



Анализ и разработка характеристик качества геоданных

Статья проводит анализ качества геоданных и ГИС. На основе анализа предлагается набор характеристик качества. Эти характеристики отражают статику и динамику процессов в геоинформатике. Характеристики качества являются универсальными, так как применимы для оценки качества геоинформационных технологий, геоинформационных систем и геоданных. Для обозначения разных объектов применяется обобщенное понятие информационная конструкция.

Ключевые слова: философия информации, информация, качество, качество геоданных, информационные системы, геоинформационные системы, геоданные



Analysis and development of the quality characteristics of geodata

The article analyzes the quality of geodata and GIS. Based on the analysis provides a set of quality characteristics. The proposed specifications reflect the statics and dynamics of processes in geoinformatics. Quality characteristics are universal. They are used to evaluate the quality of different objects: geoinformation technology, geoinformation systems and geodata. The article introduces the concept of a generalized information structure to denote different objects

Keywords: Information philosophy, information, quality, geodata quality, information systems, geoinformation systems, geodata

Введение

Необходимость получения и применения качественных геоданных и необходимость оценки качества их обработки требует разработки характеристик качества. Эта проблема обусловлена также тем, что множественность используемых данных в технологиях ГИС порождает проблему их совместного использования. Разнообразие геоданных порождает проблему их согласования, которая связана с проблемой эффективной эксплуатации баз данных ГИС и внешних баз данных, связанных с ГИС. При анализе, проводимом в данной статье, будем разграничивать пространственное моделирование и геоинформационное моделирование и геоинформационные технологии и ГИС технологии. Геоинформационные технологии более

широкий класс технологий, связанный с получением геоданных и использованием ГИС и прочих информационных систем. ГИС-технологии – часть геоинформационных технологий связанная только с применением ГИС. В настоящее время основой построения моделей в геоинформатике служат геоданные [1-5] как универсальные системные [6] конструкции.

Процессы и процедуры, требующие контроля качества в ГИС

Затраты на оценку качества. Одной из важных процедур в ГИС является процедура контроля качества (например, качества векторизации). На нее может тратиться до 30% времени от всего технологического цикла, что обусловлено тем, что часто некачественную продукцию выгоднее пе-

ределать заново, чем редактировать [7]. То есть в зависимости от уровня ошибок модель либо редактируют либо получают заново. Это приводит к необходимости создания системы контроля качества, которая, в свою очередь, должна основываться на эффективных оценках качества информации

Векторизация. При векторизации растровой информации может возникнуть проблема стоимости качества, которая обусловлена противоречием с одной стороны получения качественной (по точности и достоверности) информации, с другой уменьшением времени (и соответственно снижением стоимости) на векторизацию, которая может составлять до 60 % в технологиях создания цифровых карт [7].

Увеличение производительности (в основном за счет скорости) векторизации дает существенное снижение себестоимости геоинформационной продукции. На практике ускорить полный технологический цикл создания цифровых карт для ГИС за счет автоматической либо полуавтоматической векторизации возможно на 20-30 процентов и за счет ухудшения качества. Этот фактор требует введения процедуры принятия решений при анализе качества геоинформационных данных, которые немислимы без адекватных оценок качества информации.

Обмен библиотеками условных знаков. Обычно способы кодировки условных знаков в различных системах различны. В существующих программных продуктах ГИС для Windows могут включать типовые способы передачи знаков - через обменные форматы, clipboard и др [8]. Разработка стандартных требований к библиотекам условных знаков, обладающих свойствами мобильности (переносимости) возможны при наличии эффективных оценок качества информации в ГИС.

Проблемы классификации и организации информации. Любая геоинформационная система предполагает организацию вводимой графической и атрибутивной информации в виде совокупности "слоев", "покрытий", "композиций" и т.д. для дальнейшего ее использования. Технология организации объектов по слоям определяется чаще всего технологией конкретной ГИС или заказчиками.

В результате произвола в таком подходе одна карта может быть преобразована в разное количество слоев по тематическим или функциональным характеристикам. Это приводит к тому, что информация полученная с одной и той же карты становится плохо сопоставимой после обработки в разных ГИС одного типа, но работающих с разными технологиями и классификаторами.

Дигитализация карт разными исполнителями дает разные результаты по объему и содержанию. Эти антропогенные факторы определяют необходимость разработки обобщенных оценок пригодности конечных цифровых карт не только для используемой ГИС, но и для других.

Составление технического задания на решение конкретных задач с помощью ГИС должно опираться на систему оценок результатов решения этих задач. Составление технического задания при неполном учете факторов качества приводит в итоге к результатам работы ГИС, не удовлетворяющей заказчика. Наличие ошибок может быть следствием некорректно составленного технического задания. Это также требует разработки системы оценок качества, которые могли бы быть использованы при составлении ТЗ на проектирование ГИС или на проект информационной продукции, получаемой с помощью геоинформационных технологий.

Для того, чтобы оценка качества информации в ГИС и геоинформационных технологиях была достаточно объективной необходимо:

1. Разработать сопоставимую систему оценки качества для разных ГИС-технологий.
2. Осуществить подбор тех свойств и параметров, которые наиболее полно и объективно описывают анализируемые средства (пространство свойств),
3. Осуществить выбор количественных критериев, на основе которых будет производиться оценка качества ,
4. Разработать процедуры поддержки принятия решений при анализе информации и ее классификации.

Необходимость в последней процедуре определяется тем, что, как правило, имеется весьма узкое информационное поле об исследуемых средствах и о возможных ситуациях, которые могут появиться в процессе эксплуатации этих средств (эффект неопределенности).

Анализ оценок качества. Для реализации отмеченных выше требований предлагается введение трех типов оценок. Эти три типа включают:

- общие оценки, характеризующие полноту свойств (пространство параметров);
- специальные оценки, отражающие специфику организации данных в ГИС (по бальной системе),
- комплексные оценки (совокупность взаимосвязанных оценок, например, профили).

Общие оценки основаны на системном подходе, семиотической оценке моделей, структурном анализе систем, методах стандартизации. Набор общих оценок использует дихотомический принцип [9]. При анализе информации на основе общих оценок специалист может использовать дихотомические оппозиционные величины [10], например значения "да", "нет" или 1, 0. Это позволяет применять мультипликативный Q_m (1) и аддитивный Q_a (2) показатели качества свойств для данной технологии вида

$$Q_m = Q_1 * Q_2 * Q_3 * \dots * Q_n \quad (1)$$

$$Q_a = Q_1 + Q_2 + Q_3 + \dots + Q_n \quad (2)$$

Здесь n - набор характеристик качества, определяющих качество данной технологии. Этот набор согласовывается с заказчиком на основе существующих стандартов. Особенность данного показателя является его универсальность и адаптивность [11] для разных технологий. Для любого набора любых технологий:

$Q_m=1$ - набор технологий обладает качественной полнотой;

$Q_m=0$ - набор технологий не обладает качественной полнотой.

Во втором случае работает аддитивный показатель $Q_{a1} > Q_{a2}$. Это означает что набор технологий 1, которому соответствует Q_{a1} более качественный в сравнении с набором технологий Q_{a2} .

Оценка качества может проводиться не только для технологий, но и для формирования моделей [12] или информационных единиц. Формулы (1) и (2) применимы и для этих случаев. Поэтому для обобщения анализа технологий и моделей будем употреблять обобщающий термин информационная конструкция [13].

Удобство показателей Q_m и Q_a в том, что они допускает варьирование разных технологий (моделей) и наборов показателей качества, но оценка полноты получается одинаково для разных систем и технологий. Он упрощает и исключает дальнейший анализ качества в случае неполноты. Этим экономится время на ненужный дополнительный анализ. Такой показатель достаточно прост и его оценка может выполняться специалистами операционного уровня или независимыми экспертами [7].

Подобная оценка применима для случая, когда характеристики качества носят закрытый характер и к ним ограничен доступ. В этом случае к полной информации может быть допущен ограниченный круг лиц, в то время как исходные данные собираются независимо разными лицами.

Вторая группа оценок требует количественных расчетов, что позволяет получать числовые меры оценки и последующего анализа полученных значений оценок. Этот тип оценок назовем количественно-аналитическими. Такие оценки требуют для анализа привлечения специалистов среднего звена.

Третий вид оценок назовем экспертными, поскольку они могут определяться набором разнотипных значений, например, метод защиты данных может иметь разный тип, разный уровень защиты или несколько методов защиты одновременно. Такие оценки требуют когнитивного моделирования и анализа.

Наиболее ярким представителем оценок этого уровня является профиль [14], определяемый как совокупность нескольких стандартов и/или других нормативных документов, предназначенная для реализации функции или группы функций по оценке качества в заданном пространстве свойств. На базе одной и той же

совокупности характеристик и совокупностей количественных оценок второго уровня могут формироваться различные профили для разных типов ГИС-технологий и областей применения ГИС. Примером такого подхода является теория предпочтений.

Примером использования профилей для оценки качества могут служить профили защиты информации [14]. Профили оценки качества информации в ГИС унифицируют и регламентируют большую часть требований, характеристик, показателей качества объектов и процессов, выделенных и формализованных на базе стандартов и нормативных документов.

Оценка качества геоданных

Рассмотрим общие оценки, определяющие характеристики для определения качества геоданных. Модель геоданных, основанная на семиотическом подходе независимо от способов построения и видов моделей требует наличия в каждой модели (данных) информации трех ее качественных частей:

- синтаксис - правила построения и критерии принадлежности к виду;
- семантику - методы отображения информации объекта в информацию модели и саму эту информацию;
- прагматику - полезность, методы оценки полезности модели.

Отсутствие какой либо из трех упомянутых частей приводит $Q_m=0$ – данная информационная конструкция не обладает качественной полнотой. То есть говорит о неполном построении данной информационной конструкции.

Рассматривая геоданные как информационный продукт, можно применить типовые оценки качества информации [15-17]. Основными характеристиками качества информации, наряду с общепринятыми для других видов продукции, являются: репрезентативность, содержательность, прагматизм, достаточность, точность, актуальность, устойчивость сертификат безопасности, надежность.

Репрезентативность информации (геоданных) связана с правильностью ее отбора и формирования в целях адекватного отражения реальности или свойств объектов.

Содержательность информации (геоданных) определяется либо коэффициентом информативности, т.е. отношением количества синтаксической информации к ее общему объему, либо коэффициентом содержательности отношением семантической информации к ее общему объему.

Полнота информации (геоданных) характеризует необходимый набор данных для решения задач или принятия решений.

Актуальность информации (геоданных) характеризует ее временной аспект. Она определяется степенью ценности информации на момент ис-

пользования.

Точность информации (геоданных) оценивается чаще всего с помощью числовых мер. Она определяется степенью соответствия данных к реальному состоянию процесса или объекта.

Сертификат безопасности [18] информации (геоданных) – документ или задаваемая совокупность параметров, удостоверяющие соответствие мер защиты информации набору определенных требований по защите от несанкционированного доступа к информации. СБИ дает право разработчику и заказчику на использование информации как совокупности данных, защищенных на основе существующих стандартов информационной безопасности.

Рассматривая геоданные как продукцию вообще можно дать определение их надежности на основе международного стандарта ISO 9000 - 9004. Надежность геоданных - свойство сохранять в течение требуемого интервала времени способность правильно выполнять заданные спецификацией правила переработки информации в реальных условиях эксплуатации.

Рассматривая геоданные как продукцию производственно-технического назначения можно дать другое определение надежности. Согласно такому подходу, надежность геоданных (ГД) – комплексное свойство, состоящее как и в случае технических объектов из набора характеристик. Это – корректность, устойчивость, восстанавливаемость и исправляемость.

Корректность ГД определяется как соответствие реальных методов обработки заданным алгоритмам при отсутствии ошибок, превышающих допуски. Как известно, вычисления с помощью компьютера всегда содержат ошибки, обусловленные с одной стороны влиянием техники (вычислительные погрешности), с другой - спецификой методов вычислений (методические погрешности). Если эти погрешности в сумме не снижают точность выходных данных до недопустимых границ можно говорить о корректности данного вида ГД.

В процессе функционирования ГД подвергаются несанкционированным воздействиям со стороны программы (П), техники (Т), дополнительной информации (ДИ) и человека - оператора (Ч). Типы этих воздействий следующие: П → ГД; Т → ГД; ДИ → ГД; Ч → ГД.

Способность геоданных и программных продуктов противостоять этим воздействиям определяет устойчивость ГД. Устойчивость геоданных - это свойство давать правильные результаты в процессе их использования при наличии воздействий разных типов: (П → ГД; Т → ГД; ДИ → ГД; Ч → ГД). В значительной степени этот показатель определяется и взаимосвязан с интеллектуальностью интерфейса ГИС.

ГИС содержит в своем составе базу данных. Кроме того, многие процессы, например тематическое картографирование, построены на

системе запросов. Эти процессы длительны по времени. По этой причине можно допустить возможность случаев, когда происходят сбои в работе. Для такой ситуации важно не начинать процесс обработки данных заново, а вернуться к этапу обработки, предшествующему сбою.

Восстанавливаемость определяет свойство ГД, задаваемое технологическими возможностями ГИС, сохранять промежуточные данные, что позволяет возобновлять процесс информационного взаимодействия с этапом, предшествующего прерыванию обработки и остановки работы программного средства. На практике это свойство реализуется с помощью механизма создания набора временных файлов (например в Windows) или механизма “откатов” в СУБД.

В процессе хранения ГД довольно часто возникает необходимость изменения или дополнения для актуализации данных. Это обуславливает необходимость их частичной переработки. Повторение технологии получения набора данных при актуализации экономически не выгодно. Для решения подобной проблемы ГД должно обладать свойством исправляемости.

Исправляемость – свойство ГД, создающее возможность модернизации или изменений данных (цифровых моделей, цифровых карт) без их радикальной переработки. На практике это свойство может быть реализовано разными путями, например структуризацией данных или методами обработки, основанными на концепции “открытых систем”, позволяющей дополнять и обрабатывать данные с помощью различных приложений.

Таким образом, для обеспечения надежности ГД в документацию, описывающую его возможности и показатели, должны обязательно входить рассмотренные выше корректность, устойчивость, восстанавливаемость и исправляемость.

Одно из традиционных средств повышения надежности информации - избыточность. Она обычно реализуется за счет контроля и коррекции входных данных, промежуточных и окончательных результатов. Отсутствие или недостаточность этого контроля является одним из основных факторов утраты работоспособности систем обработки данных и управления (СОДУ), вызывает до 50% всех программных отказов, связанных с ошибками в данных.

Другим видом избыточности является использование механизма контрольных точек процесса обработки, фиксирующего состояние вычислительной среды в определенные моменты времени и позволяющего выполнять рестарты для промежуточных результатов вычислений.

Заключение

Значение оценки качества существенно возрастает при переходе от информационных к ин-

теллектуальным технологиям [20, 21]. Отчасти это обусловлено проблемой больших данных [22], которая заостряет проблему качества геоданных. Предлагаемый в статье набор оценок

качества ГД и ГИС является универсальным. Он позволяет оценивать системные признаки качества и проводить сравнительный анализ качества разных информационных конструкций.

ЛИТЕРАТУРА

1. Омельченко А. С. Геоданные как инновационный ресурс // Качество, инновации, образование. 2006. №1. С.12-14.
2. Матчин В.Т. Формирование геоданных // Славянский форум, 2015. 2(8). С.185-193.
3. Коваленко Н. И. Глобализация, пространственная информация, геоданные // Славянский форум, 2015. 4(10). С.147-156.
4. Маркелов В. М. Добыча данных и геоданных // Образовательные ресурсы и технологии. 2015. №2 (10). С.126-131.
5. Цветков В.Я. Модель геоданных для управления транспортом // Успехи современного естествознания. 2009. №4. С.50-51.
6. Савиных В.П., Цветков В.Я. Геоданные как системный информационный ресурс // Вестник Российской Академии Наук, 2014, том 84, № 9, С. 826–829. DOI: 10.7868/S0869587314090278.
7. Дышленко С. Г. Разработка технологии адаптивного проектирования ГИС: дисс. ... канд. техн. наук: 25.00.35 – Геоинформатика. М.: МГУПС, 2009. 138 с.
8. Цветков В.Я., Дышленко С.Г. Применение ГИС «Панорама» при инженерных изысканиях // Инженерные изыскания. 2009. № 12. С.46-48.
9. Tsvetkov V.Ya. Dichotomous Systemic Analysis. Life Science Journal 2014, 11(6). pp.586-590.
10. Tsvetkov V. Ya. Opposition Variables as a Tool of Qualitative Analysis // World Applied Sciences Journal. 2014. 30 (11). p.1703-1706.
11. Дышленко С.Г., Адаптивное проектирование ГИС для изысканий с использованием GIS ToolKit // Инженерные изыскания. 2010. №5. С.48-51.
12. Дружинин Г. В., Сергеева И. В. Качество информации. М.: Радио и связь, 1990.
13. Tsvetkov V. Ya. Information Constructions // European Journal of Technology and Design, 2014, Vol (5), № 3. p.147-152.
14. Иванников А.Д., Кулагин В.П., Тихонов А.Н., Цветков В.Я. Информационная безопасность в геоинформатике. М.: МаксПресс 2004. 336 с.
15. Цветков В.Я. Качество экономической информации //Успехи современного естествознания. 2008. №7. С. 84-85.
16. Titman S., Trueman B. Information quality and the valuation of new issues // Journal of Accounting and Economics. 1986. V 8, № 2. p.159-172.
17. V.Ya. Tsvetkov. Assessment of Quality of Vectorization Based Information // Biosciences biotechnology research Asia, November 2014. Vol. 11(Spl. Edn.), p. 211-213. doi: <http://dx.doi.org/10.13005/bbra/1464>
18. Мельников Д. Информационная безопасность открытых систем. – Litres, 2015.
19. Tsvetkov V. Ya. Information interaction // European Researcher, 2013, Vol.(62), № 11-1. p.2573-2577.
20. Шайтура С.В. Интеллектуальный анализ геоданных // Перспективы науки и образования. 2015. №6. С.24-30.
21. Железняков В.А. Интеллектуальное обновление информации в банке геоданных // Инженерные изыскания. 2012. № 5. С.58-61.
22. Павлов А.И. Большие данные в фотограмметрии и геодезии // Образовательные ресурсы и технологии. 2015. №4 (12). С. 96-100.

Информация об авторе
Дышленко Сергей Геннадьевич
 (Россия, Москва)
 Кандидат технических наук
 Начальник отдела ЗАО КБ «Панорама»
 E-mail: dishlenko@yandex.ru

Information about the author
Dyshlenko Sergey Gennadievich
 (Russia, Moscow)
 PhD in Technical Sciences
 Head of Department, JSC Design Bureau "Panorama"
 E-mail: dishlenko@yandex.ru