

УДК 37.0

ТЕНДЕНЦИИ ИЗУЧЕНИЯ БОЛЬШИХ ДАННЫХ В РЕГИОНАЛЬНЫХ ИССЛЕДОВАНИЯХ

Оспанова Улжан Абаевна

uljansbox@mail.ru

Докторант факультета международных отношений, кафедры «Регионоведение»,

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

Научный руководитель – Ж. Нурбаев

Введение. Интерес к большим данным в последние годы стал значимым и устойчивым трендом в глобальном научном пространстве. Научный дискурс, связанный с большими данными в социальных науках неоднозначен: с одной стороны, научное сообщество признает невероятный научный и практический потенциал больших данных, с другой, критикует склонность придавать слишком большое внимание техническим элементам анализа при нивелировании роли теоретического обоснования изучаемого феномена.

Ситуация осложняется отсутствием единогласия по поводу однозначного определения термина «большие данные». В научном дискурсе присутствуют различные вариации

трактовки термина [1], где основное семантическое ядро термина содержит не только субстантивную часть - «большие массивы данных», но и дескриптивную, относящуюся к основным характеристикам, и утилитарную компоненты, относящуюся к методологии анализа. Расхождения в онтологии термина можно проследить по трансформации его концептуальных характеристик. В ходе эволюции научного дискурса они расширились от трех «V» - Volume- объем, variety – разнообразие, velocity – скорость, предложенных Д. Лейни [2], до пяти «V» с добавлением таких характеристик как veracity – правдивость и value - ценность, а затем до десяти «V» [3]. Также некоторые авторы продолжают попытки концептуализировать другие свойства больших данных [4].

Поскольку методология работы с большими данными основывается на статистических и вычислительных методах, работающих в основном с количественными данными и измеримыми данными, научная результативность социальных наук, методологический арсенал которых не предусматривает генерацию таких типов данных в больших количествах, закономерно ограничена.

Актуальность настоящего исследования обусловлена растущим научным интересом к изучению больших данных, а новизна – малым количеством научных работ, посвященных теме исследования. Исходя из точки зрения, что этот сравнительно новый концепт все еще находится в процессе осмысления, мы предлагаем сфокусироваться не на дебатах относительно концепта и его роли в науке, а на реальной картине изучения больших данных в региональных научных исследованиях с отдельным фокусом на исследования в области международных отношений и политических исследованиях. Данная статья рассматривает статистику научных публикаций в области региональных исследований, связанных с использованием или изучением больших данных, с целью определения тенденций в общем объеме научных публикаций, тематических областях и основных направлений научного дискурса в области международных отношений и политических исследований за последние пять лет.

Методология и данные. В данном исследовании были использованы данные, генерируемые в рамках базы данных научных публикаций Web of Science, которая является одним из наиболее авторитетных собраний научных источников. Анализ проводился по информационному сдвоенному запросу «регион» и «большие данные» на английском языке во встроенной поисковой системе базы данных Web of Science Core Collection. В результате были собраны метаданные научных публикаций по данному запросу за пятилетний период 2017-2021 гг., общее количество составило 8400 научных публикаций.

Метаданные о научных публикациях включили 69 видов библиометрических сведений о публикациях, таких как данные об авторах публикаций, информацию об издании, цитировании, и самой публикации.

Критериями анализа общего объема данных выступили дескриптивная статистика цитирования, количества публикаций, тематические области науки, страновая принадлежность.

Отдельный фокус при анализе статей в научных областях международных отношений и политических наук был сделан на тематический контент анализ. Контент-анализ, используемый в данном исследовании, использует рекомендации Кохен, Манион и Моррисон [5] и состоит из нескольких этапов: выборка текстов; определение единиц анализа, кодов и категорий; интерпретация результатов. были выделены первоначальные коды, которые были объединены в категории для последующего контент анализа [6]. Для сохранения качества и достоверности анализа, каждый этап процесса прочтения текста, кодирования и обобщения информации в категории был пройден множество раз и в разной последовательности, а на его финальной стадии выделенные под-темы были обобщены в крупные категории.

Таким образом в настоящем исследовании мы попытались ответить на следующие исследовательские вопросы: 1) каков общий объем научной производительности в области изучения больших данных в региональных исследованиях? 2) к каким разделам науки относятся такие научные исследования? 3) какие страны наиболее продуктивны в области проведения таких исследований? 4) основные темы таких исследований в научных областях международных отношений и политических наук.

Результаты. За исследуемый период (2017-2022 гг.) наблюдается устойчивая тенденция роста научной производительности в области изучения больших данных в региональных исследованиях (Рис. 1). Мы можем наблюдать, что ежегодный прирост количества издаваемых публикаций составляет около 17%, при этом наибольший прирост отмечен в 2018 г., а наименьший в 2021 г. При этом ежегодный прирост неравномерен.

Цитируемость, как один из показателей научной производительности, демонстрирует более равномерную динамику роста, а ее экстремумы не совпадают в динамике с ростом количества публикаций: по данной категории наибольший годовой пророст произошел в 2021 г., а наименьший – в 2018 г.

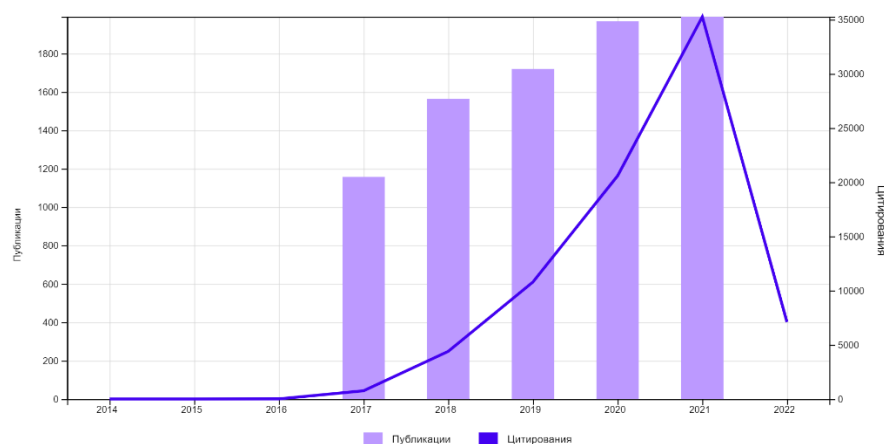


Рисунок 1 – Динамика научной производительности в области изучения больших данных в региональных исследованиях

Анализ научной производительности представлен в разрезе областей науки представлен в древовидной диаграмме на Рисунке 2. Он показывает, что наибольшее количество исследований проводится в областях, связанных с ИКТ и компьютерными науками. Также высокая научная производительность наблюдается в области междисциплинарных исследований. Третья по производительности научная область может быть объединена в общую категорию естественных наук, со значительным преобладанием исследований в области экологии, наук об окружающей среде и устойчивости. При анализе данных в разрезе отдельных годов исследуемого периода общая картина осталась в целом аналогичной, и также характеризуется доминированием технических наук.



Рисунок 2 – Топ 20 областей науки по научной производительности в региональных исследованиях о больших данных.

Наибольшую научную продуктивность в разрезе страновой представленности (Рис. 3) за исследуемый период продемонстрировали Китай (3653 публикации), США (1975 публикаций), и Англия (745 публикаций). На три эти страны приходится почти половина всех региональных исследований по большим данным. Примечательно, что Казахстан по научной продуктивности за исследуемый период занял 65 место из 169 стран - 30 публикаций.

Такие дисциплинарные области науки как международные отношения и политология продемонстрировали невысокую производительность: количество относящихся к ним публикаций за исследуемый период составило всего 13 и 9 публикаций соответственно.

Тематический контент-анализ научных публикаций в области политологии выявил следующие области: коррупция, конфликты, политическая репрезентативность в выборных процессах. К примеру, Б. Мадлович и Б. Магьяр представили исследования четырех уровней коррупции посткоммунистических стран на основе международного индекса восприятия коррупции Transparency International (CPI) и предложили уточнения индекса путем использования больших данных на примере Венгрии [7]. Н. Захаров исследовал взаимосвязь между коррупцией и инвестициями в основной капитал в условиях высокой коррумпированности с применением методов анализа больших данных [8]. Д. Канаш и соавторы провели исследование по 84 регионам Южной и Северной Америки о восприятии коррупции с учетом эмпирического опыта и психологической предрасположенности с применением методов анализа больших данных [9]. Ю. Костюченко и М. Ющенко рассмотрели подходы к использованию больших данных (контента социальных сетей) для понимания социального поведения в зонах конфликтов и анализа динамики незаконных вооруженных формирований [10]. О.Дж Уолтер и соавторы представили новый показатель пространственной динамики конфликтов (SCDi) на основе комбинации точечного анализа и территориально-пространственного анализа на примере Северной и Западной Африки [11].

А. Михалев представил исследование споров о разделе воды (трансграничные реки) во Внутренней Азии с использованием методов анализа больших данных, затронув вопрос политизации экологических проблем и перспективу изменения баланса сил в регионе [12]. Дж. Креспо провел тестирование теории партийных предпочтений типов избирательных систем в региональных парламентах Испании, заключив, что ожидания относительно партийных предпочтений в отношении создания институциональных правил не должны сводиться к совокупному размеру партии [13]. Ш. Уордани и В. Субекти посвятили исследование представленности женщин из политических династий, избранных в национальный парламент Индонезии [14].



Рисунок 3 – Научная продуктивность в разрезе стран

Тематический контент-анализ научных публикаций с использованием больших данных в области региональных исследований в международных отношениях позволил выявить неоднородность затрагиваемых тем, обобщение которых имеет смысл только на уровне под-отраслей науки и отдельных общих категорий. Можно выделить под-тему региональных интеграционных объединений и их отдельных аспектов, в частности - инициативу «Один пояс – один путь» [15, 16], НАТО [17], ЕС [18], СААРК [19], и др. Также можно выделить под-тему геопространственных данных [11, 17, 20-22], при этом важно отметить, что в такие исследования чаще всего носят междисциплинарный характер и часто связаны с другими областями науки, такими как экология, география и другие схожие области. Кроме этого, выделяется под-тема экономики и ее отдельных аспектов [23-26]. И поскольку традиционно международные отношения связаны с внешней политикой, отдельной крупной тематической областью выступает внешняя политика, которую так или иначе затрагивают практически все работы в данной категории.

Заключение и обсуждение. В данной статье приведен краткий анализ публикаций, посвященных изучению больших данных в региональных исследованиях и собранных из базы данных Web of Science Core Collection за период 2017-2021 гг. Анализ библиометрических показателей 8400 научных публикаций, относящихся к данной категории выявил тенденцию устойчивого роста популярности данной темы по двум основным показателям научной продуктивности – количеству публикуемых материалов и цитируемости. При этом, если прирост количества происходил неравномерно, то цитируемость демонстрирует более ровную и значимую в абсолютном выражении

динамику роста из года в год. Самый низкий прирост в количестве научных публикаций произошел в 2021 г., а причиной такового предположительно стала пандемия COVID-19, когда научное сообщество, как и многие отрасли, было вынуждено адаптироваться к многочисленным ограничениям. Влияние пандемии на научную производительность отмечалось авторами исследований, однако изучение этого феномена требует отдельного внимания.

Изучение больших данных в региональных исследованиях в разрезе дисциплинарных областей науки характеризуется значительным превалированием так называемых технических дисциплин. Совокупно на различные отрасли компьютерных и телекоммуникационных наук приходится большая часть научных публикаций. Социальные науки значительно отстают, и представлены в большей части исследованиями, связанными с геоданными, в области наук об окружающей среде. Одной из основных причин такого разрыва в научной продуктивности предположительно выступают отсутствие доступности данных, как отмечалось выше, в том числе за счет того, что предметная область и методологические подходы, превалирующие во многих социальных науках, не предусматривают генерацию больших объемов разнообразных и непрерывных данных. Также одной из важных причин может выступать то, что методы анализа больших данных требуют от исследователя владения навыками кодировки и программирования вычислений, мало или не изучаемых в рамках традиционных академических программ социальных наук.

В разрезе страновой представленности с большим отрывом лидирует Китай, что связано как с демографическими характеристиками КНР, так и с значительными усилиями Китая в области цифровизации всех сфер жизнедеятельности человека, которая в результате повысила доступность больших данных.

Более детальный фокус, направленный на тематический контент-анализ изучения больших данных в региональных исследованиях в области политических наук и международных отношений позволил выявить такие крупные темы как: коррупция, конфликты, политическая репрезентативность в выборных процессах, аспекты деятельности в рамках региональных интеграционных объединений, изучение геопространственных данных, экономика и внешняя политика.

При этом важно отметить, что многие из таких исследования в области социальных наук носят междисциплинарный характер. Еще одной выявленной тенденцией в категории социальных наук стало наблюдение о том, что нередко авторы, затрагивают тему больших данных косвенно и описательно, при этом не используя их в качестве эмпирических данных для собственной исследовательской работы.

Список использованных источников

1. 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
2. D.Laney. 3D-Data Management Controlling Data Volume Velocity and Variety. Application Delivery Strategies. //Stanford: The Meta Group Research Note, 2001
3. A. Gandomi, M. Haider. Beyond the hype: Big data concepts, methods, and analytics //International journal of information management. – 2015. – Т. 35. – №. 2. – С. 137-144.
4. 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.
5. L. Cohen, L. Manion, and K. Morrison, Research method in education. London and New York: Routledge: Taylor and Francis Group, 2007. 638 с.

6. Ford J. M. Content analysis: An introduction to its methodology //Personnel Psychology. – 2004. – Т. 57. – №. 4. –1110 с.
7. Magyar B., Madlovics B. From petty corruption to criminal state: a critique of the Corruption Perceptions Index as applied to the post-Communist region //Intersections. East European Journal of Society and Politics. – 2019. – Т. 5. – №. 2.
8. Zakharov N. Does corruption hinder investment? Evidence from Russian regions //European Journal of Political Economy. – 2019. – Т. 56. – С. 39-61.
9. Canache D. et al. Who sees corruption? The bases of mass perceptions of political corruption in Latin America //Journal of Politics in Latin America. – 2019. – Т. 11. – №. 2. – С. 133-160.
10. Kostyuchenko Y. V., Yuschenko M. Toward Approaches to Big Data Analysis for Terroristic Behavior Identification: Child Soldiers in Illegal Armed Groups During the Conflict in the Donbas Region (East Ukraine) //International Journal of Cyber Warfare and Terrorism (IJCWT). – 2017. – Т. 7. – №. 1. – С. 1-12.
11. Walther O. J. et al. Introducing the Spatial Conflict Dynamics indicator of political violence //Terrorism and Political Violence. – 2021. – С. 1-20.
12. Михалев А. В. Региональный порядок и трансформация трансграничного водного режима в современной Внутренней Азии //Сравнительная политика. – 2019. – Т. 10. – №. 4. – С. 155-167.
13. Crespo J. C. Qué sistemas electorales prefieren los partidos? El caso de los parlamentos regionales en España (1977-1983) //Revista de estudios políticos. – 2017. – №. 175. – С. 67-110.
14. Wardani S. B. E., Subekti V. S. Political dynasties and women candidates in Indonesia's 2019 election //Journal of Current Southeast Asian Affairs. – 2021. – Т. 40. – №. 1. – С. 28-49.
15. Layton P. Artificial intelligence, big data and autonomous systems along the belt and road: towards private security companies with Chinese characteristics? //Small Wars & Insurgencies. – 2020. – Т. 31. – №. 4. – С. 874-897.
16. Huo, D., Ouyang, R., Sun, B., Wei, L., Hung, K., & Wang, H. (2019). Complex network of aviation e-services in the belt and road initiative: a heuristic study of small data based on block modeling. *Emerging Markets Finance and Trade*, 55(14), 3151-3165.
17. Homberg, M. V. D. (2017). Toward a Balkans' Data for Disaster Management Collaborative?. In *Implications of climate Change and disasters on Military Activities* (pp. 11-18). Springer, Dordrecht.
18. Raźniak P., Winiarczyk-Raźniak A., Nowotnik D. Central and Eastern European Cities in globalized world. – 2015.
19. Галищева Н. В. Продовольственная безопасность в Южной Азии: основные проблемы и пути решения //Вестник МГИМО Университета. – 2018. – №. 1 (58). – С. 148-168.
20. Halliday W. D. et al. Tourist vessel traffic in important whale areas in the western Canadian Arctic: Risks and possible management solutions //Marine Policy. – 2018. – Т. 97. – С. 72-81.
21. Rumson A. G., Hallett S. H., Brewer T. R. Coastal risk adaptation: the potential role of accessible geospatial Big Data //Marine Policy. – 2017. – Т. 83. – С. 100-110.
22. Latifi R. et al. Telemedicine for Disaster Management: Opportunities, Challenges, and Requirements //A Multinational Telemedicine Systems for Disaster Response: Opportunities and Challenges. – IOS Press, 2017. – С. 112-129.

23. Yu L. et al. Trade direction, market access and regional wage inequality: An explanation based on asymmetric trade costs //The World Economy. – 2022. – Т. 45. – №. 3. – С. 895-913.
24. Layton P. Artificial intelligence, big data and autonomous systems along the belt and road: towards private security companies with Chinese characteristics? //Small Wars & Insurgencies. – 2020. – Т. 31. – №. 4. – С. 874-897.
25. Карминский А. М., Хон О. Д. Факторы залогового обеспечения для управления рисками банков: региональный аспект //Вестник МГИМО Университета. – 2018. – №. 1 (58). – С. 169-185.
26. Ermolin I., Svolkinas L. Assessment of the sturgeon catches and seal bycatches in an IUU fishery in the Caspian Sea //Marine Policy. – 2018. – Т. 87. – С. 284-290.