ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ НА ПРОСТРАНСТВЕ "БОЛЬШИХ ДАННЫХ"

DOI: 10.36724/2072-8735-2022-16-4-21-28

Статьев Вячеслав Юрьевич,
"Фонд транспортной безопасности", Москва, Россия,
articleglory@yandex.ru

Докучаев Владимир Анатольевич,
МТУСИ, Москва, Россия, ***v.a.dokuchaev@mtuci.ru***

Маклачкова Виктория Валентиновна,
МТУСИ, Москва, Россия, ***v.v.maklachkova@mtuci.ru***

Manuscript received 12 March 2022;
Accepted 05 April 2022

Ключевые слова: *данные, информация, знания, DIKW-модель, большие данные, информационная безопасность, информационная система, цифровая трансформация, цифровой двойник, обработка данных, информационная граница*

Результатом исследования, проведенного в данной работе, является разработка основ нового подхода к проблематике информационной безопасности в контексте технологий "больших данных". Цель получения и обработки сведений о предметной области состоит в их последующем использовании для формирования целенаправленных воздействий на объект управления. Наблюдение и осознание происходящего с объектом ранее осуществлялось через некоторую систему информатизации деятельности субъекта, т.е. опосредовано, а реализация принятого решения осуществлялась в рамках объекта управления, т.е. напрямую. В условиях цифровой трансформации воздействие на объект управления осуществляется не напрямую, а на его цифрового двойника с последующим редуцированием, при необходимости, этого воздействия на физический объект управления. Таким образом, появляется система цифровизации деятельности субъекта. При создании таких систем одним из основных принципов выступает принцип системной безопасности (обеспечение информационной безопасности). Однако его реализация затрудняется тем, что технологии "больших данных" связаны не столько с их большими объемами, сколько с необходимостью обработки разнообразных консолидированных данных, имеющих различную структуру, схемы классификации и индексации. В связи с этим предлагается проанализировать DIKW-модель и сущности входящих в неё понятий, что позволит перейти к DIKW-обработке при использовании технологий "больших данных". При этом возникает необходимость решения двух проблем: оценка влияния степени консолидации данных, информации и знаний на возрастание их важности в обеспечении информационной безопасности субъекта пространства "больших данных" и оценка критической массы данных, информации и знаний, доступных в пространстве "больших данных", с точки зрения их комплексной совместной обработки и получения на их основе более важных сведений. Для установления необходимого баланса между потребностью в информационном обмене и допустимыми ограничениями распространения данных, информации и знаний вводится новое виртуальное понятие - "информационная граница", а вопрос обеспечения информационной безопасности сводится к задаче прагматически обоснованного ограничения доступа для всех уровней DIKW-модели с введением понятия "DIKW-безопасность". В данном исследовании предложен интенсиональный подход для контроля "информационных границ" в процессе использования технологий "больших данных", что обеспечивает переход от статических решений вопросов обеспечения информационной безопасности субъекта к реализации динамических процедур такого контроля с использованием элементов искусственного интеллекта, мониторинга этого пространства и процедур прогнозирования своего "места" в пространстве "больших данных".

Информация об авторах:
Статьев Вячеслав Юрьевич, *к.т.н., эксперт "Фонд транспортной безопасности", Москва, Россия*
Докучаев Владимир Анатольевич, *д.т.н., профессор, заведующий кафедрой "Сетевые информационные технологии и сервисы" МТУСИ, Москва, Россия*
Маклачкова Виктория Валентиновна, *старший преподаватель кафедры "Сетевые информационные технологии и сервисы" МТУСИ, Москва, Россия*

Для цитирования:
Статьев В.Ю., Докучаев В.А., Маклачкова В.В. Информационная безопасность на пространстве "больших данных" // T-Comm: Телекоммуникации и транспорт. 2022. Том 16. №4. С. 21-28.

For citation:
Statev V.Yu., Dokuchaev V.A., Maklachkova V.V. (2022) Information security in the big data space. *T-Comm*, vol. 16, no.4, pp. 21-28. (*in Russian*)

Введение

Процесс информатизации мирового сообщества развивается стремительно и зачастую непредсказуемо [1]. Информатизация ведет к созданию единого мирового информационного пространства, в рамках которого производится накопление, обработка, хранение и обмен данными (информацией) между субъектами этого пространства – личностями, обществом и государством. Прогнозы экспертов показывают, что объём «цифровой вселенной» к 2025 г. может превысить 40 зеттабайт, к 2023 г. стоимость индустрии «больших данных» (англ. Big Data) вырастет примерно в 5 раз – до 54,3 млрд. долл. [2,3]. Интернет-трафик (объём информации), передаваемый через Сеть в течение одного часа в 2020 г., равнялся всему трафику, передаваемому через Интернет в 2008 г. Прямые инвестиции в цифровую трансформацию организаций (англ. DX) растут с совокупным годовым темпом роста (англ. CAGR) 15,5% в период с 2020 по 2023 гг. и, как ожидается, достигнут 6,8 трлн. долл., поскольку компании опираются на существующие стратегии и инвестиции, превращаясь в масштабные цифровые экосистемы [4].

В условиях цифровой трансформации экспоненциально возрастает роль информационно-телекоммуникационных систем, которые строятся на основе космических спутников, используемых для связи, безопасности, разведки, коммерции и т.п. По данным [5] на начало 2022 г. на орбите находилось 4852 спутника, и их количество будет постоянно возрастать, поскольку стоимость запуска спутников на низких околоземных орбитах значительно снизилась. Это открывает границы космоса для крупных инициатив частного сектора по запуску с такими компаниями, как SpaceX, Blue Orgin и многими другими.

По мере превращения околоземной спутниковой системы в неотъемлемую часть глобальной информационной системы в части создания, обработки и хранения всё новых объёмов данных, а также по мере того, как отдельные подсистемы и элементы космической группировки становятся элементами архитектуры различных информационно-телекоммуникационных сетей (например, глобальная спутниковая система Starlink), возрастают и потенциальные информационные риски для государства, общества, личности. Всё это требует разработки новых подходов к обеспечению информационной безопасности космических систем и также требует особого внимания со стороны уполномоченных Федеральных Регуляторов, занимающихся вопросами безопасности объектов критической информационной инфраструктуры.

При этом собственниками информационных ресурсов могут быть различные субъекты (государство, общество, личность), в полном объеме реализующие полномочия владения, пользования, распоряжения этими ресурсами в пределах, установленных законом [6,7].

В этих условиях информационные ресурсы не упорядочены, далеко не все одинакового качества и разбросаны по бесчисленным базам данных по всему миру. На этом информационном пространстве и получили свое развитие технологии «больших данных».

Развитие технологий «больших данных» требует нового осмысления проблематики информационной безопасности в контексте этих технологий. Использование термина «информационная безопасность» носит весьма широкое распространение. При этом семантика этого понятия лежит в очень широких диапазонах в зависимости от того, где используется это понятие. В этой ситуации приходится задаваться вопросом: «Информационная безопасность чего или кого?».

В общем случае можно говорить, что цель получения и обработки некоторых сведений о предметной области состоит в их последующем использовании для формирования целенаправленных воздействий на объект управления. При этом до последнего времени наблюдение и осознание происходящего с объектом осуществлялось через некоторую *систему информатизации деятельности субъекта* (СИДС), т.е. опосредовано, а реализация принятого решения осуществлялась в рамках объекта управления, т.е. напрямую.

Таким образом, между субъектом, принимающим решение, и объектом, на который это решение направлено, стоит некоторая совокупность информационных моделей в составе СИДС. Это является потенциальной угрозой принятия ошибочного решения, так как эти информационные модели могут неправильно отображать объект управления.

В настоящее время осуществляется цифровизация всех областей жизнедеятельности современного государства, общества и личности, включая экономические, социальные, политические и прочие аспекты. Создаются цифровые двойники (англ. Digital Twins), представляющие собой цифровые модели физических объектов, физических процессов, процессов принятия решений, а также физических лиц - личностей этих объектов и процессов [8-10]. Теперь воздействие на объект управления осуществляется не напрямую, а на его цифрового двойника с последующим редуцированием, при необходимости, этого воздействия на физический объект управления.

Таким образом, появляется *система цифровизации деятельности субъекта* (СЦДС), которая включает в свой состав традиционную СИДС. Развитие технологий цифровизации и сфер их применения в различных областях жизнедеятельности современного государства, общества и личности порождают значительное разнообразие множества негативных воздействий на их цифровые модели [11,12]. Потенциальная возможность реализации этих негативных воздействий зависит от конкретных условий использования цифровых двойников [13,14].

Поэтому одним из основных принципов, реализуемым в процессе создания некоторой СЦДС, выступает принцип системной безопасности (обеспечение информационной безопасности). Этот принцип определяет, что информационно-коммуникационные услуги, реализованные в рамках СЦДС с определенным уровнем качества, являются безопасным продуктом. В свою очередь, это означает, что, во-первых, использование этих услуг не приводит к нежелательным для их потребителей последствиям в процессе выработки ими требуемого решения, и, во-вторых, эти услуги не могут быть использованы третьей стороной во вред как самой СЦДС, так и легитимным потребителям ее услуг.

Технологии «больших данных» связаны не столько с большими объемами этих данных, сколько с необходимостью обработки разнообразных консолидированных данных, доступных из огромного числа источников, имеющих различную структуру, схемы классификации и индексации.

Различные субъекты этого информационного пространства, автоматизируя свою деятельность и создавая соответст-

вующие СЦДС, хотят использовать в этих системах данные, информацию и знания, необходимые им для успешной реализации своей деятельности, при этом возможность использования технологий «больших данных» позволяет этим системам проанализировать огромное количество разнородных данных, стремясь к обработке практически всех данных, касающихся исследуемой предметной области, не ограничиваясь случайными выборками данных об этой предметной области.

Еще до момента появления технологий «больших данных» отмечалось, что «консолидированная информация является общественным знанием, подвергшимся специальному отбору, анализу, оценке, а также возможной реструктуризации и видоизменению («переупаковке») с целью быть пригодной для непосредственного решения проблем и удовлетворения других информационных потребностей определенных лиц или социальных групп, которые иначе не имели бы прямого доступа к этим знаниям и не могли бы ими эффективно воспользоваться, так как они рассеяны по многим документам и труднодоступны в своей оригинальной форме...» [15].

Указанные системы в общем случае позволяют агрегировать первичные «данные» и консолидировать «информацию», порождая их интерпретацию с последующей переработкой этой «информации» в «знания».

В связи со сказанным необходимо рассмотреть четыре понятия, которые широко используются в области информационно-коммуникационных технологий, а именно, понятия «данные», «информация», «знания» и «мудрость», а также их взаимосвязь. Обсуждение сущности этих понятий и их взаимосвязи имеет место в значительном количестве публикаций [16-27].

DIKW-модель и обработка «больших данных»

Для целей понимания перехода от «данных» к «информации» и далее к «знаниям» и «мудрости» в рамках технологий «больших данных» с точки зрения вопросов информационной безопасности представляется полезной DIKW-модель [28,29], в рамках которой иерархическая взаимосвязь указанных понятий представлена на рисунке 1.

Рис. 1. Уровни иерархии модели DIKW

DIKW (англ. Data, Information, Knowledge, Wisdom – данные, информация, знания, мудрость) – информационная иерархия, где каждый уровень добавляет определённые свойства к предыдущему уровню [30]. В основании иерархии находится уровень «данных». Уровень «информация» добавляет контекст. Уровень «знания» добавляет механизм использования. Уровень «мудрость» добавляет условия использования. Модель демонстрирует пути получения ценности в процессе обработки данных [28,30], при этом с каждым уровнем данные становятся более структурированными и пригодными для использования, вместе со степенью переработки данных, информации, знаний растет ценность полученного результата:

- ata («данные») – набор разрозненных фактов, символов (числа, слова, визуальные данные). Находится на дне иерархии и является материалом для обработки, из которого можно получить что-то ценное. Сами по себе данные не несут никакой пользы;
- Information («информация») – объединённые по смыслу данные. На этом уровне базовые кирпичики фактов образуют связи. В отличие от данных, информация несёт в себе пользу, т. к. описывает процессы и явления. Позволяет ответить на базовые вопросы: «кто?», «что?», «где?», «когда?». Информации недостаточно для решения каких-либо проблем;
- nowledge («знания») – результат фильтрации информации, которая переработана таким образом, что возникает возможность делать выводы. Связанные между собой факты образуют полную картину явления или процесса, с помощью которой можно делать выводы;
- isdom («мудрость» или иногда употребляют термин «общепринятое мнение») – верхушка пирамиды. На этом этапе обработки данных к знанию добавляется понимание. Если информация отвечает на вопрос «что?», знание – как?», то мудрость говорит нам «зачем?». Понимание позволяет выйти за границы интересующего нас явления или процесса, чтобы использовать его для более масштабных целей.

Похожая модель консолидации обработки от данных к знаниям была описана в [31] и представлена на рисунке 2.

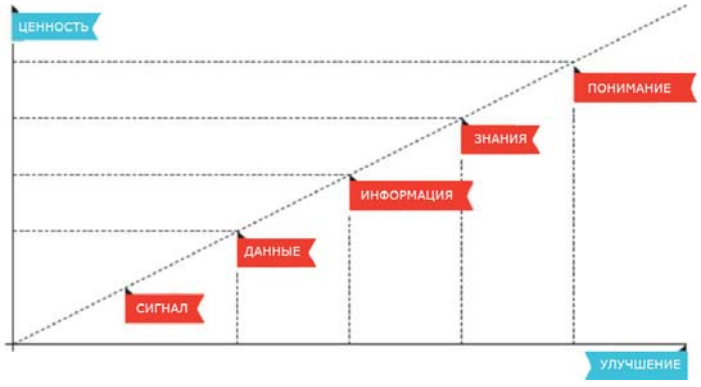

Рис. 2. Модель консолидации обработки

Обобщенное описание роста ценности результата обработки в рамках DIKW-модели представлено на рисунке 3.

Таким образом, учитывая вышеизложенное, можно сделать вывод, что воспринимаемые и фиксируемые факты окружающего мира представляют собой данные. Так в [32] данные рассматриваются как совокупность зафиксированных на физическом носителе сведений (фактов) об объекте, событии предметной области в форме, пригодной для их постоянного хранения, передачи и обработки.

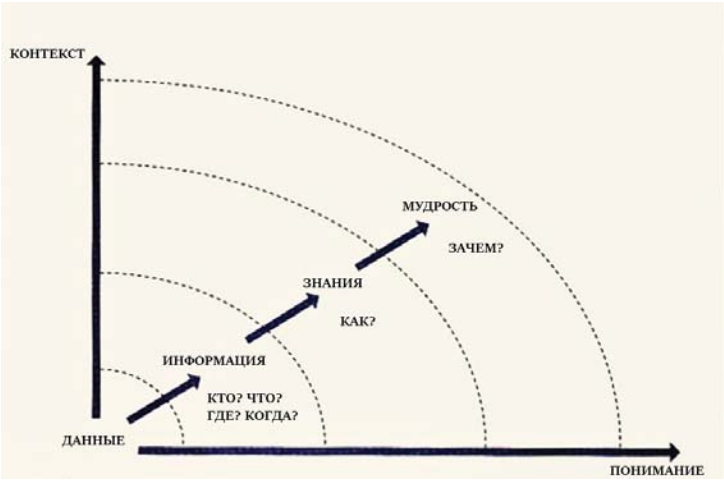

Рис. 3. Обобщенное описание роста ценности результата обработки

При использовании данных в процессе решения конкретных задач появляется информация. Результаты решения задач, информация (сведения), обобщенная в виде законов, моделей, теорий, совокупностей взглядов и представлений, представляет собой знания [16]. В этом смысле будем говорить о DIKW-обработке при использовании технологий «больших данных».

В современных условиях любая правомерная деятельность субъектов, принятие значимых и эффективных решений в конкретной предметной области требует использования всего множества знаний об этой области. На получение этих знаний и направлено применение технологий «больших данных». Эти технологии, по сути, предназначены для прогнозирования. Их рассматривают как часть компьютерной науки под названием «искусственный интеллект» (англ. Artificial Intelligent, AI), при этом происходит смещение акцента знаний от понимания причинности в пользу выявления корреляций [33], т.е. знание «почему» меняется на знание «что именно». Корреляции не могут сказать нам точно, почему происходит то или иное событие, зато предупреждают о том, какого оно рода и чего ждать от его наступления.

Однако существующие подходы к вопросам обеспечения информационной безопасности носят статический характер, связаны прежде всего с уровнем «данных» DIKW- модели и такими понятиями, как «разграничение доступа» и «матрица доступа». В этих подходах, как правило, определение уровня доступа (ограничение доступа сообразно грифу, роли, метке доступа) направлено на единичные данные, их агрегации (совокупности) и элементарные операции над ними (ввод, чтение, удаление, модификация).

Угрозы в области обеспечения информационной безопасности деятельности любого субъекта и процесса принятия им решений, используемых им информационных и телекоммуникационных технологий, сохранности и конфиденциальности информационных ресурсов являются следствием двух основных причин: специфики самого процесса информатизации и целенаправленных негативных действий со стороны иных субъектов, которые в принципе могут действовать в рамках предоставленного им уровня доступа и ничего при этом не нарушать.

Так как процесс принятия решений базируется на обработке «больших данных», то при получении доступа к ним возникают угрозы нарушения информационной безопасности субъекта, а также появляется возможность прогнозирования принимаемых им решений и, следовательно, влияния на эти решения. То есть возникает возможность для лиц, обладающих доступом к «большим данным» определенного уровня значимости (открытым и/или закрытым сведениям), получить доступ к информации и знаниям, имеющим чрезвычайно большую значимость, чем отдельные элементы данных. В [34] показано, что использование «больших данных» в области персональных данных, в принципе, в большинстве случаев может обеспечить требуемую идентификацию субъекта персональных данных практически с вероятностью 100% даже в случае применения технологий обезличивания к отдельным совокупностям этих данных.

Из всего сказанного выше при использовании технологий «больших данных» следует необходимость решения двух основных проблем:

1. Оценка влияния степени консолидации данных, информации и знаний на возрастание их важности в обеспечении информационной безопасности субъекта пространства «больших данных» (государства, общества, личности);

2. Оценка критической массы данных, информации и знаний, доступных в пространстве «больших данных», с точки зрения их комплексной совместной обработки и получения на их основе более важных сведений.

Перечисленные проблемы особенно актуальны, когда собственники информационных ресурсов (государство, общество, личность) могут непреднамеренно ошибочно определять степень важности этих ресурсов. Например, можно считать для себя, что отдельно взятые ресурсы являются открытыми и общедоступными. Однако при их накоплении и сопоставлении с другими такими же ресурсами, взятыми от других собственников и из смежных предметных областей, при их комплексной обработке и анализе может выясниться, что они несут семантически новые, более закрытые и ценные сведения, утечка или разрушение которых может нанести ущерб интересам субъекта пространства «больших данных».

Таким образом, возникают проблемы установления необходимого баланса между потребностью в информационном обмене и допустимыми ограничениями распространения данных, информации и знаний. Появляется новое виртуальное понятие – «информационная граница», которое нуждается в четком определении и способах обеспечения ее контроля [6]. Под «информационной границей» предлагается понимать зону ответственности (степень информированности) источника или потребителя результатов обработки «больших данных», нарушение которой может причинить моральный, психологический, материальный или физический ущерб государству, обществу или личности,

В данной постановке вопрос обеспечения информационной безопасности сводится к задаче прагматически обоснованного ограничения доступа для всех уровней DIKW-модели. В этом смысле будем говорить о «DIKW-безопасности».

Походы к контролю «информационных границ»

Можно предложить два подхода к контролю «информационных границ» – это экстенсиональный и интенсиональный подходы.

В рамках экстенсионального подхода [6,35] используются методы фильтрации для элементов DIKW-обработки, управления их потоками в рамках пространства «больших данных». Наиболее наглядно эти методы проявляются применительно к средствам массовой информации и социальным сетям, которые в значительном числе случаях не столько обеспечивают контроль «информационной границы», сколько пытаются ее задать для личности, общества или государства, исходя из собственных целей.

В рамках интенсионального подхода упор делается на контент и содержательную часть консолидированного результата DIKW-обработки. В этом случае задание информационных границ предлагается определять с помощью функционала информационной безопасности [6,34].

Характеристика «консолидация» определяет потенциальную возможность получения допустимой степени объединения элементов DIKW-обработки в процессе их агрегации/обобщения (верхний уровень консолидации) или декомпозиции/специализации (нижний уровень консолидации) и связана, прежде всего, с уровнем значимости (категорирования) доступной их совокупности.

Опишем соображения, лежащие в основе рассматриваемой модели.

Постановка задачи

Субъект пространства «больших данных» в процессе своей деятельности запрашивает те или иные элементы DIKW-обработки и методы их обработки (составляющие обработки). Каждая порция запрашиваемых составляющих обработки с точки зрения информационной безопасности характеризуются параметром, который будем называть исходным уровнем значимости G. Значения этого параметра естественно предполагать упорядоченными, но рассматриваемая модель пригодна и при отсутствии структуры на множестве возможных значений G. Существенно, что субъекту разрешен доступ лишь к составляющим обработки, исходный уровень значимости которых принадлежит некоторому подмножеству G_A. Данные с исходным уровнем значимости вне этого множества G_A не выдаются при условии прямого запроса.

Задача состоит в следующем. Применяя доступные ему методы, субъект может получать из запрошенных им элементов DIKW-обработки новые элементы. При этом он может, вообще говоря, получить и такие элементы, исходный уровень значимости которых выходит за пределы разрешенного множества G_A. Именно диагностика этой ситуации несанкционированного косвенного доступа и является основной целью данной модели. Поскольку ситуация в такой неформальной общности представляется необозримой, опишем основные типы угроз, отслеживание которых возможно в рамках предлагаемой модели. Это две основные угрозы: неправомочное агрегирование/обобщение и неправомочная детализация (верхний и нижний уровень консолидации).

Один сценарий реализации такого типа угроз состоит в том, что косвенно доступными становятся элементы DIKW-обработки, относящиеся либо к объектам управления более высокого уровня, чем положено, либо, наоборот, из агрегированных элементов более высокого уровня становятся косвенно доступными элементы о конкретных объектах управления более низкого уровня.

Второй сценарий заключается в том, что, имея сведения о части элементов DIKW-обработки, относящихся к некоторому объекту управления, субъект может получить сведения о других элементах, которые не входят в число разрешенных для него.

Функционал информационной безопасности C, определяющий связь уровня значимости с уровнями консолидации элементов DIKW-обработки, схематично приведен на рисунке 4. Данный функционал показывает, что для нахождения в рамках допустимых «информационных границ» необходимо, например, понижать уровень агрегации элементов DIKW-обработки при повышении уровня обработки информации. Это означает, что если субъект пытается в процессе несанкционированного косвенного анализа определенной задачи управления применить более интеллектуальные модели и методы обработки, то для него может потребоваться уменьшение уровня агрегации данных (например, глубины временной ретроспективы или состава доступных ему элементов DIKW-обработки).

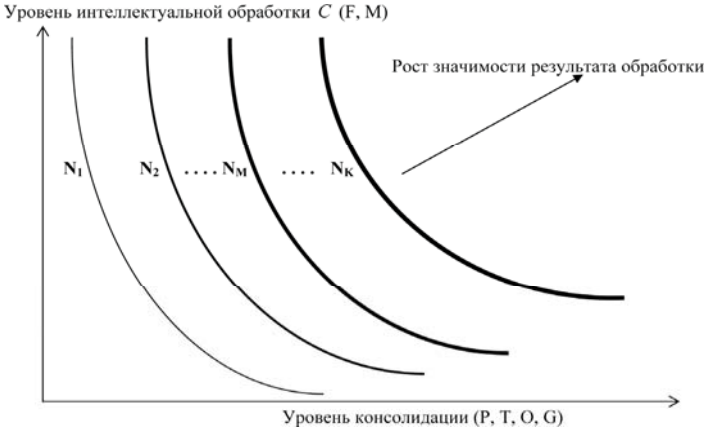

Рис. 4. Функционал информационной безопасности

При таком подходе постановка задачи обеспечения контроля информационных границ имеет следующий вид:

«Для i-го субъекта $(i=1,I)$ ***ЗАДАНЫ*** N_k *и* N_m *уровни значимости, определяющие верхний и нижний уровни консолидации элементов DIKW-обработки. Нижний уровень консолидации – максимальная степень детализации элементов. Верхний уровень консолидации – максимальная степень агрегации и обобщения элементов.*

Требуется ***ОБЕСПЕЧИТЬ*** $N_m < C_i < N_k$, *с учетом (1):*

$$C = FUN(P,T,O,G,F,M), \quad (1)$$

где:

P – элементы DIKW-обработки;

T – глубина временной ретроспективы;

O – объекты управления;

G – исходные уровни значимости;

F – функции управления;

M – используемые модели и методы обработки информации.

Построение функционала информационной безопасности является нетривиальной задачей, однако в [34] были рассмотрены и исследованы подходы, а также частные решения для его построения.

Заключение

Рассмотренный интенсиональный подход для контроля «информационных границ» в процессе использования технологий «больших данных» означает переход от статических решений вопросов обеспечения информационной безопасности субъекта к реализации динамических процедур такого контроля с использованием элементов искусственного интеллекта, мониторинга этого пространства и процедур прогнозирования своего «места» в пространстве «больших данных».

Предложенный подход приобретает ещё большую актуальность по мере возникновения новых потенциальных рисков, связанных с развитием квантовых вычислений и широким включением околоземного космического пространства в состав глобальной информационной системы. Это приведёт к ещё большему взрывному росту объёма обрабатываемых разнородных данных, содержащих конфиденциальную информацию о государстве, обществе и личности, и попыткам злоумышленников получить к ним несанкционированный доступ с целью нанесения ущерба как непосредственно субъекту, так и компрометации его цифрового двойника.

Литература

1. *Dokuchaev V.A.* Digital transformation: New drivers and new risks // 2020 International Conference on Engineering Management of Communication and Technology, EMCTECH 2020 - Proceedings, Vienna, 20-22 октября 2020 года. Vienna, 2020. P. 9261544. DOI 10.1109/EMCTECH49634.2020.9261544. EDN VWIIZW.
2. Что вы знаете о цифровой вселенной? [Электронный ресурс] URL: https://www.osp.ru/dobrodata/news/2020-08-24/13055554# (дата обращения: 07.05.2022).
3. Big Data факты 2022 г. [Электронный ресурс] URL: https://alakris.ru/blog/big-data-fakty-2015-g/ (дата обращения: 07.05.2022).
4. IDC FutureScape: Worldwide Digital Transformation 2021 Predictions By: Shawn Fitzgerald, Daniel-Zoe Jimenez, Serge Findling, Yukiharu Yorifuji, Megha Kumar, Lianfeng Wu, Giulia Carosella, Sandra Ng, Robert Parker, Philip Carter, Meredith Whalen. [Электронный ресурс] URL: https://www.idc.com/getdoc.jsp?containerId=US46880818 (дата обращения: 07.05.2022).
5. Chuck Brooks. The Urgency To Cyber-Secure Space Assets. [Электронный ресурс] URL: https://www.forbes.com/sites/chuckbrooks/2022/02/27/the-urgency-to-cyber-secure-space-assets/?sh=1b9dc71051b1 (дата обращения: 07.05.2022).
6. *Статьев В.Ю., Ефремов В.А., Солянкин Н.Н.* Информационная безопасность при подготовке и принятии управленческих решений // Экономика и производство, №10-12. 1999.
7. *Докучаев В.А., Кальфа А.А., Маклачкова В.В.* Архитектура центров обработки данных. М.: Горячая линия-Телеком, 2020. 240 с. ISBN 978-5-9912-0849-9. EDN BHARSE.
8. *Докучаев В.А., Маклачкова В.В., Статьев В.Ю.* Цифровизация субъекта персональных данных // T-Comm: Телекоммуникации и транспорт. 2020. Т. 14. № 6. С. 27-32. DOI 10.36724/2072-8735-2020-14-6-27-32. EDN XVWYJP.
9. *Докучаев В.А., Маклачкова В.В., Статьев В.Ю.* Идентификация субъекта – ключевой момент в процессе обработки персональных данных // Технологии информационного общества : Сборник трудов XIV Международной отраслевой научно-технической конференции, Москва, 18-19 марта 2020 года. Москва: Издательский дом Медиа Паблишер, 2020. С. 273-274. EDN ADBXFS.
10. *Dokuchaev V.A., Maklachkova V.V., Statyev V.Yu.* Data subject as augmented reality // Synchroinfo Journal. 2020. Vol. 6. No 1. P. 11-15. DOI 10.36724/2664-066X-2020-6-1-11-15. – EDN ULPVZC.
11. *Dokuchaev V.A., Maklachkova V.V., Statyev V.Yu.* Classification of personal data security threats in information systems // T-Comm. 2020. Vol. 14. No 1. P. 56-60. DOI 10.36724/2072-8735-2020-14-1-56-60. EDN QOGYHH.
12. *Докучаев В.А., Маклачкова В.В., Статьев В.Ю.* Требования к информационным системам при работе с «цифровым образом» субъекта // III Научный форум телекоммуникации: теория и технологии ТТТ-2019 : Материалы XXI Международной научно-технической конференции, Казань, 18-22 ноября 2019 года. Казань: Казанский государственный технический университет им. А.Н. Туполева, 2019. С. 296-297. EDN HRNDBS.
13. *Dokuchaev V.A., Maklachkova V.V., Makarova D.V., Volkova L.V.* Analysis of Data Risk Management Methods for Personal Data Information Systems // 2020 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, 19-20 марта 2020 года. Moscow: Institute of Electrical and Electronics Engineers Inc., 2020. P. 9078547. DOI 10.1109/IEEECONF48371.2020.9078547. EDN XXGAVW.
14. *Maklachkova V.V., Dokuchaev V.A., Statev V.Y.* Risks identification in the exploitation of a geographically distributed cloud infrastructure for storing personal data // 2020 International Conference on Engineering Management of Communication and Technology, EMCTECH 2020 – Proceedings, Vienna, 20-22 октября 2020 года. Vienna, 2020. P. 9261541. DOI 10.1109/EMCTECH49634.2020.9261541. EDN EQOLNB.
15. *Сарацевич Т., Вуд Ю.* Консолидация информации: Руководство по оценке, реструктуризации и видоизменению научной и технической информации, Париж, ЮНЕСКО, 1981. 327с.
16. Понятия информации, даваемые различными науками. Отличие от сведений данных и знаний. [Электронный ресурс] URL: https://vuzlit.ru/1039237/ponyatiya_informatsii_davaemye_razlichnymi_naukami_otlichie_svedeniy_dannyh_znaniy (дата обращения: 07.05.2022).
17. Информация. Электронный ресурс] U L: https://science.wikia.org/ru/wiki/Информация (дата обращения: 07.05.2022).
18. Информация. Электронный ресурс] U L: https://investments.academic.ru/1012/Информация (дата обращения: 07.05.2022).
19. Исследование соотношения понятий "знание" и "информация". [Электронный ресурс] URL: https://vuzlit.ru/1583288/issledovanie_sootnosheniya_ponyatiy_znanie_i_informatsiya_ (дата обращения: 07.05.2022).
20. *Орехов В.Д.* Определение понятия знание. [Электронный ресурс] URL: https://world-evolution.ru/opredelenie-ponyatiya-znanie/ (дата обращения: 07.05.2022).
21. *Шемакин Ю.И.* Основы систематики // Открытое образование, №6, 2009.
22. *Курилкина В.Н.* Философский и общенаучный анализ понятия информации // Вестник СВФУ, 2014. Т. 11, № 1.
23. *Урсул А.Д.* Информация, информатика, глобалистика // Открытое образование. №6, 2011.
24. *Воскресенский А.К.* Понятие «информация»: Философские аспекты (Аналитический обзор) // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Сер. 3, Философия: РЖ / РАН. ИНИОН. Центр гуманит. науч.-информ. исслед. Отд. философии. М., 2012. № 1. С. 5-28.
25. *Колин К.К.* Природа информации и философские основы информатики // Открытое образование», №2, 2005.
26. *Петров М.А.* О соотношении понятий "знание" и "информация", диссертация по ВАК РФ 09.00.01., 2005, Электронная библиотека диссертаций (www.dissercat.com).
27. *Турчевская Б.К.* Трансформация знания и информации: Информационные процессы и барьеры, диссертация по ВАК РФ 09.00.01.,2002, Электронная библиотека диссертаций. [Электронный ресурс] URL: https://www.dissercat.com/content/transformatsiya-znaniya-i-informatsii-informatsionnye-protsessy-i-barery
28. DIKW. Электронный ресурс] URL https://dic.academic.ru/dic.nsf/ruwiki/119809 (дата обращения: 07.05.2022).
29. *Ackoff R.L.* From Data to Wisdom // Journal of Applies Systems Analysis. Vol. 16, 1989, pp. 3-9.
30. DIKW модель. [Электронный ресурс] URL: https://simpleone.ru/glossary/dikw-model/ (дата обращения: 07.05.2022).
31. Lankow, J. and Ritchie, J. and Crooks, R. Chapter 7. Data Visualization Interfaces Infographics: The Power of Visual Storytelling. Wiley, 2012.
32. *Горбань Е.В., Докучаев В.А., Маклачкова В.В., Статьев В.Ю.* Оценка качества обработки больших объёмов данных в высоконагруженных инфокоммуникационных системах // Телекоммуникационные и вычислительные системы - 2018 : Международный форум информатизации (МФМ-2018); Международный конгресс (CTN-2018) "Коммуникационные технологии сети", Москва, 21 ноября 2018 года. М.: Горячая линия – Телеком, 2018. С. 25-28. EDN VWQCCQ.
33. *Виктор Майер-Шенбергер, Кеннет Кукьер.* Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. М.: Издательство «Манн, Иванов и Фербер», 2014.
34. *Левин В.К., Статьев В.Ю.*, и др. Отчет по НИР «Формула», Академия криптографии Российской Федерации, 2003.
35. *Феоктистов Г.Г.* «Информационная безопасность общества», социально-политический журнал. №5, 1996.

INFORMATION SECURITY IN THE BIG DATA SPACE

Vyacheslav Yu. Statev, *Transport Security Fund, Moscow, Russia,* ***articleglory@yandex.ru***
Vladimir A. Dokuchaev, *MTUCI, Moscow, Russia,* ***v.a.dokuchaev@mtuci.ru***
Victoria V. Maklachkova, *MTUCI, Moscow, Russia,* ***v.v.maklachkova@mtuci.ru***

Abstract
The purpose of this work is to develop the foundations of a new approach to the problem of information security in the context of "Big Data" technologies. The purpose of obtaining and processing information about the subject area is their subsequent use for the formation of targeted impacts on the control object. At the same time, observation and awareness of what is happening with the object was previously carried out through a certain system of informatization of the subject's activity, i.e. indirectly, and the implementation of the decision was carried out within the framework of the control object, i.e. directly. Under the conditions of Digital Transformation, the impact on the control object is carried out not directly, but on its "Digital Twin" with subsequent reduction, if necessary, of this impact on the physical control object. Thus, a system of digitalization of the subject's activity appears. When implementing such systems, one of the main principles is the principle of system security (ensuring information security). However, its implementation is hampered by the fact that "Big Data" technologies are associated not so much with their large volumes, but also with the need to process a variety of consolidated data that have a different structure, classification and indexing schemes. In this regard, it is proposed to analyze the DIKW-model and analyze the essence of the concepts included in it, which will make it possible to proceed to DIKW processing using "Big Data" technologies. At the same time, it becomes necessary to solve two problems: assessing the impact of the degree of consolidation of data, information and knowledge on the increase in their importance in ensuring the information security of the subject of the "Big Data" space and assessing the critical mass of data, information and knowledge available in the "Big Data" space, with from the point of view of their complex joint processing and obtaining more important information on their basis. To establish the necessary balance between the need for information exchange and the permissible restrictions on the dissemination of data, information and knowledge, a new virtual concept is introduced - "Information Boundary" and the issue of ensuring information security is reduced to the task of pragmatically justified access restriction for all levels of the DIKW-model and the concept of "DIKW - security. An intensional approach is proposed for controlling "Information Boundaries" in the process of using "Big Data" technologies, which ensures the transition from static solutions to the issues of ensuring the subject's information security to the implementation of dynamic procedures for such control using elements of Artificial Intelligence, monitoring this space and predicting one's "place" in the "Big Data" space.

Keywords: *data, information, knowledge, DIKW-model, Big Data, information security, information system, digital transformation, digital twin, data processing, information boundary.*

References

1. V.A. Dokuchaev (2020). Digital transformation: New drivers and new risks. *2020 International Conference on Engineering Management of Communication and Technology, EMCTECH 2020*. Proceedings, Vienna, 20-22 october 2020. Vienna, 2020. P. 9261544. DOI 10.1109/EMCTECH49634.2020.9261544. EDN VWIIZW.
2. What do you know about the digital universe? URL: https://www.osp.ru/dobrodata/news/2020-08-24/13055554# (date of access: 07.05.2022).
3. Big data facts 2022 URL: https://alakris.ru/blog/big-data-fakty-2015-g/ (date of access: 07.05.2022).
4. Shawn Fitzgerald, Daniel-Zoe Jimenez, Serge Findling, Yukiharu Yorifuji, Megha Kumar, Lianfeng Wu, Giulia Carosella, Sandra Ng, Robert Parker, Philip Carter, Meredith Whalen. IDC FutureScape: Worldwide Digital Transformation 2021 Predictions By: URL: https://www.idc.com/getdoc.jsp?containerId=US46880818 (date of access: 07.05.2022).
5. Chuck Brooks. The Urgency To Cyber-Secure Space Assets. URL: https://www.forbes.com/sites/chuckbrooks/2022/02/27/the-urgency-to-cyber-secure-space-assets/?sh=1b9dc71051b1 (date of access: 07.05.2022).
6. V.Yu. Statev, V.A. Efremov, N.N. Solyankin (1999), "Information security in the preparation and adoption of managerial decisions", *Economics and production*, no. 10-12.
7. V.A. Dokuchaev, A.A. Kalfa, V.V. Maklachkova (2020). Architecture of data processing centers. Moscow: Hot Line-Telecom. 240 p. ISBN 978-5-9912-0849-9. EDN BHARSE.
8. V.A. Dokuchaev, V.V. Maklachkova, V.Yu. Statiev (2020). Digitalization of the subject of personal data. *T-Comm*. Vol. 14. No. 6, pp. 27-32. DOI 10.36724/2072-8735-2020-14-6-27-32. EDN XVWYJP.
9. V.A. Dokuchaev, V.V. Maklachkova, V.Yu. Statev (2020). Identification of the subject - a key moment in the process of processing personal data Technologies of the Information Society: *Proceedings of the XIV International Industry Scientific and Technical Conference,* Moscow, March 18-19, 2020. Moscow: Media Publisher, pp. 273-274. EDN ADBXFS.
10. V. A. Dokuchaev, V. V. Maklachkova, V. Yu. Statev (2020). Data subject as augmented reality. *Synchroinfo Journal*. Vol. 6. No 1, pp. 11-15. DOI 10.36724/2664-066X-2020-6-1-11-15. EDN ULPVZC.
11. V.A. Dokuchaev, V.V. Maklachkova, V.Yu. Statev (2020). Classification of personal data security threats in information systems. T-Comm. Vol. 14. No 1, pp. 56-60. DOI 10.36724/2072-8735-2020-14-1-56-60. EDN QOGYHH.

12. V.A. Dokuchaev, V.V. Maklachkova, V.Yu. Statev (2019). Requirements for information systems when working with the "digital image" of the subject. *III Scientific Forum of Telecommunications: Theory and Technologies TTT-2019 : Proceedings of the XXI International Scientific and Technical conference,* Kazan, November 18-22, 2019. Kazan: Kazan State Technical University. A.N. Tupolev, pp. 296-297. EDN HRNDBS.
13. V.A. Dokuchaev, V.V. Maklachkova, D.V. Makarova, L.V. Volkova (2020). Analysis of Data Risk Management Methods for Personal Data Information Systems. *2020 Systems of Signals Generating and Processing in the Field of on Board Communications*, Moscow, 19-20 March 2020. Moscow: Institute of Electrical and Electronics Engineers Inc., 2020. P. 9078547. DOI 10.1109/IEEECONF48371.2020.9078547. EDN XXGAVW.
14. V.V. Maklachkova, V.A. Dokuchaev, V.Y. Statev (2020). Risks identification in the exploitation of a geographically distributed cloud infrastructure for storing personal data. *2020 International Conference on Engineering Management of Communication and Technology, EMCTECH 2020*. Proceedings, Vienna, 20-22 oktober 2020. Vienna, 2020. P. 9261541. DOI 10.1109/EMCTECH49634.2020.9261541. EDN EQOLNB.
15. T. Saracevic, Y. Wood (1981), "Information Consolidation: A Guide to Assessing, Restructuring, and Reshaping Scientific and Technical Information", Paris, UNESCO. 327p.
16. The concepts of information given by various sciences. Difference from information data and knowledge. [Electronic resource] URL:https://vuzlit.ru/1039237/ponyatiya_informatsii_davaemye_razlichnymi_naukami_otlichie_svedeniy_dannyh_znaniy (date of access:05/07/2022).
17. URL: https://science.wikia.org/ru/wiki/Информация (date of access: 07.05.2022).
18. URL: https://investments.academic.ru/1012/Информация (date of access: 07.05.2022).
19. Study of the relationship between the concepts of "knowledge" and "information". URL: https://vuzlit.ru/1583288/issledovanie_soot-nosheniya_ponyatiy_znanie_i_informatsiya_ (date of access: 07.05.2022).
20. V.D. Orekhov. Definition of the concept of knowledge. URL: https://world-evolution.ru/opredelenie-ponyatiya-znanie/ (date of access: 07.05.2022).
21. Yu.I. Shemakin (2009), "Fundamentals of Systematics", *Open Education*, No. 6.
22. V.N. Kurilkina (2014), "Philosophical and general scientific analysis of the concept of information", NEFU BULLETIN, vol. 11, no. 1.
23. A.D. Ursul (2011), "Information, Informatics, Globalistics", *Open Education*, No. 6.
24. A.K. Voskresensky (2012). The concept of "information": Philosophical aspects (Analytical review). *Social and humanitarian sciences. Domestic and foreign literature.* Ser. 3, Philosophy: RJ / RAS. INION. Humanitarian Center. scientific-inform. research department philosophy. No. 1, pp. 5-28.
25. K.K. Colin (2005), Nature of Information and Philosophical Foundations of Informatics, *Open Education*, No. 2.
26. M.A. Petrov (2005), On the relationship between the concepts of "knowledge" and "information", dissertation according to the Higher Attestation Commission of the Russian Federation 09.00.01., Electronic Library of Dissertations (www.dissercat.com).
27. B.K. Turchevskaya (2002), "Transformation of knowledge and information: Information processes and barriers", dissertation on the Higher Attestation Commission of the Russian Federation 09.00.01.,2002, Electronic Library of Dissertations. [Electronic resource] URL: https://www.dissercat.com/content/transformatsiya-znaniya-i-informatsii-informatsionnye-protsessy-i-barery
28. DIKW. URL: https://dic.academic.ru/dic.nsf/ruwiki/119809 (дата обращения: 07.05.2022).
29. R.L. Ackoff (1989), "From Data to Wisdom", Journal of Applies Systems Analysis. Vol. 16, pp 3-9.
30. DIKW модель. URL: https://simpleone.ru/glossary/dikw-model/ (date of access: 07.05.2022).
32. E.V. Gorban, V.A. Dokuchaev, V.V. Maklachkova, V.Yu. Statev (2018). Evaluation of the quality of processing large amounts of data in highly loaded infocommunication systems. *Telecommunication and Computing Systems - 2018*: International Informatization Forum (IFM -2018); International Congress (CTN-2018) "Communication Network Technologies", Moscow, November 21, 2018. Moscow: Hot Line-Telecom. P. 25-28. EDN VWQCCQ.
33. Viktor Mayer-Schenberger, Kenneth Kukier (2014), "Big data. A revolution that will change the way we live, work and think", Mann, Ivanov and Ferber Publishing House, Moscow.
34. V.K. Levin, V.Yu. Statev, etc. (2003). Report on research work "Formula", Academy of Cryptography of the Russian Federation.
35. G.G. Feoktistov (1996). "Information security of society", *Socio-political magazine.* No. 5.

Information about authors:

Vyacheslav Yu. Statev, *PhD, expert "Transport Security Fund", Moscow, Russia*

Vladimir A. Dokuchaev, *DSc (Tech), Professor, Head of the Department "Networks Information Technologies and Services" MTUCI, Moscow, Russia*

Victoria V. Maklachkova, *Senior Lecturer of the Department "Networks Information Technologies and Services" MTUCI, Moscow, Russia*