

УДК 338.24:338.32

**ПРОБЛЕМА ОБЕСПЕЧЕНИЯ КАЧЕСТВА ДАННЫХ  
В ИНФОРМАЦИОННЫХ СИСТЕМАХ ГОСУДАРСТВЕННОГО УПРАВЛЕНИЯ****Старовойтов В.Г., Кузнецов Н.В., Котова Н.Е., Лапенкова Н.В.***Финансовый университет при Правительстве Российской Федерации, Москва,  
e-mail: nkuznetsov@fa.ru*

Статья посвящена вопросу обеспечения качества данных, содержащихся в информационных системах используемых при принятии решений в сфере государственного управления. Рассмотрены различные подходы к трактованию понятия «качество данных в информационных системах». Показано, что с точки зрения задач государственного управления в качестве ключевых критериев качества должны выступать полнота, достоверность, интерпретируемость, согласованность, затраты ресурсов на сбор и обработку данных, а также доступность и своевременность их предоставления пользователям. Обозначена ключевая проблема качества данных, используемых для решения задач государственного управления федерального уровня – информация, как правило, существенно зашумлена, а иногда и целенаправленно искажена. Рассмотрены основные подходы к повышению качества данных, при этом рассматриваемые методы структурированы на прямые (непосредственное сравнение данных из различных источников) и косвенные (агрегация различной по содержанию информации об одном и том же и последующая оценка ее согласованности). Рассмотрена зависимость сложности модели оценки качества данных от объема анализируемого информационного поля. Предложено создание экспертной системы нового поколения, которая сможет эффективно агрегировать в том числе противоречивые оценки, обеспечивая множественные прямые и косвенные оценки согласованности данных. Данная система должна являться первым уровнем всех прочих систем поддержки принятия решений, обеспечивая для них подготовку исходных данных.

**Ключевые слова:** государственное управление, информационные системы, качество данных, критерии качества, методы повышения качества, поддержка принятия решений, экспертная система

**THE PROBLEM OF ENSURING DATA QUALITY IN PUBLIC  
ADMINISTRATION INFORMATION SYSTEMS****Starovoytov V.G., Kuznetsov N.V., Kotova N.E., Lapenkova N.V.***Financial University under the Government of the Russian Federation, Moscow,  
e-mail: nkuznetsov@fa.ru*

The article is devoted to the issue of ensuring the quality of data contained in information systems used in decision-making in public administration. Various approaches to the interpretation of the concept of «data quality in information systems» are considered. It is shown that, from the point of view of public administration tasks, the key quality criteria should be completeness, reliability, interpretability, consistency, the cost of resources for collecting and processing data, as well as the availability and timeliness of their provision to users. The key problem of the quality of the data used to solve the tasks of public administration at the federal level is identified – the information is usually significantly noisy, and sometimes purposefully distorted. The main approaches to improving the quality of data are considered, while the methods under consideration are structured into direct (direct comparison of data from various sources) and indirect (aggregation of information on the same information that is different in content and a subsequent assessment of its consistency). The dependence of the complexity of the model for assessing data quality on the volume of the analyzed information field is considered. It is proposed to create a new generation expert system that can effectively aggregate, including conflicting estimates, providing multiple direct and indirect estimates of data consistency. This system should be the first level of all other decision support systems, providing them with the preparation of initial data.

**Keywords:** public administration, information systems, data quality, quality criteria, quality improvement methods, decision support, expert system

В последнее время кратно умножилось количество всевозможных информационных систем, собирающих, хранящих и обрабатывающих самую разнообразную информацию, касающуюся социально-экономической жизни общества. Одновременно с этим значительно возросло требование к качеству данных, их надежности, достаточности, полноте и согласованности. На наш взгляд, это обуславливается двумя причинами. Во-первых, постепенное усложнение и глобализация социально-экономических явлений и процессов приводит

к тому, что «цена» ошибочных управленческих решений становится все выше и выше. Использование некачественных данных, как основы для принятия решений, способно свести к нулю возможный позитивный эффект. А во-вторых, широкое использование в управлении современных информационных технологий, искусственного интеллекта и экономико-математических моделей требуют их наполнения исходными данными. Однако ручная проверка достоверности аналитических расчетов с ростом объемов информации и усложнением используе-

мых методов вычисления часто оказывается невозможной.

Проблемой качества данных довольно давно занимаются не только ученые, но также и государственные учреждения, и представители бизнес-сообщества. Следует отметить большую роль Научно-методологического совета Росстата (на сегодняшний день 83% его состава являются представителями науки и бизнеса), который выступил инициатором проведения большого объема научных исследований, направленных на формирование и совершенствование учетно-аналитической методологии на основе отечественного и международного опыта [1]. Однако эти исследования еще далеки от завершения. Сегодня можно констатировать, что координация методологии сбора, обработки и распространения информации вне федерального плана статистических работ осуществляется слабо [2], многие данные собираются вручную, отсутствуют четкие алгоритмы их обработки. Кроме того, присутствует сильный информационный шум, а иногда и намеренное искажение информации. Все это значительно осложняет аналитическую деятельность, приводит к ошибкам в оценке и прогнозы и, как следствие, приводит к значительному снижению качества принимаемых управленческих решений.

В условиях цифровизации экономики происходят кардинальные изменения и в системе взаимоотношений в сфере государственного управления. Национальная программа «Цифровая экономика Российской Федерации» в качестве одного из проектов предусматривает «Формирование цифровой платформы для взаимодействия в сфере стратегического управления в целях согласованности действий участников стратегического планирования на всех уровнях государственного управления в достижении стратегических приоритетов» [3]. Эта платформа предусматривает не только создание единого информационного поля для всех участников процесса управления, но также и применение технологий больших данных и искусственного интеллекта. Очевидно, что внедрение такой системы требует предварительной детальной проработки ее информационного обеспечения.

Цель исследования: определение подходов и ключевых методических принципов повышения качества данных при формировании информационной базы для решения прикладных задач в сфере государственного управления на федеральном уровне.

#### **Материалы и методы исследования**

Теоретическую основу исследования составили теоретические труды в сфере анализа и оценки ка-

чества информации, построения информационных систем, а также их применения для решения прикладных задач государственного управления.

В исследовании анализируются источники данных, используемые в системах государственного управления источники данных (такие как, например, ГАС «Управление», информационные системы Банка России, Федерального казначейства, Росстата и др.). Для этого использовался системный подход, методы научной абстракции, группировки, сравнения, синтеза, факторного анализа, систематизации и формализации.

#### **Результаты исследования и их обсуждение**

Прежде всего следует отметить, что в мировой практике нет общепринятого определения качества данных. Понятие «качество» применительно к любой информации субъективно и ситуационно, а следовательно, и классификация данных на «качественные» и «некачественные» является исключительно результатом относительно рационального консенсуса. Тем не менее в настоящее время де-факто установилось и превалирует понятие качества, основанное на оценке корректности методологии, применяемой в процессе сбора, обработки, анализа и использования данных (в том числе соответствие методологии принятым международным стандартам). При этом в качестве ключевых критериев качества данных в информационных системах выступают полнота, достоверность, интерпретируемость и согласованность [4].

В понятие полноты входит охват всех возможных источников информации в контексте решения поставленной задачи. Достоверность – это степень адекватности данных реальной социально-экономической ситуации, которая может быть подтверждена экспертными методами и верификацией. Интерпретируемость отражает возможность понимания данных пользователем и их соотнесения с объективной реальностью и во многом определяется строгостью используемых в методологии понятий и определений. Согласованность означает степень логической взаимосвязи данных, содержащихся в различных информационных системах, в том числе и полученных различными методами с использованием различных целевых совокупностей (согласованные данные непротиворечивы, сопоставимы и дополняют друг друга в сопряженных наблюдениях).

Другим компонентом, часто используемым при трактовании понятия «качество данных», является степень их востребованности пользователями. С точки зрения авторов настоящей статьи, этот критерий характеризует не столько сами данные, сколько

их полезность для решения конкретной задачи в конкретный момент времени (то, что сегодня кажется невостребованным, вполне может завтра оказаться абсолютно необходимой информацией) и, следовательно, он не может быть отнесен к критериям качества данных. Единственным исключением здесь, пожалуй, являются выраженные в стоимостных и временных показателях затраты ресурсов на сбор и обработку данных [1].

Важными характеристиками качества данных являются, на наш взгляд, их доступность и своевременность, определяющие возможность получения пользователями данных в требуемые сроки. При этом доступность характеризует не только готовность данных к распространению, но также и приемлемость формы их представления, а также наличие необходимой дополнительной информации (в том числе и информации о возможности и способах получения этих данных). Своевременность отражает приемлемость временного периода от момента возникновения у пользователя потребности в каких-либо данных до их фактического получения. Она характеризуется предельно допустимым интервалом времени, на протяжении которого информация продолжает оставаться актуальной и востребованной.

Все эти элементы качества должны быть сбалансированы между собой для достижения максимального положительного эффекта.

Данные, используемые для решения задач государственного управления (особенно федерального уровня), имеют определенную специфику, которая является следствием специфичности самого объекта исследования, доступный информационный след которого, как правило, существенно зашумлен, а иногда и целенаправленно искажен. В связи с этим возникает необходимость в применении специальных методов повышения качества данных. При этом рассмотренные критерии качеств применительно к конкретной прикладной задаче могут иметь различную значимость. Так в предметных областях, связанных с анализом, прогнозом и выработкой решений в условиях высокой неопределенности, наиболее важными характеристиками данных являются их полнота и достоверность. В то же время критерий своевременности не является критичным, а критерий согласованности является техническим и служит основой для оценки достоверности данных в условиях невозможности получения абсолютно надежной информации.

Можно выделить два основных типа методов повышения качества данных: прямые и косвенные [5].

**Прямые методы.** Согласно основной идее таких методов необходимо обеспечить получение одних и тех же по содержанию данных из различных источников и, используя достаточно тривиальные процедуры их сравнения, формировать заключение об их качестве и осуществлять выбор конкретных значений для применения в моделях (характерным примером является использование правила «два из трех»). При этом прямой подход используется в случаях, когда имеется возможность получения информации из различных независимых друг от друга источников и при этом отсутствует достаточное время на интеллектуальную обработку входного потока или имеющиеся у экспертов знания предметной области не позволяют сформировать достаточно качественные суждения. В качестве примеров такого рода отраслей можно привести военное управление, управление технологическими процессами, управление социально-экономическими системами и т.п.

**Косвенные методы.** В основе этих методов лежит агрегация различной по содержанию информации об одном и том же явлении (различные ракурсы наблюдаемого явления) и последующая оценка ее согласованности на основе знания предметной области. Косвенный подход используется в случаях, когда получение информации из множества источников связано с неприемлемыми затратами (финансовыми или временными) или отсутствует возможность эффективной обработки требуемого массива информации. Примером использования такого подхода является формирование материального баланса, когда в условиях невозможности получения необходимого количества независимых измерений по всем необходимым точкам мониторинга, решается задача нелинейной оптимизации для восстановления значений характеристик материальных потоков в местах, где прямые измерения не проводятся.

На практике для повышения качества данных могут применяться одновременно оба эти подхода (в том числе и в рамках одной системы) [5].

Ключевым принципом, который является основой методики обеспечения качественными исходными данными задач государственного управления является принцип, который можно условно назвать «принцип компенсации отсутствия информации знаниями» и который можно описать следующим образом: если информации много, то модель оценки ее качества может быть относительно простой, но если информации мало, то для обеспечения того же качества оценки данных модель должна быть

обширной и глубокой [6]. В конечном итоге ключевой технологической основой определения качества данных является выявление степени и характера рассогласованности содержательно идентичных синтетических параметров, получаемых, как результат применения различных моделей над анализируемыми данными.

Эффективность такого подхода и специфика его реализации достаточно очевидны для предметных областей, в которых анализ данных осуществляется на основе детерминированных моделей, а применяемые в них взаимосвязи между параметрами являются строго доказанными (геометрия, кинематика, многие производственные процессы и т.п.). Однако в области социально-экономических наук отсутствуют строгие модели, поэтому прямой перенос в эту область технологий оценки качества данных из «технических» областей зачастую не является рациональным. Здесь необходимо учитывать слабую изученность предметных областей, превалирование мнений над знаниями при оценке возможных последствий и/или возможных причин. Основной проблемой при этом является принципиальная невозможность дать строгую локальную оценку причины выявленной рассогласованности данных: является ли она следствием их некачественности или причиной этой рассогласованности является несовершенство моделей, применяемых для оценки качества данных? Для ответа на этот вопрос в каждом конкретном случае требуется проведение отдельного детального исследования.

Отдельную проблему представляет собой информатизация процесса оценки качества данных. Можно констатировать, что в настоящий момент у экспертов практически отсутствует технология, позволяющая действительно эффективно объединять разнородные экспертные знания для эффективного решения задач комплексной оценки обстановки, прогнозирования ее развития и выработки решений по эффективному воздействию на нее. Формально в настоящее время существует около 10 базовых технологий экспертного анализа и около 200 их вариантов, однако все они слабо ориентированы на практическое решение сложных задач. Отсутствие эффективной технологии интеграции экспертных мнений создало в экспертной среде отношение к этим технологиям как к технологиям, главным эффектом применения которых является частичный перенос ответственности за итоговое решение с эксперта на технологию [7].

В этих условиях, лица, принимающие решения, не могут полностью довериться результатам работы и данным экспертных

систем, поскольку лежащая в основе любой системы модель отражает субъективное понимание действительности ее разработчиком. На сегодняшний день сформировался достаточно отрицательный исторический опыт применения таких систем (например, связанный с созданием и применением моделей системной динамики Форрестера). Также необходимо отметить, что сложные и закрытые системы, из-за отсутствия возможностей понимания их содержания для тех, кто не является их разработчиком, лишают руководителя управленческой воли (руководитель из субъекта воли частично превращается в «придаток» экспертной системы). Системный ущерб этого явления очевиден.

Для эффективного повышения качества данных с целью проведения модельных исследований в сфере государственного управления должна быть создана экспертная система нового поколения, которая сможет эффективно агрегировать в том числе противоречивые оценки, обеспечивая множественные прямые и косвенные оценки согласованности данных. Такой подход к модельному обеспечению является, несомненно, более сложным для освоения конечным пользователем и менее контролируемым потенциальным бенефициаром результатов модельных исследований, однако обеспечивает существование более высокой автоматизации (соотношение прикладных решений к объему вводимых пользователем данных и правил). Одним из ключевых преимуществ подхода, базирующегося на экспертных системах, является его потенциальная возможность выявлять «слабые звенья» в предлагаемых интерпретациях обстановки не только на этапе анализа данных, но и на этапе решения прикладных задач анализа и прогноза.

Ключевой технологической основой определения качества данных в такой системе является выявление степени и характера рассогласованности содержательно идентичных синтетических параметров, получаемых, как результат применения различных моделей над анализируемыми данными. При этом эти модели условно разделить на несколько категорий [5]: элементарные (направленные на локальное исследование отдельных элементов информации на предмет соответствия содержания некоторым ограничениям, например, базовым ограничениям предметной области); системные (направленные на выявление несогласованности характеристик элементов в некоторой системе); мульти-системные (направленные на выявление рассогласования состояний нескольких

взаимосвязанных систем); временные (направленные на выявление рассогласованности данных во времени); эволюционные (направленные на выявление рассогласований в состояниях, связанных с определенными фазами эволюции связанных элементов/систем). Спецификой этих моделей является то, что конечным результатом моделирования является оценка расхождения значений определенных параметров. При этом модели, применяемые для повышения качества данных, также отражают понимание объекта исследования исследователем, однако, в отличие от целевых моделей (моделей анализа, прогноза), модели для анализа исходных данных, как правило, более локальны.

Такая экспертная система будет являться первым уровнем всех прочих систем поддержки принятия решений, обеспечивая для них подготовку исходных данных [8].

### Заключение

Рассмотренные выше подходы направлены на систематизацию и повышение качества информационного обеспечения государственного управления.

Очевидно, что деятельность по сбору, обработке и распространению информации должна основываться на научно обоснованной методологии. При этом методологическая база должна быть гармонизирована с принятыми международными стандартами и принципами использования информации. Этому требованию сегодня в наибольшей мере отвечают данные информационных систем, которые создаются под руководством федеральных органов исполнительной власти (Федеральная служба государственной статистики, Банк России, Федеральное казначейство России, Минэкономразвития России, Минфин России и др.). Их работа регулируется нормативными правовыми актами, а сами данные соответствуют утвержденным стандартам, техническим условиям и пр. При этом анализ показывает, что наибольшее качество имеют данные ГАС «Управление», Центрально-го Банка и Росстата.

Для обеспечения эффективности функционирования системы государственного управления в нее должны быть включены механизмы анализа качества источников информации, а также опыта эксплуатации используемых информационных технологий, программно-инструментальных средств, конкретных методов и методик обработки информации, вынесения оценок и суждений. Кроме того, как показывает практика, качество предоставляемой пользователям информации возрастает при регламентации основных требований, предъявляемых к данным. При этом должен соблюдаться принцип прозрачности используемой регламентной и методологической базы, призванный гарантировать объективность и соблюдение этических норм при осуществлении наблюдений и анализа.

*Статья подготовлена по результатам исследований, выполненных за счет бюджетных средств по государственному заданию Финуниверситету.*

### Список литературы

1. Кузнецова Е.И., Гамоненко С.С. Финансовый мониторинг как метод контроля в управлении экономической безопасностью // Вестник Московского университета МВД России. 2015. № 1. С. 74–79.
2. Борзых Л.А. Мониторинг показателей экономической безопасности России в социальной сфере // Социально-экономические явления и процессы. 2015. Т. 10. № 7. С. 34–41.
3. Об утверждении программы «Цифровая экономика Российской Федерации»: [расп. Правительства РФ от 28.07.2017 № 1632-р] // Собрание законодательства РФ. 2017. 07 авг. № 32. ст. 5138.
4. Виссарионов А.Б., Гумеров Р.Р. Об использовании предельных (пороговых) значений индикаторов экономической безопасности Российской Федерации // Управленческие науки. 2017. № 3. С. 38–45.
5. Селиванов А.И. Мониторинг экономической безопасности в условиях развития: постановка проблемы // Микроэкономика. 2017. № 4. С. 84–89.
6. Якимов И.М. Оценка достоверности результатов имитационного моделирования по результатам аналитического моделирования // Вестник Казанского технологического университета. 2015. № 17. С. 19–22.
7. Логвиненко Е.В., Полякова М.Е. Макроэкономическое моделирование инфляции // Национальная Ассоциация Ученых. 2015. № 2. С. 143–146.
8. Чувилов Д.А. Методика объединения экспертной системы и системы имитационного моделирования // Промышленные АСУ и контроллеры. 2017. № 3. С. 11–18.