

## ПОДХОДЫ К АВТОМАТИЗИРОВАННОЙ ОБРАБОТКЕ АНАЛИЗА СТАТИСТИЧЕСКИХ ДАННЫХ

М. Н. Лавринович, Е. М. Янкевич

*Витебский государственный университет имени П. М. Машерова,  
Республика Беларусь*

*Статья посвящена проблеме обработки статистической информации, представленной в виде электронных таблиц и PDF-документов, и поиску путей ее решения. Рассмотрены трудности, связанные с низкой совместимостью данных продолжительных динамических рядов и сложностью их автоматизированной обработки. Анализируются техническая реализация, преимущества и компромиссы каждого из подходов с точки зрения сложности разработки и уровня автоматизации для конечного пользователя.*

**Ключевые слова:** статистическая информация, совместимость, стандартизация, программное обеспечение, автоматизация.

## APPROACHES TO AUTOMATED ANALYSIS PROCESSING STATISTICS

E. M. Yankevich, M. N. Lavrinovich

*Vitebsk State University after P. M. Masherov, Republic of Belarus*

*The article addresses the problem of processing statistical information presented in the form of spreadsheets and PDF documents and explores potential solutions. The challenges associated with low compatibility of data from different periods and the complexity of its automated processing are considered. The technical implementation, advantages, and compromises of each approach are analyzed from the perspective of development complexity and the level of automation for the end user.*

**Keywords:** statistical information, compatibility, standardization, software, automatization.

С каждым годом Национальный статистический комитет Республики Беларусь совершенствует базу данных. Работает интерактивная информационно-аналитическая система, которая позволяет производить экспорт информации для обработки данных в различные форматы от CSV до SDMX-CSV [1, 2]. Однако, если учеными осуществляются дополнительные запросы в статистические органы с детализацией информации по районам, у которых представление информации постоянно совершенствуется и выделяются уже не только районы, но и районные центры, что осложняет сопоставление. Цель – исследовать возможности и подходы к обработке информации программными средствами.

Исследование динамики показателей за более продолжительное количество лет предполагает обработку показателей представленных в единообразных показателях и на основе одинаковой методики их расчета, что на деле не всегда так. Нами предлагаются два взаимодополняющих возможных решения данной проблемы. Первое заключается во введении и соблюдении стандартов, обеспечивающих наибольшую обратную совместимость статистических отчетов, и, соответственно приемлемую скорость обработки содержащейся в них информации без использования продвинутого функционала программных средств. Такое решение видится нам оптимальным, поскольку позволяет заинтересованным лицам быстро работать с уже существующими программными средствами ознакомившись лишь с базовым функционалом этих средств и не прибегая к установке и изучению дополнительного программного обеспечения [3]. Тем не менее введение подобных стандартов подразумевает труд-

ности, связанные с их разработкой, внедрением на тех или иных уровнях и адаптацией уже существующих объемов структурированных данных к новому виду. Однако соблюдение даже не слишком строгих правил структуризации и оформления статистической информации при сборе ее в сводки и таблицы, может значительно упростить реализацию второго возможного решения. Оно, в свою очередь, заключается в создании программного обеспечения, ориентированного либо на приведение информации к единому стандарту, либо непосредственно на анализ и экспорт готовой информации в таблицу в приемлемом виде.

Более того, с точки зрения технической реализации, первый вариант представляет собой конвейер данных, который включает этапы извлечения информации из разнородных источников (таблицы разных лет, PDF-документы с использованием OCR), ее трансформации с помощью алгоритмов для нормализации структуры и исправления несоответствий, и последующей загрузки в унифицированный формат. Для пользователя же этот сложный технический процесс сводится к интуитивно понятным действиям – выбору исходных файлов и нажатию кнопки «Конвертировать», что делает его простым и доступным для специалистов без глубоких технических знаний.

Второй вариант, программное обеспечение для непосредственного анализа и представления информации, технически является более комплексным решением, так как включает в себя функционал первого и дополняет его мощным вычислительным модулем и интерфейсом визуализации. Такая система не только приводит данные к общему виду, но и автоматически строит сводные отчеты и диаграммы. С точки зрения простоты использования этот подход предлагает наивысшую степень автоматизации, где задача пользователя сводится лишь к загрузке исходных данных и выбору типа отчета, полностью абстрагируя его от всех промежуточных этапов обработки.

Как видно, оба программных подхода, являются взаимодополняющими, предлагают разные компромиссы между сложностью разработки и уровнем автоматизации для конечного пользователя, при этом оба они служат одной цели – преодолению хаоса ретроспекции статистических данных.

Таким образом, преодоление проблем обработки представленной органами статистики данных требует комплексного подхода. Наиболее эффективным путем представляется синергия двух направлений: разработки унифицированных стандартов оформления отчетности и создания специализированного программного обеспечения. Стандартизация закладывает основу для долгосрочной упорядоченности данных, в то время как программные решения, будь то инструменты для конвертации или комплексные аналитические платформы, предлагают практические средства для интеграции и преодоления существующего хаоса здесь и сейчас. Ключевым критерием успеха обоих подходов является ориентация на конечного пользователя, обеспечивающая доступность технологий без необходимости обладания глубокими техническими знаниями, что в конечном итоге способствует повышению эффективности и скорости работы со статистической информацией.

#### Литература

1. Сайт Национального статистического комитета Республики Беларусь – URL: <https://www.belstat.gov.by/> (дата обращения: 10.10.2025).
2. База статистических данных ИАС БД – URL: <https://dataportal.belstat.gov.by/osids/home-page> (дата обращения: 05.10.2025).
3. Лавринович, М. Н. Возможности интеграции интерактивной информационно-аналитической системы статистической информации / М. Н. Лавринович // Молодость. Интеллект. Инициатива : материалы XIII Междунар. науч.-практ. конф. студентов и магистрантов, Витебск, 25 апр. 2025 г. : в 2 т. / Витеб. гос. ун-т ; редкол.: Е. Я. Аршанский (гл. ред.) [и др.]. – Витебск : ВГУ им. П. М. Машерова, 2025. – Т. 2. – С. 61–62.