

ИСПОЛЬЗОВАНИЕ ВОЗМОЖНОСТЕЙ «БОЛЬШИХ ДАННЫХ» В КРИМИНАЛИСТИКЕ

В настоящей статье рассматриваются возможности и перспективы использования в криминалистических целях общедоступной информации из открытых источников (открытые данные), как структурированной, так неструктурированной. Автор обосновывает необходимость криминалистического использования ресурсов «больших данных» для решения криминалистических и иных задач оперативно-розыскной деятельности и уголовного судопроизводства. Также обозначены возможности аналитического исследования общедоступной информации для решения идентификационных и диагностических задач криминалистики, дан примерный перечень открытых ресурсов информации для обнаружения/использования необходимых данных, направлений и способов обращения с ними, перечислены уже применяемые при раскрытии и расследовании преступлений оперативные и процессуальные формы работы с общедоступными данными сети Интернет.

Ключевые слова: «большие данные», цифровая криминалистика, общедоступная информация, открытые данные, электронно-цифровые следы.

S.A. Mashkov

USING THE OPPORTUNITIES OF «BIG DATA» IN CRIMINALISTIK

The article describes the possibilities and prospects of using publicly available information from open sources (open data), both structured and unstructured, for forensic purposes. The author substantiates the need for the criminalistic use of «big data» resources to solve the tasks of operational investigative activities and criminal proceedings. To this end, designates the possibilities of analytical research of publicly available information to solve identification and diagnostic tasks of criminalistic, gives an approximate list of open information resources for the detection/use of necessary data and search methods for handling them, lists the operational and procedural forms of work with publicly available Internet

© Машков С.А., 2024

Keywords: «big data», digital criminalistic, publicly available information, open data, electronic digital traces.

В настоящее время информация, под которой законодательство Российской Федерации понимает «сведения (сообщения, данные) независимо от формы их представления»¹, – это одна из самых важных ценностей. Информация позволяет получать преимущество, принимать верные решения, основанные на этих данных, и действовать на опережение, в то время пока остальные находятся в неведении. Данный постулат актуален и в сфере борьбы с преступностью, особенно при решении задач оперативно-розыскной деятельности, к которым, в частности, относятся: «выявление, предупреждение, пресечение и раскрытие преступлений, а также выявление и установление лиц, их подготавливающих, совершающих или совершивших»². В определенной степени это характерно и для уголовного преследования, то есть «процессуальной деятельности стороны обвинения в целях изобличения подозреваемого, обвиняемого в совершении преступления»³.

Современные цифровые реалии и массив знаний, накопленный разными науками, позволяет не только и не столько получать новую (ранее неизвестную) информацию о фактах и событиях, но куда более значима, точна и содержательна стала информация «старая», то есть уже известная и проверенная. Эта информация накоплена, систематизирована и может быть изучена по любым интересующим критериям, что позволяет характеризовать (как диагностически, так и идентификационно) любые объекты, явления и процессы, делать обоснованные выводы и строить точные прогнозы, то есть формировать «производную», «выводную», «дополнительную» информацию.

Цифровой формат информации снял количественные и качественные ограничения её хранения и обработки. Уже большая часть

¹ Об информации, информационных технологиях и о защите информации : федер. закон от 27 июля 2006 г. № 149-ФЗ // Собрание законодательства Российской Федерации. 2006. № 31.1. Ст. 3448. (Ст. 2, п. 1.)

² Об оперативно-розыскной деятельности : федер. закон от 12 августа 1995 г. № 144-ФЗ // Собрание законодательства Российской Федерации. 1995. № 33. Ст. 3349. (Ст. 2.)

³ Уголовно-процессуальный кодекс Российской Федерации : федер. закон от 18 декабря 2001 г. № 174-ФЗ // Там же. 2001. № 52 (ч. I). Ст. 4921. (Ст. 5, п. 55.)

зафиксированной информации «оцифрована», а массивы информации растут высокими темпами. Если в начале нулевых годов говорили о ежегодном удвоении объема информации, то в 2020 году этот показатель составлял уже ежегодное увеличение зафиксированной информации в 5 раз. В следующие пять лет информация достигнет цифры, которая в 250 раз превысит количество песчинок песка на всех пляжах мира [1], и составит 175 Зеттабайт [2]. Информации становится всё больше, при этом она разнообразна по содержанию и формам представления: текстуальная, графическая, фото, видео, идентификационная, координатная и многая другая.

Цифровые технологии привели к тому, что каждый человек, используя современные многофункциональные коммуникативные устройства и технологии ввода-вывода информации (смартфоны, компьютеры, планшеты, навигаторы, видеокамеры, сеть Интернет, банковские карточки и др.), сам постоянно продуцирует информацию в цифровой реальности как о себе самом, так и об окружающей действительности. Это же правило действует и при осуществлении преступной деятельности на всех ее стадиях, в процессе которой преступники оставляют множество следов, в том числе, следы цифровые.

Доступ к информации (возможность получения информации и ее использования¹) при этом всё более упрощается и становится «открытым», то есть свободным, ничем не ограниченным и позволяющим использовать ее любым способом и в любых целях.

Законодательство России использует понятия «общедоступная информация» и «открытые данные», которые определяются как общеизвестные сведения и иная информация, доступ к которой не ограничен, в том числе, информация, размещаемая ее обладателями в сети Интернет в формате, допускающем автоматизированную обработку без предварительных изменений человеком в целях повторного ее использования (общедоступная информация, размещаемая в форме открытых данных)².

¹ Об информации... (п. 6 ст. 2.)

² Об информации... (ч. 1 и ч. 4 ст. 7).; «Типовые условия использования общедоступной информации, размещаемой в информационно-телекоммуникационной сети «Интернет» в форме открытых данных» (утв. протоколом заочного голосования Правительственной комиссии по координации деятельности открытого правительства от 19 сентября 2016 г. № 6) // СПС «КонсультантПлюс».

То есть, как отмечает А.И. Савельев [4, с. 73], законодательно установлена «...презумпция открытости информации: общедоступной является любая информация, кроме той, к которой ограничен доступ. Отнесение информации к категории общедоступной в самом общем виде означает, что любое лицо без указания причин и целей может получать такую информацию и использовать по своему усмотрению...». Сеть Интернет является неограниченным ресурсом открытых данных, которая является доступной и допускающей ее автоматизированную обработку.

С 2008 года стал применяться в мире и сегодня широко используется в России термин «большие данные» (англ. «big data»), которым обозначают большие массивы данных, имеющих такие характеристики как объем, разнообразие, скорость обработки и/или вариативность, и требующие использования технологии масштабирования для эффективного хранения, обработки, управления и анализа. Массивы данных представляют собой идентифицируемые совокупности данных, к которым можно получить доступ или скачать в одном или нескольких форматах¹. При этом «данные» понимаются как представление информации в формализованном (формальном) виде, пригодном для коммуникации, передачи, интерпретации или обработки людьми и компьютерами².

Большие данные могут быть структурированными и неструктурированными, их невозможно эффективно обработать с использованием традиционных методов, и для этого применяют специальные технологии и программное обеспечение: системы распределенных хранилищ данных; технологии обработки потоков данных; методы машинного обучения; алгоритмы анализа, позволяющие определять ценность и значение информации и извлекать её из огромных объемов данных [5; 6].

Следовательно, «большие данные» являются общедоступной информацией, размещаемой в форме открытых данных. «При этом огромные объемы информации можно использовать

¹ Национальный стандарт Российской Федерации (ГОСТ Р ИСО/МЭК 20546-2019) «Информационные технологии. Большие данные. Обзор и словарь.» // Москва. Стандартинформ. 2020. (п. 3.1.2, п. 3.1.5, п. 3.1.11.)

² Межгосударственный стандарт (ГОСТ 33707-2016 (ISO/IEC 2382:2015)). Информационные технологии. Словарь. // Межгосударственный совет по стандартизации, метрологии и сертификации. 2016. (п. 4.259.)

для решения задач, требующих высокой точности прогнозов, поиска обоснований для тех или иных решений, персонализации и так далее» [6].

К ресурсам такого рода получения информации практика относит следующие [7, с. 7; 8, с. 63–64]:

- информационные материалы (интервью, статьи, новости, заметки) в средствах массовой информации (газеты, журналы, радио, телевидение);

- научные исследования, опубликованные в специализированных изданиях; профессиональные и академические отчёты, конференции, доклады; книги, энциклопедии, справочники, ме-муары и т.д.;

- социальные сети (данные, фото, посты, лайки, группы, комментарии и т.д.), веб-сообщества и контент, созданный пользователями (видеохостинги, вики-справочники, блоги, веб-форумы);

- информация из переписи; публичные отчёты правительства, официальные данные о бюджетах, демографии и т.д.;

- документы из открытых государственных и негосударственных архивов; информация из открытых реестров (прав на недвижимое имущество, юридических лиц, индивидуальных предпринимателей и т.п.);

- публичные коммерческие данные (доход, прибыль, убыток, рост, стоимость акций и т.д.);

- результаты публичных опросов;

- материалы пресс-конференций, различные публичные заявления; данные со спутников дистанционного зондирования Земли и самолетов аэрофотосъемки, радиомониторинг;

- опубликованные полицейские и судебные документы и другие источники; массовых утечки из сервисов и социальных сетей;

- метаданные файлов (дата создания документа, имена пользователей, модели принтеров, программное обеспечение, установленное на компьютерах, иногда геолокация);

- непубличная документация, находящаяся в открытом до-ступе;

– данные о домене, e-мейлы, телефоны, факсы, технологии, на которых построен сайт, криптографические сертификаты, суб-домены;

– индексация интернет-оборудования (сервера, роутеры, камеры видеонаблюдения, вебкамеры, онлайн-накопители и т. д.).

Способами обращения с «большими данными» называют [7, с. 9]:

– сбор информации (в том числе по фотографиям) из открытых поисковых систем; анализ пользовательской активности в социальных сетях и блогах, на форумах, иных виртуальных платформах;

– поиск открытых персональных данных пользователей в социальных сетях, мессенджерах; просмотр сохраненных копий сайтов в поисковых системах, интернет-архивах;

– получение геолокационных данных с помощью общедоступных ресурсов вроде «Google Maps», «Яндекс.Карты» и других;

– сбор данных на ресурсах, доступ к которым возможен только по подписке; применение специализированных сервисов и программ; использование сервисов, сканирующих приложения, файлы или сайты на наличие вредоносного кода.

Цифровые технологии дали возможность сделать максимально удобными процессы поиска, сбора, формирования, хранения, обработки, обновления, передачи/получения и представления информации, а также позволили алгоритмизировать их, вводить любую категориальность, минимизировать вероятность ошибок, сократить время обработки данных, повысить точность и достоверность обобщений и выводов. Главное, что позволяют технологии – это быстро и эффективно работать с большими практически неограниченными объемами информации. При этом, чем больше информации и чем она разнообразнее, тем точнее будут определены интересующие закономерности.

Таким образом, современное состояние науки и техники делают возможным применять технологии и оборудования, позволяющие:

- 1) получать большие объемы информации;
- 2) из любых источников;

- 3) в любых форматах;
- 4) обрабатывать их любыми методами и способом;
- 5) использовать при этом любые критерии (отдельные и комплексные);
- 6) очень быстро и
- 7) представлять результаты в любой форме (текстуальной, графической и иных).

Такие возможности на сегодняшний день активно и очень результативно используются и объектом интереса и исследований выступают данные открытого доступа. Неограниченным ресурсом информации является сеть Интернет.

Криминалистическая деятельность – это деятельность по работе с информацией: остаточными явлениями преступного события, представленной в виде идеальных и материальных следов, к числу которых относятся и электронно-цифровые следы. Игнорирование возможностей такого неограниченного ресурса как общедоступная информация (большие данные) является недопустимым. Представляется, что перспективы вовлечения «больших данных» в практику раскрытия и расследования преступлений весьма широки. Криминалистика – «живая» наука, она развивается и «эволюционирует», ориентируется на практику, которая определенно испытывает потребность использования возможностей анализа «больших данных».

Правоприменительная практика по борьбе с преступностью также «движется» в этом направлении. Уже сегодня поисково-информационные действия в ходе оперативно-розыскной деятельности эффективно используют эти возможности и ведется поиск разнообразных по форме и содержанию открытых данных в сети Интернет при осуществлении оперативно-розыскных мероприятий (наведение справок, наблюдение, отождествление личности, получение компьютерной информации, снятие информации с технических каналов связи и др.). Оперативные и криминалистические учеты формируются в электронной форме, региональные учеты объединяются в единые базы данных, создаются поисково-аналитические системы, предусматривается возможности поиска информации в базах данных разных учетов по единому критерию и т.д.

В рамках уголовного судопроизводства привлекаются специалисты по информационным и компьютерным технологиям (консультации, заключения специалистов, допросы специалистов и т.д.). Расширяется объектный и предметный перечень «компьютерных» экспертиз: компьютерно-техническая, аппаратно-компьютерная, программно-компьютерная, компьютерно-сетевая, информационно-компьютерная экспертизы, которые исследуют соответствующие оборудование, технологии, данные и информационное содержание компьютерных систем и носителей [9, с. 17–19].

Также проводятся информационно-аналитические экспертизы, объектами которых являются любые структурированные и/или неструктурированные массивы данных, а предметом выступает анализ цифровых массивов данных, содержащих сведения о деятельности цифровых систем, устройств, отдельных индивидуумов, в целях поиска взаимосвязей отдельных элементов указанных массивов данных. При этом практическая необходимость и законодательная возможность проведения данного вида экспертного исследования бесспорна.

При проведении данных экспертиз эксперт, в том числе, исследует и общедоступную информацию (открытые данные), находящуюся в сети Интернет, что может быть обусловлено применяемой им методикой данного вида экспертиз, экспертным заданием [10, с. 107–108] или практической необходимостью, обусловленной предметом и объектом информационно-аналитической экспертизы, так как информация подлежит обнаружению и сбору, что может методологически обеспечить только сам эксперт. О возможности и необходимости использования данных, находящихся в открытых источниках, как объекта экспертизы, указывает А. Г. Себякин [11, с. 98–100], характеризуя их как общедоступные и легальные информационно-цифровые объекты и их копии.

Конечно, эти шаги органов следствия и оперативно-розыскной деятельности очень незначительны, и возможности исследования открытых данных остаются мало реализованы, но процесс идёт. Современные информационные реалии обуславливают необходимость криминалистического осмысления и ши-

рокого практического применения исследовательских возможностей «больших данных» в целях раскрытия и расследования преступлений. Перспективы использования «больших данных» в криминалистике неограниченны, как и объемы общедоступной информации. Практическая востребованность бесспорна.

Потенциальные возможности экспертных оценок и экспертиз огромных массивов информации, находящейся в открытом доступе, открывают большие перспективы для аналитики и прогнозирования при осуществлении оперативно-розыскной и уголовно-процессуальной видов деятельности. Ресурсы «больших данных» необходимо использовать в криминалистике, применять аналогичный опыт иных сфер деятельности (экономика, политика), системно привлекать лиц, обладающих специальными знаниями в информационно-аналитической деятельности, расширять практику назначения и проведения соответствующих видов судебных экспертиз.

Список использованной литературы

1. Рыжов В. От «информационного взрыва» к «информационной депрессии»: неучтенные риски четвертой ИТ-революции. Часть первая / В. Рыжов // Комсомольская правда. 2021. 26 ноября. – URL: <https://www.kp.ru/daily/28361/4509297/?ysclid=lwc4n-pecjw730301940>.

2. 175 Zettabytes by 2025. – URL: <https://www.forbes.com/sites/tomcoughlin/2018/11/27/175-zettabytes-by-2025/#b5fe21054597>.

3. Рожкова М.А. «Общедоступная информация», «открытые данные» и «персональные данные, разрешенные субъектом для распространения» – что это такое и как они между собой связаны? / М.А. Рожкова // Закон.ру. 2021. 13 января. – URL: https://zakon.ru/blog/2021/1/13/obschedostupnaya_informaciya_otkrytye_dannye_i_personalnye_dannye_razreshennye_subektom_dlya_raspros.

4. Савельев А.И. Комментарий к Федеральному закону от 27 июля 2006 г. № 149-ФЗ «Об информации, информационных технологиях и защите информации» (постатейный) / А.И. Савельев. – Москва : Статут, 2015. – 320 с.

5. Preimesberger, Chris Hadoop, Yahoo, 'Big Data' Brighten BI Future / EWeek (2011. 15 августа). – URL <https://www.eweek.com/storage/hadoop-yahoo-big-data-brighten-bi-future>.

6. Наташкин А. Что такое Big Data: как собирают и где применяют большие данные? / А. Наташкин. – URL: <https://lenta.ru/articles/2023/11/27/chto-takoe-big-data/?ysclid=lwccvcc6or627380342>.

7. Дворянкин О.А. Osint, pentest и нетсталкинг – информационные технологии интернета / О.А. Дворянкин // Национальная ассоциация ученых. – 2022. – № 84. – С. 6–13.

8. Иванов В.Ю. Использование OSINT в раскрытии и расследовании преступлений / В.Ю. Иванов // Вестник Уральского юридического института МВД России. – 2023. – № 1. – С. 62–66.

9. Смолина А.Р. Методическое и алгоритмическое обеспечение производства компьютерно-технической экспертизы: дис. ... канд. техн. наук : 05.13.19 / А.Р. Смолина. – Томск, 2017. – 132 с.

10. Егоров Н.Н. Использование специальных знаний: история, состояние и перспективы развития / Н.Н. Егоров, А.А. Протасевич // Сибирские уголовно-процессуальные и криминалистические чтения. – 2022. – № 1. – С. 101–117.

11. Себякин А.Г. Данные, находящиеся в открытых источниках информации, как объект судебной информационно-аналитической экспертизы / А.Г. Себякин // Сибирские уголовно-процессуальные и криминалистические чтения. – 2024. – № 2. – С. 95–103.

Информация об авторе

Машков Сергей Александрович – кандидат юридических наук, доцент кафедры криминалистики, судебных экспертиз и юридической психологии, Институт юстиции, Байкальский государственный университет, г. Иркутск, Российская Федерация, e-mail: msa325@mail.ru.

Author

Mashkov Sergey Alexandrovich – Ph.D. of Law, Associate Professor of the Department of Criminalistics, Forensic Examinations and Legal Psychology, Institute of Justice, Baikal State University, Irkutsk, the Russian Federation, e-mail: msa325@mail.ru.