

Л.В.ИВАНОВ, НТУ «ХПИ»

ОСОБЕННОСТИ ПРЕДСТАВЛЕНИЯ ДАННЫХ В ИНФОРМАЦИОННЫХ СИСТЕМАХ ЛАБОРАТОРИЙ, ОСУЩЕСТВЛЯЮЩИХ ХРОМАТОГРАФИЧЕСКИЙ АНАЛИЗ

В статті розглянута задача вибору та обґрунтування концепції представлення хроматографічних даних у лабораторній інформаційно-управляючій системі. Здійснено аналіз вимог належної лабораторної практики (GLP), а також проблем, пов'язаних з представленням хроматографічних даних відповідно до цих вимог. Запропоновано підхід, який об'єднує переваги сховищ даних з XML-технологіями.

In article a problem of choice and argumentation of chromatographic data representation concepts in Laboratory Information Management System is considered. Analysis of Good Laboratory Practice (GLP) requirements and concerned problems of chromatographic data representation is implemented. A research based on joining of advantages of Data Warehouses and XML technologies is proposed.

Введение. Интеграция Украины в европейское экономическое пространство требует нового подхода к контролю качества промышленной продукции. Это в первую очередь относится к так называемым регулируемым отраслям экономики, прежде всего изделий медицинского предназначения.

В настоящее время во всех странах Европы уделяется большое внимание развитию лабораторных информационно-управляющих систем (ЛИУС, или LIMS, Laboratory Information Management Systems). Данный термин точнее было бы перевести как системы управления лабораторной информацией, так как их основное предназначение заключается в обработке информации, необходимой лаборатории - регистрацию и отслеживание образцов, их распределение в обработку, ввод результатов и формирование сертификатов анализа, создание различных статистических отчетов и т.п. [1].

Достоверность результатов анализов, обеспечивающая качество продукции, важна во всех отраслях народного хозяйства. Использование ЛИУС в так называемых регулируемых отраслях экономики, прежде всего в производстве изделий медицинского предназначения, обуславливает дополнительные требования к таким системам. Все информационные процессы, происходящие в ЛИУС, должны строго соответствовать требованиям надлежащей лабораторной практики (Good Laboratory Practice, GLP) в частности дополнительным требованиям, накладываемым

на компьютеризованные системы, используемые в таких отраслях. К таким требованиям относятся: обеспечение хранения полного следа всех изменений данных без потери первичных данных, наличие связи информации с лицами, внесшими данные, и временем внесения, хранение вспомогательной информации в архиве, хранение данных, записанных вручную, как первичных, обеспечение защиты информации [2].

Среди методов лабораторного анализа, применяемых при производстве лекарственных препаратов, одним из самых распространенных и полных с точки зрения результатов является хроматографический анализ. Этот вид анализа основан на разделении веществ между двумя фазами – неподвижной и подвижной [3]. После чтения исходной, т.н. «сырой» хроматограммы, она подлежит дальнейшей обработке с целью фильтрации шумов, детекции базовой линии и пиков, в частности, разделения неразделенных пиков, выделения «пиков на хвосте» и т.д.

Для реализации обработки хроматографических данных могут использоваться различные алгоритмы, характеризующиеся, кроме всего прочего, различными наборами параметров [4]. Все это говорит о сложности задачи представления хроматографических данных и необходимости ее адекватного решения.

1. Постановка задачи исследования. Определение концепции хранения хроматографических данных включает решение ряда взаимосвязанных подзадач. Необходимо сделать выбор между использованием базы данных и файловой системы, текстовых и двоичных файлов, открытым и закрытым форматом, стандартными и специфическими форматами данных и т.д. Решение поставленных задач должно базироваться на анализе существующих подходов к их решению, в том числе, в историческом аспекте.

Практика строгого документирования результатов анализа всегда была типичной для фармакопейной отрасли всегда, в том числе и в докомпьютерную эпоху. Первый этап компьютеризации обработки данных в лабораториях связан с использованием текстовых файлов. Недостатки такого способа хранения данных очевидны: не может идти речи ни о персонализации, ни об отслеживании доступа к данным, ни о представлении необходимой дополнительной информации. В связи с невозможностью использования текстовых файлов как полноценных документов весь контроль и отслеживания информации продолжал происходить на уровне бумажных носителей. Зачастую повторные исследования требовали повторного ручного ввода исходных данных, что

кроме всего прочего увеличивало вероятность ошибок и искажений информации.

Следующий этап компьютеризации лабораторных исследований связан с появлением закрытых форматов данных, являющихся собственностью компаний-разработчиков программного обеспечения. Здесь недостатки также очевидны. Закрытость формата затрудняет процедуру валидации программного обеспечения и результатов экспериментов. Наличие в системе звеньев, недоступных для проверки делает результирующие документы невалидными с точки зрения GLP. Кроме того, стремительная эволюция операционных систем и сред программирования часто приводила к тому, что программное обеспечение оказывалось по этим причинам непригодным к повторному использованию иначе, как для воспроизведения ранее полученных данных с распечатки [5].

Современный этап развития лабораторных информационных систем характеризуется переходом к открытым форматам данных лабораторных исследований. Первые попытки стандартизации форматов аналитических данных предпринимались в восьмидесятые годы. В частности, группой ученых, занимающихся спектроскопией, был предложен формат JCAMP, основанный на использовании ASCII файлов [6]. Спецификация формата была одновременно и сложной, и неполной.

Следующая, более удачная попытка стандартизации лабораторных данных связана с созданием формата, именуемого ANDI (Analytical Data Interchange), изначально разработанного и поддерживаемого Ассоциацией аналитического инструментария (Analytical Instrumentation Association, AIA) [7]. В отличие от JCAMP, разработчики изначально нацелили свой проект на поддержку требований GLP.

Эти и другие проекты страдали общими недостатками – ориентацией на конкретные, ныне устаревшие форматы данных, невозможностью расширения. Кроме того, эти и другие открытые форматы не предусматривают хранения методик расчетов, что входит в разрез с требованиями надлежащей лабораторной практики.

Следует также учесть, что аналитические данные должны быть тесно интегрированы с системой управления лабораторией, информационными потоками, а также хорошо отражать специфику различных видов анализа, в частности хроматографического анализа.

Широкое использование хроматографии в лабораторных исследованиях, важность сохранения полной информации обо всех этапах и результатах исследований, необходимость обеспечения доступности данных в соответствии с требованиями GLP делают актуальной задачу

определения концепций представления и адекватных форматов хроматографических данных.

2. Концепция представления хроматографических данных и ее составляющие. Сравнительный анализ методов представления данных, обеспечивающих открытость, доступность, читаемость, свободно расширяемый формат и переносимость, позволяет сделать выбор в пользу XML-документов. В настоящее время такой подход уже нашел свое применение в аналитической химии. Разработано большое количество языков представления аналитических данных, в том числе ставших промышленным стандартом [5]. Многие языки претендуют на общность представления. При этом, безусловно, сокращается возможность представления данных, специфических прежде всего для хроматографического анализа. Следовательно, перспективным является подход, позволяющий расширить универсальные языки представления аналитических данных, дополнив их тегами, специфичными для хроматографии.

Вместе с тем, открытость XML-формата делает затруднительным использование XML-документов для хранения информации, позволяющей идентифицировать все этапы анализа и конкретные его участники. Будучи свободно доступной, такая информация может быть преднамеренно искажена или вовсе утеряна, а следовательно, снова всплывает необходимость дублирования информации в текстовых документах.

Требованиям GLP, предполагающим хранение всей истории проведения анализов, начиная с хорошо соответствует концепция хранилищ данных [8]. Хроматографическая информация вполне удовлетворяет требованиям хранилища данных, так как является предметно-ориентированной, интегрированной, поддерживает хронологию и практически не изменяется во времени. Корпоративное хранилище данных поддерживает историческую целостность данных. В хранилище попадает исчерпывающий набор исходных данных из различных источников, который хранится в своем первоначальном состоянии и не удаляется. Для исходных данных допускается некоторое нарушение целостности и непротиворечивости.

Недостатком традиционных хранилищ данных, построенных на реляционной модели, является неадекватность этой модели самим хроматографическим данным.

Эффективное решение задачи может быть получено за счет применения технологии непосредственного хранения XML-документов в

реляционной базе данных. Такой подход обеспечивается многими современными СУБД, в том числе Oracle [9]. При этом сами хроматограммы импортируются, хранятся и экспортируются в XML-формате. Таким образом, речь идет о сочетании преимуществ хранилищ данных и XML-технологий.

В XML-ориентированных базах данных документы записываются в своем первоначальном состоянии, фиксируются системные даты – поступления документа, установление связей между элементами различных документов, прекращения действия документа.

Динамика развития методов обработки хроматографических данных определяет появление новых более эффективных математических методов, совершенствования уже существующих и т.д. Формат хранения хроматографических данных должен допускать представление данных, как о результатах обработки, так и о промежуточных данных, в том числе параметрах методов, использованных для обработки. Однако фиксация конкретного набора методов приводит к преждевременному моральному устареванию форматов данных и снижению их универсальности.

Для разрешения указанной проблемы можно воспользоваться накопленным опытом по представлению в XML-документах процедур преобразования. Это могут быть как стандартные языки, построенные на XML (например, XSLT), так и специализированные, отражающие специфику конкретной предметной области. Такое расширение представлений о хроматографических данных, безусловно требует разработки нового программного обеспечения. Однако отделение хранения алгоритмов от программ, реализующих эти алгоритмы, увеличивает срок жизни программного обеспечения и отодвигает время его морального устаревания.

3. Выводы. В процессе исследований обозначены проблемы, связанные с представлением данных аналитических лабораторий, обозначены требования, предъявляемые к представлению результатов лабораторных исследований в регулируемых отраслях экономики, в частности в производстве изделий медицинского назначения.

Проанализированы особенности хроматографической информации, требования к представлению хроматографических данных и процедур их обработки.

Выполнен анализ эволюции концепций представления форматов аналитических данных, стандартные подходы и языки, используемые в этой отрасли, их достоинства и недостатки.

Поставлена задача исследования, заключающаяся в определении

концепций представления и адекватных форматов хроматографических данных.

В результате исследования определены основные составляющие концепции представления хроматографических данных – использование XML-формата, использование технологий корпоративных хранилищ данных, непосредственное хранение в базе данных хроматограмм в виде XML-документов, а также хранение в XML-документах информации о процедурах обработки данных.

Список литературы: 1. *Paszko C., Turner E.* Laboratory Information Management Systems Revised & Expanded, CRC Press, 2001. – 242 p. 2. Good Practices for Computerized Systems in Regulated “GxP” Environments. – Geneva, Pharmaceutical Inspection Convention, 2003. – 50 p. 3. *Столяров Б.В., Савинов И.М., Витенберг А.Г., Карцова А.А.* Практическая газовая и жидкостная хроматография. СПб.: Изд-во С.-Петербург. ун-та, 1998. – 610 с. 4. *Felinger, A.* Data Analysis and Signal Processing in Chromatography. Elsevier:Amsterdam, 1998. – 428 p. 5. *Roth A., Jopp R., Schäfer R., Kramer G.W.* Automated Generation of AnIML Documents by Analytical Instruments, Journal of the Association for Laboratory Automation, 2006, 11, p. 247. 6. *McDonald R. S., Wilks P. A.* JCAMP-DX: A Standard Form for Exchange of Infrared Spectra in Computer Readable Form, Appl. Spec., Vol. 42, 1988, p. 151. 7. ASTM E1947-98, Standard Specification for Analytical Data Interchange Protocol for Chromatographic Data. 8. *Иванов Л.В., Косвина Е.Б.* Использование intranet-технологий в компьютеризированной системе обработки хроматографических данных. // Вісник Національного технічного університету «Харківський політехнічний інститут»: Збірка наукових праць. Тематичний випуск: Системний аналіз, управління та інформаційні технології. Харків, НТУ «ХПИ». 2004. - №1. С. 93 – 99. 9. Scardina, M. , Chang, B, Wang, J. Oracle Database 10g XML & SQL: Design, Build, & Manage XML Applications in Java, C, C++, & PL/SQL. , Osborne, 2004, – 600 p.

Поступила в редколлегию 18.02.2008