данных о распространении коронавируса COVID-19, которую испытывали аудитории средств массовой информации по всему миру в марте 2020 года.<sup>3</sup>

В контексте дискуссии о «цифровой грамотности», представляющей собой самый широкий спектр навыков, необходимых для того, чтобы гражданин мог учиться и ориентироваться в обществе цифровых знаний и практик, мы рассматриваем дата-грамотность как ее подмножество, связанное с знаниями и навыками в области обработки и интерпретации количественных данных. Существуют различные определения дата-грамотности, акценты в которых варьируются в зависимости от специализации приводящих их авторов, некоторые из определений приведены ниже.

Дата-грамотность включает в себя способность понимать и оценивать информацию, которая может быть получена на основе данных [4]. Владеющий дата-грамотностью учащийся должен

<sup>1</sup> Network Readiness Index 2019 [Электронный ресурс]. URL <http://www.networkreadinessindex.org> (дата обращения 10.04.2020).

<sup>2</sup> CRISP-DM (Cross-Industry Standard Process for Data Mining) – межотраслевой стандартный процесс для исследования данных, проверенная в промышленности и наиболее распространённая методология по исследованию данных.

<sup>3</sup> Например: Koerth M., Bronner L., Mithani J. Why It's So Freaking Hard To Make A Good COVID-19 Model. [Электронный ресурс] Опубликовано 31.03.2020. URL <https://fivethirtyeight.com/features/coronavirus-case-counts-are-meaningless/>

владеть количественными и аналитическими инструментами, необходимыми для решения задачи и способностью применять их в контексте, чтобы анализировать, интерпретировать и сообщать другим результаты анализа данных [5].

Дата-грамотность лежит на пересечении областей количественного мышления и науки о данных. Количественное мышление (*quantitative reasoning*) – это способность применять математические принципы, критическое мышление и ясную логику к решению каждодневных проблем [6].

Мастерство аналитика данных основывается на понимании предметного контекста, в котором находится набор данных, знании математики, статистики и на компьютерных навыках [7, 8]. Последнее определение, как и многие другие, отражает тенденцию сдвигать фокус внимания на технические аспекты дата-грамотности. Однако данные далеко не всегда являются объективными свидетельствами естественных феноменов, независимыми от применяемых исследователем средств их получения и намерений определенным образом их использовать [9]:

1. данные всегда несут в себе следы своего технического и материального происхождения [10];
2. инструменты создания и обработки данных не нейтральны, но несут в себе ту или иную имплицитную теорию с встроенным намерением редуцировать сложные феномены к полезным формам данных [10, 11];
3. данные существуют потому, что кто-то для чего-то хотел их создать; они представляют собой ценность, актив и обладают силой убеждения, следовательно, существуют не только в предметных, но также в экономических и политических границах [12].

Fotopoulou [13] находит сходство между концепциями дата- и медиа-грамотности, в которой чтение и интерпретация медиа понимается как социокультурная практика, а отправители и реципиенты рассматриваются в связи с их принадлежностью к социальным группам и сообществам. Каждое из которых, в свою очередь, располагает собственными мотивами, символическими, материальными ресурсами и интерпретативным репертуаром [14]. Хотя и не в полной мере (практики и деятельности, связанные с данными, разнообразней, чем «чтение» в теории медиа), данные, как и медиа-сообщение, можно рассматривать как содержание и канал коммуникации.

Таким образом, взаимодействие с данными является социокультурной практикой, и большие данные могут нести в себе скрытые мотивы их кураторов [15] – достаточно привести в качестве примера электоральные данные или данные здравоохранения [16], что концептуально позволяет применять к ним соответствующий гуманитарный методологический аппарат с оговоркой: если такой подход продемонстрирует свою продуктивность. Например, Pangrazio и Sefton-Green [17] считают, что следует осторожно относиться к данным как к «тексту» в силу того, что конкретный набор данных может претерпевать слишком много циклов обработки (от чистки «сырых» данных к искусственной инженерии нужных переменных, и далее к метаданным) и использования его разными акторами в разных контекстах с разными намерениями, что делает проблематичным саму возможность анализа набора данных как неизменного во времени текста<sup>4</sup>.

Тем не менее, ряд авторов рассматривает сферу науки о данных и «дата-грамотность» в русле критической теории. Этот подход в образовательных инициативах подкрепляется «дискурсом продуктивности», сфокусированном не на углублении технических навыков анализа данных, а на массовом развитии «критической дата-грамотности» (*critical data literacy*) для общественного блага и расширения гражданских прав и возможностей [18]. Практики и исследователи этого направления часто обращаются к открытым правительственным и международным данным, касающимся широкого круга вопросов высокой общественной значимости [19]. Примерами инициатив, направленных на повышение дата-грамотности различных общественных групп, являются программы по журналистике данных [20], семинары, адресованные некоммерческим организациям и общественным активистам [13].

Однако, несмотря на риторику прозрачности, окружающую публично доступные государственные наборы данных, многие специалисты и за рубежом, и в России обращают внимание на противоречивые мотивы, сопровождающие их публикацию [21].<sup>5</sup>

<sup>4</sup> Это не касается случаев, когда собственно текст как таковой является массивом данных, обрабатываемым и исследуемым машинными средствами обработки естественного языка.

<sup>5</sup> См. также: (1) Линделл Д. Раскрытие данных о заболеваниях россиян [Электронный ресурс] / Инфокультура. Опубликовано 15.03.2020. URL <https://youtu.be/fepraC4frf4>; (2) Открытость и открытые данные в современной России. Дискуссия [Электронный ресурс] / Инфокультура. Опубликовано 20.03.2020. URL <https://youtu.be/-oXQTTEsaN8>

## 2 Спектр компетенций и образовательных программ в области науки о данных

В настоящее время в литературе описано нет так много инициатив по преподаванию науки о данных учащимся подросткового и юношеского возраста; некоторые из них, позволяющие судить о разбросе содержания, длительности и целевой аудитории, приведены в Таблице 1. В то же время существует большое количество программ профессионального образования в сфере науки о данных бакалаврского и магистерского уровня [22], а соответствующих онлайн-курсов, отвечающих растущим запросам кадрового рынка в этой сфере, еще больше, и их предметное содержание общедоступно<sup>6</sup>. Однако, как отмечают Heinemann и др. [23], программы высшего образования не применимы на уровне школы т.к. они слишком сфокусированы на углубленном изучении технической стороны анализа данных и уделяют недостаточно внимания рассмотрению вопросов влияния данных на жизнь общества и индивида, важных для подростков. Структура профессиональных курсов связана с производственным циклом стандартного проекта исследования данных, а развиваемые в них компетенции хорошо описаны (компьютерное и статистическое мышление; математические основы, алгоритмы, построение и оценка моделей машинного обучения, подготовка данных, представление результатов и т.д.) [22]. Heinemann и др. [23] резонно полагают, что копировать всю модель CRISP-DM для программы школьного уровня не имеет смысла, т.к. учащиеся не находятся в реальном производственном контексте, а стало быть, такие фазы как «понимание бизнес-требований» и «внедрение» нужно опустить или адаптировать к учебному процессу, как это, например сделано у Srikant и Aggarwal [24], хотя, в целом, цикл CRISP-DM задает вполне работающую рамку для образовательной программы и для реализации коллективных учебных проектов.

Приведение этой модели к менее абстрактным и более простым для учащихся этапам делается через фокус на более низкоуровневых практических навыках работы с данными в той или иной программной среде (например, с использованием языков Python или R). Так, Erickson и др. [25] предлагают продвигаться по фазам исследования данных, делая элементарные «шаги» (data moves) – выполняя конкретные действия с данными, без которых не обходится ни один дата-проект, например, фильтрация, группировка данных или построение диаграмм. По мнению авторов и цитирующих их специалистов, этот «ремесленный» уровень работы с данными готовит школьников к лучшему пониманию более сложного материала - индуктивной статистики и математических основ машинного обучения.

Приведем в качестве примера некоторые «шаги» (data moves) из программы интенсива для старшеклассников «Дата Кампус: основы анализа данных и машинного обучения»<sup>7</sup>: импорт-экспорт данных, проверка и исправление типов данных, кодирование категориальных переменных, переименование-добавление-удаление столбцов, устранение или заполнение пустых значений, поиск выбросов данных, вывод описательных статистик, вывод информации о наборе данных, построение диаграмм рассеивания, обращение к произвольным элементам и фильтрация данных.

### Заключение

Принципы построения образовательных программ для формирования дата-компетенций учащихся, можно обобщить в следующих пунктах.

1. Вырабатывать критическое отношение к данным, развивая у учащихся понимание того, как человеческие, технические и институциональные аспекты производства данных взаимодействуют между собой и отражаются на конкретном наборе данных [18, С.13-14].
2. Давать возможность студентам испытать сложность работы с большими, комплексными, беспорядочными наборами данных.
3. Помогать студентам искать паттерны и закономерности в данных, быть способными делать обобщения, уметь менять масштаб рассмотрения проблемы (например, от муниципального до странового).

<sup>6</sup> Например, на сайтах Coursera <https://www.coursera.org/browse/data-science> и edX <https://www.edx.org/course/subject/data-science>.

<sup>7</sup> Дерябин А.А. Дата Кампус: игры с большими данными [Электронный ресурс] // EdExpert. Эффективные решения в образовании. URL <http://edexpert.ru/datacampus> (дата обращения 10.04.2020).



4. Давать студентам представление о том, как постановка исследовательской задачи влияет на представление и трансформацию данных, и как переопределение задачи реорганизует используемые ими инструменты и техники анализа данных.

Очевидно, конкретные форматы и методы преподавания науки о данных связаны с тем, как образовательная программа или учебный курс определяет свою задачу: знакомство учащихся с наукой о данных как первый шаг на пути их профессионального становления в этой области или обучение навыкам дата-грамотности как одной из базовых компетенций, необходимых для жизни в современном обществе вне зависимости от карьерного выбора учащегося.

Таблица 1. Примеры учебных модулей науки о данных для учащихся школьного возраста.

Источник	Содержание модулей/курса	Аудитория, формат
Drafting a Data Science Curriculum for Secondary Schools [23]	Статистическое мышление и методы статистики; Большие данные и искусственный интеллект (деревья решений и нейросети); Учебные проекты, основанные на реальных данных; Наука о данных и общество.	Германия; школьники 15–18 лет. Учебный курс – часть обязательной школьной программы, 18 учащихся.
Introducing Data Science to School Kids [24]	Полный цикл машинного обучения с учителем: подготовка данных, визуализация, построение и тестирование модели.	Индия, США; школьники 10-15 лет. Одно занятие на половину дня, всего 71 учащихся в 4 городах.
A Middle-School Camp Emphasizing Data Science and Computing for Social Good [26]	Основы статистики, парное программирование (BlockPy, BBC micro:bits, Jupyter Notebooks), синтаксис Python, методы работы с данными и визуализацией; проекты: «наука о данных для социального блага».	США, средняя школа. Летний лагерь, 1 неделя.
A Middle-School Module for Introducing Data-Mining, Big-Data, Ethics and Privacy Using RapidMiner and a Hollywood Theme [27]	Поиск данных, предиктивная аналитика, визуализация, большие данные, этика и приватность, использование ПО RapidMiner, набор данных Titanic, построение рекомендательной системы про кинофильмы.	США, школьники 11-19 лет (большинство 12-14) 90-минутный практический семинар
Integrating Data Science and R Programming at an Early Stage [28]	(1) введение в науку о данных через игры в сбор данных, чистку, визуализацию. Базовая информация о машинном обучении (классификация, кластеризация) через групповые активности без компьютера; (2) введение в программирование на языке R.	США, школьники 11 лет. Практический семинар на 15 дней по 2 часа; 100 учащихся.
Machine Learning Exercises for High School Students [29]	Написание кода для обработки изображений и передачи в готовую модель классификации изображений. План урока: введение в обработку изображений (15 мин), описание метода (20 мин), групповая работа (30 мин), работа с приложением машинного обучения (15 мин), эксперименты (20 мин).	Румыния, школьники 13 и 19 лет. 2-часовой урок
Exploring design principles for data literacy activities to support children's inquiries from complex data [30]	Полевые исследования с сбором и анализом данных на тему «умный город» (использование солнечной энергии и потребление электричества).	Великобритания, дети 10-14 лет. 4 полевых исследования длительностью несколько недель. Всего 67 учащихся.

## Литература

1. ISO/IEC 20546:2019 Information technology – Big data – Overview and vocabulary. [Электронный ресурс] URL <https://www.iso.org/standard/68305.html> (дата обращения 09.10.2020).
2. Chapman P. и др. CRISP-DM 1.0: Step-by-step data mining guide. [Электронный ресурс]. / CRISP-DM Consortium, 2000. URL <https://www.the-modeling-agency.com/crisp-dm.pdf> (дата обращения 10.04.2020)
3. Matzner T. Data science education as contribution to media ethics. // Biehler R. и др. Paderborn Symposium on Data Science Education 2017: The Collected Extended Abstracts. Paderborn: University of Paderborn, 2018.
4. Schield M. Information literacy, statistical literacy and data literacy. // IASSIST Quarterly, 28(2), 6–11, 2018.
5. Gibson J. P., Mourad T. The growing importance of data literacy in life science education. American Journal of Botany, 105(12), 1–4, 2018.
6. Steen L. A. Achieving quantitative literacy: An urgent challenge for higher education / Washington, DC: Mathematical Association of America, 2004 - 124 с.
7. Conway D. The data science venn diagram. [Электронный ресурс]. September 30, 2010. URL <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram> (дата обращения 10.04.2020).
8. Finzer, W. The data science education dilemma [Электронный ресурс]. // Technology Innovations in Statistics Education, 2013, 7(2). URL <https://escholarship.org/uc/item/7gv0q9dc> (дата обращения 10.04.2020).
9. Hardy L., Dixon C., Hsi S. From Data Collectors to Data Producers: Shifting Students' Relationship to Data // Journal of the Learning Sciences, 2020, 29(1). С. 104–126.
10. Ihde D. Heidegger's philosophy of technology // Technics and Praxis. Boston studies in the philosophy of science, 1979, Vol. 24. С. 103–129.
11. Tala S. Unified view of science and technology for education: Technoscience and technoscience education // Science & Education, 2009, 18(3–4). С. 275–298.
12. Latour B. Science in action: How to follow scientists and engineers through society. Harvard University Press, 1987. - 288 с.
13. Fotopoulou A. Conceptualising critical data literacies for civil society organisations: agency, care, and social responsibility dilemma // Information Communication and Society, 2020. doi:10.1080/1369118X.2020.1716041
14. Kellner D., Share J. Toward critical media literacy: Core concepts, debates, organizations, and policy // Discourse: Studies in the Cultural Politics of Education, 2005, 26(3). С. 369–386.
15. Crawford K. The hidden biases in big data [Электронный ресурс] // Harvard Business Review Blog Network, 2013, April 1. URL <http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/> (дата обращения 10.04.2020).
16. Bowker G. C. Data flakes: An afterword to 'raw data' is an oxymoron. In Gitelman L. (Ed.), 'Raw data' is an oxymoron. Cambridge, MA: MIT Press, 2013. С. 167–171.
17. Pangrazio L., Sefton-Green J. The social utility of 'data literacy.' // Learning, Media and Technology, 2019. doi:10.1080/17439884.2020.1707223
18. Wise A. F. Educating Data Scientists and Data Literate Citizens for a New Generation of Data // Journal of the Learning Sciences, 2020, 29(1). С. 165–181.
19. Pangrazio L., Selwyn N. 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data // New Media and Society, 2019, 21(2). С. 419–437.
20. Hewett J. Learning to teach data journalism: Innovation, influence and constraints // Journalism, 2016, 17(1). С. 119–137.
21. Halford S., Pope C., Weal M. Digital futures? Sociological challenges and opportunities in the emergent semantic web // Sociology, 2013, 47(1). С. 173–189.
22. De Veaux, R. и др. Curriculum guidelines for undergraduate programs in data science // Annual Review of Statistics and Its Application, 2017, 4(1). С. 15–30.
23. Heinemann B. и др. Drafting a data science curriculum for secondary schools / ACM International Conference Proceeding Series, 2018. С. 1–5. doi:10.1145/3279720.3279737
24. Srikant, S., & Aggarwal, V. (2017). Introducing data science to school kids / Proceedings of the Conference on Integrating Technology into Computer Science Education, ITiCSE, 561–566. doi:10.1145/3017680.3017717

25. Erickson T. и др. Data Moves: one key to data science at school level / Proceedings of the International Conference on Teaching Statistics (ICOTS-10), 2018. [Электронный ресурс]. URL [https://iase-web.org/Conference\\_Proceedings.php?p=ICOTS\\_10\\_2018](https://iase-web.org/Conference_Proceedings.php?p=ICOTS_10_2018) (дата обращения 10.04.2020).
26. Bryant C. и др. A middle-school camp emphasizing data science and computing for social good // SIGCSE 2019 - Proceedings of the 50th ACM Technical Symposium on Computer Science Education, 2019. С. 358–364. doi:10.1145/3287324.3287510
27. Dryer A., Walia N., Chattopadhyay A. A middle-school module for introducing data-mining, big-data, ethics and privacy using rapidminer and a Hollywood theme / SIGCSE 2018 - Proceedings of the 49th ACM Technical Symposium on Computer Science Education, 2018. С. 753–758. doi:10.1145/3159450.3159553
28. Datta S., Nagabandi V. (2018). Integrating data science and R programming at an early stage / IEEE 4th International Conference on Soft Computing and Machine Intelligence, ISCMi 2017, 2018. С. 1–5. doi:10.1109/ISCMi.2017.8279587
29. Mariescu-Istodor R., Jormanainen I. Machine Learning Exercises for High School Students [Электронный ресурс]. / Proceedings of the 19th Koli Calling International Conference on Computing Education Research, 2019. URL [http://www.cs.columbia.edu/~CS4HS/talks/ml\\_for\\_hs.pdf](http://www.cs.columbia.edu/~CS4HS/talks/ml_for_hs.pdf) (дата обращения 10.04.2020).
30. Wolff A., Wermelinger M., Petre M. Exploring design principles for data literacy activities to support children’s inquiries from complex data // International Journal of Human Computer Studies, 2019, 129 (March). С. 41–54. doi:10.1016/j.jhcs.2019.03.006

## DATA LITERACY AS A NEW DIGITAL COMPETENCE

**Deryabin, Andrey Aleksandrovich**

*MSc Social Psychology*

*Russian Presidential Academy of National Economy and Public Administration, Federal Education Development Institute, researcher*

*Moscow, Russia*

*deryabin-aa@ranepa.ru*

**Popov, Aleksandr Anatolyevich**

*Dr. Sci. (Philos.), Associate Professor*

*Russian Presidential Academy of National Economy and Public Administration, Federal Education Development Institute, Head of Open Education section*

*Moscow City University, Institute of System Projects, head of Competency Practices Lab*

*Novosibirsk State Technical University, professor of Sociology & Mass Communications Dept.*

*Moscow, Russia*

*popov-aa@ranepa.ru*

### Abstract

*The concept of 'data literacy' as a core competence in the digital society is reviewed altogether with Data Science as a strand of STEM education, and as a sociocultural practice and empowering resource for community activism.*

### Keywords

*Big data; data literacy; data science education; machine learning; artificial intelligence; school education; computer science education; curriculum development*

### References

1. Ward J.S., Baker A. Undefined By Data: A Survey of Big Data Definitions. [Электронный ресурс]. // Cornell University Library, 20 Sept. 2013. URL <http://arxiv.org/abs/1309.5821> (дата обращения 10.04.2020)
2. Чарман Р. и др. CRISP-DM 1.0: Step-by-step data mining guide. [Электронный ресурс]. / CRISP-DM Consortium, 2000. URL <https://www.the-modeling-agency.com/crisp-dm.pdf> (дата обращения 10.04.2020)
3. Matzner T. Data science education as contribution to media ethics. // Biehler R. и др. Paderborn Symposium on Data Science Education 2017: The Collected Extended Abstracts. Paderborn: University of Paderborn, 2018.
4. Schield M. Information literacy, statistical literacy and data literacy. // IASSIST Quarterly, 28(2), 6–11, 2018.
5. Gibson J. P., Mourad T. The growing importance of data literacy in life science education. American Journal of Botany, 105(12), 1–4, 2018.
6. Steen L. A. Achieving quantitative literacy: An urgent challenge for higher education / Washington, DC: Mathematical Association of America, 2004 - 124 с.
7. Conway D. The data science venn diagram. [Электронный ресурс]. September 30, 2010. URL <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram> (дата обращения 10.04.2020).
8. Finzer, W. The data science education dilemma [Электронный ресурс]. // Technology Innovations in Statistics Education, 2013, 7(2). URL <https://escholarship.org/uc/item/7gv0q9dc> (дата обращения 10.04.2020).
9. Hardy L., Dixon C., Hsi S. From Data Collectors to Data Producers: Shifting Students' Relationship to Data // Journal of the Learning Sciences, 2020, 29(1). С. 104–126.
10. Ihde D. Heidegger's philosophy of technology // Technics and Praxis. Boston studies in the philosophy of science, 1979, Vol. 24. С. 103–129.
11. Tala S. Unified view of science and technology for education: Technoscience and technoscience education // Science & Education, 2009, 18(3–4). С. 275–298.
12. Latour B. Science in action : How to follow scientists and engineers through society. Harvard University Press, 1987. - 288 с.



13. Fotopoulou A. Conceptualising critical data literacies for civil society organisations: agency, care, and social responsibility dilemma // *Information Communication and Society*, 2020. doi:10.1080/1369118X.2020.1716041
14. Kellner D., Share J. Toward critical media literacy: Core concepts, debates, organizations, and policy // *Discourse: Studies in the Cultural Politics of Education*, 2005, 26(3). С. 369–386.
15. Crawford K. The hidden biases in big data [Электронный ресурс] // *Harvard Business Review Blog Network*, 2013, April 1. URL <http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/> (дата обращения 10.04.2020).
16. Bowker G. C. Data flakes: An afterword to ‘raw data’ is an oxymoron. In Gitelman L. (Ed.), ‘Raw data’ is an oxymoron. Cambridge, MA: MIT Press, 2013. С. 167–171.
17. Pangrazio L., Sefton-Green J. The social utility of ‘data literacy.’ // *Learning, Media and Technology*, 2019. doi:10.1080/17439884.2020.1707223
18. Wise A. F. Educating Data Scientists and Data Literate Citizens for a New Generation of Data // *Journal of the Learning Sciences*, 2020, 29(1). С. 165–181.
19. Pangrazio L., Selwyn N. ‘Personal data literacies’: A critical literacies approach to enhancing understandings of personal digital data // *New Media and Society*, 2019, 21(2). С. 419–437.
20. Hewett J. Learning to teach data journalism: Innovation, influence and constraints // *Journalism*, 2016, 17(1). С. 119–137.
21. Halford S., Pope C., Weal M. Digital futures? Sociological challenges and opportunities in the emergent semantic web // *Sociology*, 2013, 47(1). С. 173–189.
22. De Veaux, R. и др. Curriculum guidelines for undergraduate programs in data science // *Annual Review of Statistics and Its Application*, 2017, 4(1). С. 15–30.
23. Heinemann B. и др. Drafting a data science curriculum for secondary schools / *ACM International Conference Proceeding Series*, 2018. С. 1–5. doi:10.1145/3279720.3279737
24. Srikant, S., & Aggarwal, V. (2017). Introducing data science to school kids / *Proceedings of the Conference on Integrating Technology into Computer Science Education, ITiCSE*, 561–566. doi:10.1145/3017680.3017717
25. Erickson T. и др. Data Moves: one key to data science at school level / *Proceedings of the International Conference on Teaching Statistics (ICOTS-10)*, 2018. [Электронный ресурс]. URL [https://iase-web.org/Conference\\_Proceedings.php?p=ICOTS\\_10\\_2018](https://iase-web.org/Conference_Proceedings.php?p=ICOTS_10_2018) (дата обращения 10.04.2020).
26. Bryant C. и др. A middle-school camp emphasizing data science and computing for social good // *SIGCSE 2019 - Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 2019. С. 358–364. doi:10.1145/3287324.3287510
27. Dryer A., Walia N., Chattopadhyay A. A middle-school module for introducing data-mining, big-data, ethics and privacy using rapidminer and a Hollywood theme / *SIGCSE 2018 - Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, 2018. С. 753–758. doi:10.1145/3159450.3159553
28. Datta S., Nagabandi V. (2018). Integrating data science and R programming at an early stage / *IEEE 4th International Conference on Soft Computing and Machine Intelligence, ISCMi 2017*, 2018. С. 1–5. doi:10.1109/ISCMi.2017.8279587
29. Marinescu-Istodor R., Jormanainen I. Machine Learning Exercises for High School Students [Электронный ресурс]. / *Proceedings of the 19th Koli Calling International Conference on Computing Education Research*, 2019. URL [http://www.cs.columbia.edu/~CS4HS/talks/ml\\_for\\_hs.pdf](http://www.cs.columbia.edu/~CS4HS/talks/ml_for_hs.pdf) (дата обращения 10.04.2020).
30. Wolff A., Wermelinger M., Petre M. Exploring design principles for data literacy activities to support children’s inquiries from complex data // *International Journal of Human Computer Studies*, 2019, 129 (March). С. 41–54. doi:10.1016/j.ijhcs.2019.03.006