

УДК 37.0

БОЛЬШИЕ ДАННЫЕ В ОБРАЗОВАНИИ: НАЦИОНАЛЬНАЯ ОБРАЗОВАТЕЛЬНАЯ БАЗА ДАННЫХ

Оспанова Улжан Абаевна¹, Атанаева Мираим Кажмухамбетовна²

¹uljansbox@mail.ru

¹Докторант Факультета международных отношений, кафедры «Регионоведение»,

¹ЕНУ им. Л.Н. Гумилева, Нур-Султан, Казахстан

¹Научный руководитель – Ж. Нурбаев, А. Оспанова

²Заместитель президента АО «Информационно-аналитический центр», Нур-Султан,
Казахстан

Введение. Развитие технологий, характеризующее современность, и скорость, с которой оно трансформирует многие сферы жизни обсуждается в контексте многих социальных исследований. Сфера образования не стала исключением из общего тренда, и ее трансформация под влиянием технологического прогресса требует пересмотра многих аспектов: фундаментальных теоретических основ [1], педагогических практик [2], мастерства [3] и навыков [4], управления образованием [5], поиска новых форматов преподавания [6], образовательного контента [7,8], новых механизмов взаимодействия стейкхолдеров образования [9] и новых институциональных форм организации образования [10] и др. Помимо целого ряда новшеств и позитивных изменений, появление больших данных принесло и новые вызовы основными из которых выступают необходимость обеспечения безопасности и конфиденциальности персональных данных, этические вопросы использования данных [11].

В образовании по сравнению с другими дисциплинарными областями социальных наук генерирует сравнительно большое количество данных, которые можно анализировать. Первенство остается за данными массовых открытых онлайн-курсов, позволяющих проводить предиктивный анализ [12], анализ образовательных траекторий [13], анализ поведенческих особенностей обучающихся [14] и многое другое.

Отдельной категорией источников данных выступают государственные образовательные информационные системы [15]. В таких информационных системах происходит генерация структурированных данных, что значительно облегчает их анализ и «извлечение смысла» и способствует принятию политических решений на основе данных [16]. Достоверность данных управляемых государством информационных систем также выше, чем в других видах источников данных, соответственно выше и доверие к ним.

В Казахстане цифровизация образования находится в фокусе политической повестки дня довольно давно, начиная с 1997 г. [17] с попыток автоматизировать некоторые процессы в образовании, в том числе процесс внедрения онлайн-обучения. В 2013 г. «Национальный центр образовательной статистики и оценки» (упраздненный в настоящее время) начал вести Национальную образовательную базу данных (далее – НОБД). Национальная образовательная база данных (НОБД) – это информационная система Министерства образования и науки Республики Казахстан, агрегирующая административные статистические данные системы образования. На сегодняшний день НОБД выступает основным источником административных данных в области дошкольного, среднего и технического и профессионального образования, однако подробного анализа НОБД как источника больших данных с точки зрения теории в отечественной научной литературе не было проведено, что обуславливает актуальность и новизну настоящего исследования.

Целью данной статьи выступает проведение анализа НОБД на предмет соответствия общепринятым характеристикам больших данных с приведением соответствующих обоснований заключения о соответствии.

Методология и данные.

Методология настоящего исследования представляет собой комбинацию контент-анализа официальных документов и данных, полученных в ходе глубинных интервью экспертов (6 чел.), работающих с НОБД. Соответственно, эмпирическими данными, на которых эта статья выступают тексты официальных документов и транскрипты глубинных интервью. Изложение результатов исследования представлено в виде сводной таблицы, где основной фокус будет сделан на анализ соответствия НОБД основным общепринятым характеристикам больших данных. Предусмотрены три категории соответствия: полное соответствие, частичное соответствие, несоответствие с приведением обоснования.

Результаты. В научном дискурсе присутствуют различные вариации трактовки термина «большие данные» [18], при этом подчеркивается его субстантивный (большие массивы данных), но и дескриптивный (основные характеристики больших данных), а также утилитарный (аналитические методы и инструменты) компоненты. Результаты анализа НОБД с точки зрения ее соответствия основным общепринятым в научной литературе характеристикам больших данных – так называемым 5«V», представлены в Таблице 1. Первые три «V» характеристики были предложены Д. Лейни, они включают: 1) объем данных (англ. volume), 2) разнообразие данных (англ. variety) и 3) скорость (англ. velocity) [19]. Эти три характеристики являются общепринятыми, однако с развитием научной мысли был предложен расширенный набор характеристик с включением дополнительных параметров таких как 4) правдивость (англ. veracity) и 5) ценность (англ. value) [20]. Хотя в научной литературе рассматриваются и другие характеристики больших данных [21], названные пять все же выступают общепринятой концептуальной основой.

Таблица 1 – Анализ соответствия НОБД основным характеристикам больших данных

Характеристика	Описание характеристики*	Анализ соответствия
Объем	Относится к большим объемам данных (Тбайты, Пбайты), работа с которыми требует больших	<i>Полностью соответствует.</i> В НОБД регистрируются данные о 21 286 организациях образования, более 5,5 млн воспитанниках/учащихся и персонале всех образовательных учреждений (учащиеся, работники, администрация), составляющих около 1 млн. чел.

	вычислительных мощностей	
Разнообразии	Разнообразие типов данных. К таким данным могут относиться непрерывные, категориальные, ординальные данные, записи логов, фото и видео и др.	<p><i>Полностью соответствует.</i></p> <p>Информация об организациях образования представлена следующими разделами:</p> <ul style="list-style-type: none"> • Регистрационные сведения - основные сведения об организации образования, такие как: виды организации, форма собственности, местоположение • Материально-техническая база <ul style="list-style-type: none"> - Сведения о зданиях - информация по каждому из зданий школ (его характеристика и техническое состояние; обеспеченность инженерными коммуникациями) - Библиотека/учебники - информация по библиотеке (количество книг, читателей) и учебникам, имеющимся на балансе организации (количество учебников по классам, предметам) - Кабинеты/классы/аудитории - информация по кабинетам (количество и виды кабинетов) - Компьютеризация - информация по количеству компьютерной и иной техники, ее распределению среди учителей и учащихся - Столовая - информация о имеющиеся в школе столовой (количество мест, имеющемся оборудовании) - Дополнительные сведения по МТБ - различная информация по материально-технической базе организации (спортзалы, учебно-материальная база) • Информация об образовательном процессе - языки обучения, углубленное и полиязычное изучение предметов, имеющиеся кружки и секции • Информация об объекте образования - информация о форме обучения и классах, используемых электронных ресурсах • Общежитие/интернат - информация об общежитии организации образования (площади, проектной мощности, техническом состоянии) • Дополнительные сведения <ul style="list-style-type: none"> - о контингенте - о кадрах - о количестве и трудоустройстве выпускников - о попечительском совете организации образования (состав совета, информация о счете и средствах)

		<ul style="list-style-type: none"> • Ответственный за паспортизацию - информация об ответственном за заполнение НОБД сотруднике организации • Контингент - паспорта на каждого из учащихся организации • Персонал - паспорта на каждого из сотрудников организации <p>Информация о контингенте учащихся содержат следующие показатели:</p> <ul style="list-style-type: none"> • общие показатели (ФИО, пол, возраст, гражданство) • информация об обучении в организации образования (параллель, литера, форма и язык обучения) • обеспечение (охват питанием, подвозом; посещение кружков и секций) • наличие устройств для дистанционного обучения; наличие доступа к сети интернет • результаты обучения (годовые оценки по предметам ЕНЦ) • достижения (участие в конкурсах, олимпиадах и других мероприятиях) • отношение к социально-уязвимым категориям (является ли сиротой; ребенком, без попечения родителей; относится ли категориям, которым оказывается материальная и финансовая помощь; получает ли АСП) • особые образовательные потребности (инвалидность; наличие нарушений; обучение в специальных классах). <p>Информация о персонале содержит следующие показатели:</p> <p>общие показатели (ФИО, пол, возраст, гражданство)</p> <ul style="list-style-type: none"> • качественные показатели сотрудника (образование, степень, категория, стаж) • информация о приеме на работу (дата приема на работу, приказ, информация о трудовом договоре) • информация о работе в организации (должность, предметы, ставка) • наличие устройств для дистанционного обучения; наличие доступа к сети интернет • прохождение курсовой подготовки <p>В структуре данных присутствуют непрерывные, категориальные, ординарные, интервальные, бинарные данные. С точки зрения формата данных НОБД преобладают текстовые и числовые данные, незначительно представлены</p>
--	--	--

		изображения (прикрепление планировок зданий, фото учредительных документов и др.) и данные журнала пользовательских логов.
Скорость	Относится к скорости генерации/обновления данных, к примеру в режиме реального времени; и скорость анализа данных, реализуемая часто за счет встроенных в информационную систему инструментов аналитики. В комбинации данные должны предоставлять информацию о реальной картине на текущий момент.	<p><i>Полностью соответствует.</i></p> <p>Ранее сбор данных осуществлялся два раза в год – сентябрь-октябрь и актуализация данных в марте – апреле.</p> <p>С 3 апреля 2019 года была запущена модифицированная онлайн система, которая доступна 24/7, позволившая обеспечить актуализацию данных в режиме реального времени.</p> <p>Необходимо отметить, что для отдельных категорий данных установлен максимальный срок занесения данных в НОБД: до 3-х рабочих дней.</p>
Правдивость	Относится к степени точности и достоверности данных. Критично исключение регистрации ложных данных и максимальное исключение ошибок за счет человеческого фактора, таких как опечатки, двусмысленность и др.	<p><i>Частично соответствует</i></p> <p>Независимых исследований по вычислению стандартных ошибок НОБД не проводилось. Предусмотрены различные меры обеспечения надежности данных, однако однозначно утверждать об их достаточности невозможно.</p> <p>Существуют законодательные инструменты обеспечения правдивости данных в области порядка сбора административных данных [22], форм сбора административных данных [23].</p> <p>В организациях образования ответственное лицо, вход в систему которого подтверждается ЭЦП, на основании подтверждающих документов вносит данные контингента и персонала, заполняет регистрационные сведения, основные сведения об образовательном процессе (язык обучения, количество кружков и спортивных секций, смен, специальных классов и т.д.), материально-техническую базу (сведения о здании, компьютеризации, библиотеке, столовой и т.д.). Внесенные данные подтверждаются ЭЦП руководителя организации образования.</p> <p>Также среди дополнительных инструментов обеспечения надежности данных можно назвать</p>

		<p>верификацию данных в рамках интеграции с другими государственными информационными системами (более 100 систем), например, с Государственной базой данных физических лиц Министерства Юстиции РК (ГБДФЛ), Государственной базой данных юридических лиц Министерства Юстиции РК (ГБДЮЛ), базой данных о получателях адресной социальной помощи Министерства труда и социальной защиты населения РК, и др. Это позволяет минимизировать ручной ввод данных.</p> <p>Также инструментами обеспечения надежности данных являются встроенные элементы форматно-логического контроля, автоматизируются некоторые показатели и обеспечивая перекрестную верификацию данных и т.д.</p> <p>Эти меры направлены на создание условий для заполняющих, позволяющих максимально качественно заполнить данные в НОБД, и исключить человеческий фактор ошибки данных.</p> <p>Кроме вышеуказанных мер в целях обеспечения надежности данных лицам, ответственным за заполнение НОБД, регулярно оказывается консультационная поддержка по горячей линии и проводится обучение по заполнению НОБД и выгрузке данных.</p>
<p>Ценность</p>	<p>Относится к ценности данных с точки зрения применимости для последующего анализа, извлечению пользы, в том числе для решения прикладных задач.</p>	<p><i>Полностью соответствует</i></p> <p>В целом, НОБД позволила более чем в три раза сократить количество форм административной отчетности (с 467 в 2011 г. до 126 в 2021 г.) и экономит время на сдачу отчетности регионами.</p> <p>Внедрение системы исключило непосредственный контакт между лицами, участвующими в генерации данных, и получателями информации, устранив возможность манипулирования данными.</p> <p>Сократилась отчетность в организациях образования. НОБД предусматривает заполнение форм и внесение изменений (актуализацию данных) ответственными лицами в организациях образования, отчетность же генерируется автоматически системой с извлечением необходимой информации из паспортов в соответствии с установленными формами.</p> <p>С целью исключения дублирования статистики образования, все данные заполняются по ИИН и БИН. Это дало возможность «вычистить»</p>

		<p>статистику образования, исключить дублирование контингента и обеспечить прозрачность и противодействие коррупции в связи с начислением подушевого финансирования на каждого конкретного учащегося.</p> <p>Ввод НОБД значительно сэкономил время на сбор необходимой информации, практически полностью исключил участие учителей в процессе формирования отчетности и статистики. Благодаря развитию НОБД только на бизнес-процессе сдачи отчетности, ежегодная экономия бюджетных средств превысила 750 млн.тенге.</p> <p>Кроме того, НОБД позволил автоматизировать порядка 70% государственных услуг. Доступ к системе имеют все работники сферы образования (организации образования, отделы и Управления образования, МОН РК и ее подведомственные организации), а также статистические органы и другие госорганы в части своих компетенций.</p> <p>По данным НОБД ежегодно публикуется национальный сборник «Статистика системы образования Республики Казахстан», предоставляются данные для Национального доклада и т.д.</p>
--	--	--

**Примечание: все описания характеристик больших данных, приведенные в таблице адаптированы из статьи Б. Кумара [20], дополненных разъяснениями от автора настоящей статьи*

Заключение и обсуждение. Анализ соответствия НОБД основным характеристикам больших данных показал полное соответствие четырем из пяти основных параметров и частичное соответствие по параметру «правдивость» (надежность).

Существует ряд перспективных направлений совершенствования НОБД. Одним из основных является совершенствование структуры данных: существует необходимость расширения показателей НОБД, в особенности непрерывными данными, что существенно позволит расширить аналитические инструменты ИС (к примеру, финансовые данные). Также отдельного внимания заслуживает необходимость нормативного закрепления этических аспектов использования данных НОБД. В настоящее время предприняты меры по защите персональных данных [24] и требованиям информационной безопасности [25, 26]. Еще одним актуальным направлением является повышение доступности данных: в настоящий момент к данным НОБД имеет доступ ограниченное количество авторизованных лиц, использование данных НОБД в исследовательских целях происходит по официальному запросу и нуждается в отдельном регулировании.

В целом НОБД можно оценивать как успешный продукт – среди государственных ИС в области образования на пространстве СНГ, система выгодно отличается от аналогов других стран и опытом создания и оперирования НОБД интересуются представители России, Узбекистана, Кыргызстана и др.

Список использованных источников

1. Collins A., Halverson R. Rethinking education in the age of technology: The digital revolution and schooling in America. – New York: Teachers College Press, 2018. С.167
2. Newhouse C. P., Lane J., Brown C. Reflecting on teaching practices using digital video representation in teacher education //Australian Journal of Teacher Education (Online). – 2007. – Т. 32. – №. 3. – С. 51-62.
3. Kim M. K., Xie K., Cheng S. L. Building teacher competency for digital content evaluation //Teaching and Teacher Education. – 2017. – Т. 66. – С. 309-324.
4. Hashim H. Application of technology in the digital era education //International Journal of Research in Counseling and Education. – 2018. – Т. 2. – №. 1. – С. 1-5.
5. Williamson B. Digital education governance: An introduction //European Educational Research Journal. – 2016. – Т. 15. – №. 1. – С. 3-13.
6. Xie K. et al. Teacher professional development through digital content evaluation //Educational Technology Research and Development. – 2017. – Т. 65. – №. 4. – С. 1067-1103.
7. Reynolds R. Trends influencing the growth of digital textbooks in US higher education //Publishing research quarterly. – 2011. – Т. 27. – №. 2. – С. 178-187.
8. Drijvers P. Digital technology in mathematics education: Why it works (or doesn't) //Selected regular lectures from the 12th international congress on mathematical education. – Springer, Cham, 2015. – С. 135-151.
9. Sindi H. F. A machine learning approach for intelligent tutoring systems //WSEAS Transactions on Systems. – 2005. – Т. 4. – №. 7. – С. 1053-1057.
10. Cox D. D., McLeod S. Social media marketing and communications strategies for school superintendents //Journal of Educational Administration. – 2014 -Т. 52. – №. 6. – С. 850–868.
11. Wang Y. Big opportunities and big concerns of big data in education //TechTrends. – 2016. – Т. 60. – №. 4. – С. 381-384.
12. Tang J. K. T., Xie H., Wong T. L. A big data framework for early identification of dropout students in MOOC //International Conference on Technology in Education. – Springer, Berlin, Heidelberg, 2015. – С. 127-132.
13. Maldonado-Mahauad J. et al. Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses //Computers in Human Behavior. – 2018. – Т. 80. – С. 179-196.
14. Zhong S. H. et al. A computational investigation of learning behaviors in MOOCs //Computer Applications in Engineering Education. – 2017. – Т. 25. – №. 5. – С. 693-705.
15. Kim G. H., Trimi S., Chung J. H. Big-data applications in the government sector //Communications of the ACM. – 2014. – Т. 57. – №. 3. – С. 78-85.
16. Thirathon U. et al. Big data, analytic culture and analytic-based decision making evidence from Australia //Procedia computer science. – 2017. – Т. 121. – С. 775-783.
17. Национальный доклад о состоянии и развитии системы образования РК (по итогам 2018 года) / М. Атанаева, М. Амангазы, Г. Ногайбаева, и др. – Нур-Султан: МОН РК, АО «Информационно-аналитический центр», 2019. – 364 с.
18. F. Diebold. On the Origin(s) and Development of the Term 'Big Data'// PIER Working Paper No. 12-037, September 21, 2012. SSRN: Электронный ресурс.

<https://ssrn.com/abstract=2152421> или <http://dx.doi.org/10.2139/ssrn.2152421>. Дата обращения 21.03.2022

19. D.Laney. 3D-Data Management Controlling Data Volume Velocity and Variety. Application Delivery Strategies. //Stanford: The Meta Group Research Note, 2001

20. Kumar B. An encyclopedic overview of 'big data' analytics //International Journal of Applied Engineering Research. – 2015. – Т. 10. – №. 3. – С. 5681-5705.

21. Z.Sun, K. Strang, R. Li. Big data with ten big characteristics //Proceedings of the 2nd International Conference on Big Data Research. – 2018. – С. 56-61.

22. . Приказ Министра образования и науки Республики Казахстан от 12.11.2014№ 459 «Об утверждении правила осуществления образовательного мониторинга»

23. Приказ Министра от 27.12.2012 г. №570 «Об утверждении форм административной данных в рамках образовательного мониторинга».

24. Закон Республики Казахстан «О персональных данных и их защите» от 24.05.2013 №94-V

25. Постановление Правительства Республики Казахстан от 20.12.2016 № 832 «Об утверждении единых требований в области информационно-коммуникационных технологий и обеспечения информационной безопасности».

26. Стандарт Республики Казахстан СТ РК ISO/IEC 27002-2015 «Информационная технология. Методы и средства обеспечения безопасности. Свод правил по средствам управления защитой информации».