

ISSN 0869-6063

Номер 1

Январь - Февраль - Март
2005



РОССИЙСКАЯ АРХЕОЛОГИЯ



<http://www.malk.ru>



“НАУКА”

ЦИФРОВОЙ АРХИВ МАТЕРИАЛОВ ИССЛЕДОВАНИЙ ГНЁЗДОВСКОГО АРХЕОЛОГИЧЕСКОГО КОМПЛЕКСА

© 2005 г. М. О. Жуковский, Т. А. Пушкина

Московский государственный университет

Внедрение компьютерных технологий в обработку археологической информации в России началось более десяти лет назад. С начала 1990-х годов вышли в свет многочисленные статьи и ряд сборников, посвященные теоретическим и методическим проблемам применения информационных технологий в исторических исследованиях, разработке археологических баз данных, экспертных и информационных систем, компьютерной исторической и археологической картографии (Базы данных..., 1995; Методы..., 1995; см. также сборники конференций ассоциации «История и компьютер» за 1994–2001 гг.). Работы в этой области ведутся и на кафедре археологии исторического факультета МГУ им. М.В. Ломоносова (Ениосова и др., 2003. С. 88–103).

Несколько лет назад авторами был разработан проект, призванный объединить в многофункциональной информационной системе (ИС) весь массив научных знаний о Гнёздовском археологическом комплексе для обеспечения их надежного хранения, свободного доступа, аналитической обработки, представления и распространения в научных и образовательных целях.

Цель данной статьи – показать технологию и методику построения ИС, охватывающих разнородный археологический материал, с применением новейших компьютерных технологий и продемонстрировать богатейший потенциал, который они предоставляют, выводя исследования на принципиально новый уровень.

В настоящее время Гнёздовский археологический комплекс – один из крупнейших комплексов археологических памятников периода формирования Древнерусского государства – состоит из шести курганных групп, нескольких одиночных курганов, двух городищ и одного селища, связанных временем возникновения и существования и характером найденных при исследованиях материалов (АКР, 1997. С. 82–92). Общая площадь, занимаемая памятниками, насчитывает около 92 га. История археологического изучения Гнёздова насчитывает 130 лет (Авдусин, 1999). В 1874 г. начались раскопки курганов, первым исследователем которых был М.Ф. Кущинский. В 1881–1900 гг.

курганы раскапывались В.И. Сизовым, В.Д. Соколовым, А.С. Уваровым. В 1898–1901 гг. С.И. Сергеев исследовал значительную часть курганов на территории, отчуждаемой железной дорогой. С 1905 по 1922 гг. небольшое количество курганов в разных частях могильника с перерывами было раскопано И.С. Абрамовым, В.А. Городиловым, Е.Н. Клетновой. Если вплоть до конца XIX в. в Гнёздове исследовались только курганы, то уже в 1900–1901 гг. С.И. Сергеев впервые приступил к раскопкам Центрального поселения – им произведена прорезка вала городища. Затем в 1905 г. Ольшанское городище было исследовано И.С. Абрамовым, который прорезал вал городища и заложил несколько небольших шурфов на его площадке. Проведенное в 1922–1923 гг. А.Н. Лявданским сплошное обследование Гнёздова и его ближайшей окрестности привело к открытию селищ вокруг Центрального и Ольшанского городищ. Однако внимание исследователей по-прежнему было обращено в первую очередь на курганы и Центральное городище. В 1940 г. экспедиция Смоленского краеведческого музея и Смоленского педагогического института предприняла раскопки нескольких курганов и площадки Центрального городища. В 1949 г. раскопки в Гнёздове начал Д.А. Авдусин (Смоленская археологическая экспедиция МГУ). Большое значение имели стационарные исследования западной части центрального поселения, проведенные И.И. Ляпушкиным (экспедиция ЛОИА АН СССР) в 1967–1968 гг. С 1970 г. Смоленская экспедиция МГУ ведет работы одновременно в разных курганных группах и на поселении (Авдусин, 1999. С. 12–20). В 1996 г. экспедицией был открыт неизвестный ранее участок древнерусского селища, расположенный в затопляемой ныне пойме Днепра (Мурашева и др., 2001).

В результате проведенных исследований было изучено более 1270 курганных насыпей: 560 – до-революционными раскопками и 714 – за период с 1949 по 2003 г. В ходе раскопок поселения было заложено более 80 шурфов и раскопов общей площадью более 6 тыс. м². За длительный период масштабного изучения памятников Гнёздова накопился огромный архивный материал, представ-

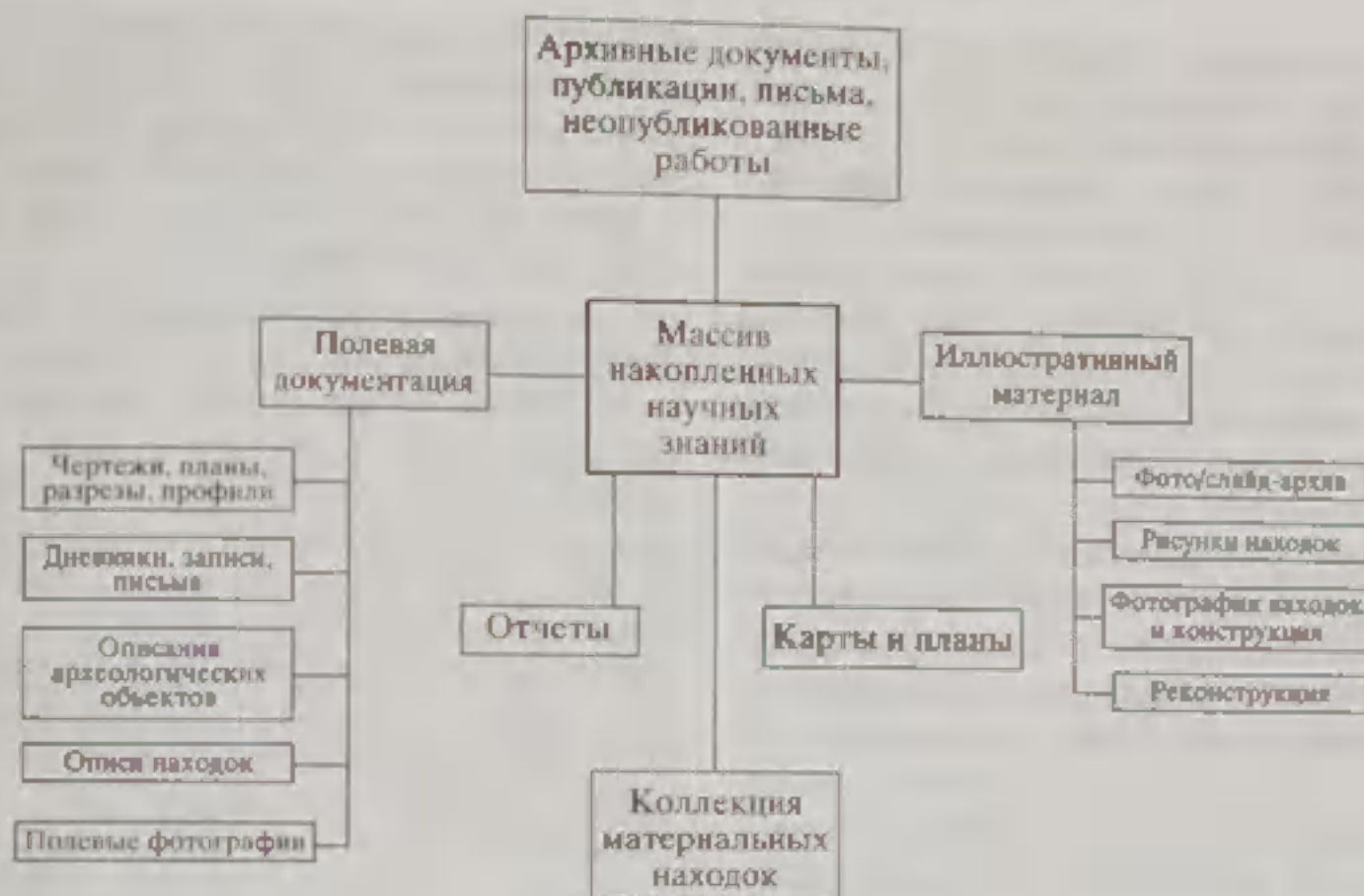


Рис. 1. Структура источников, включенных в информационную систему.

ленный подлинными полевыми дневниками, записками, письмами, отчетами, чертежами, описаниями, фотографиями и рисунками находок, археологическими и топографическими картами, планами и схемами, архивными картами и рисунками. Все данные, собранные различными исследователями и в разное время, можно разделить на несколько основных групп, представленных на рис. 1.

Анализ источниковой базы позволяет выделить ключевые факторы, определяющие актуальность проекта:

1. Полнота, качество, доступность и сохранность архивных документов, хранящихся в разобщенном состоянии в архивах ИИМК и ИА РАН, архиве кафедры археологии исторического факультета МГУ, Государственном Историческом музее, Государственном архиве Смоленской области и РГАДА, очень неравнозначны. Большинство материалов существуют в единственных экземплярах и находятся в ветхом и малопригодном для многократного использования состоянии. Необходимы сбор, каталогизация, определение объема и состояния накопленных данных, а также обеспечение надежности хранения и резервное копирование информации.

2. Отсутствие сводного реестра материальных находок, происходящих с территории Гнёздовского археологического комплекса за все время его изучения, фотографий вещевого материала. Коллекции находок хранятся в Государственном Историческом музее, Государственном Эрмитаже, на кафедре археологии исторического факультета МГУ им. М.В. Ломоносова, в Смоленском государственном музее-заповеднике, часть ее попросту

недоступна, часть депаспортизована¹. Помимо этого, в течение долгого времени значительная часть коллекций не сверялись.

3. Отсутствие полноценной картографической подосновы на территорию Гнёздовского археологического комплекса. Археологические топографические планы и схемы Гнёздовского археологического комплекса, выполненные в разные годы, фрагментарны, разобщены и требуют уточнения. До сих пор в распоряжении исследователей Гнёздова не было крупномасштабных общегеографических, тематических (геологических, почвенных, климатических, гидрологических и т.д.) и топографических карт и планов, а также данных дистанционного зондирования земной поверхности.

4. Насущная необходимость интеграции и обобщения накопленной археологической информации. Длительный период изучения Гнёздовского археологического комплекса привел к накоплению огромного массива разнородных и дезинтегрированных данных о памятнике, собранных разными исследователями по различным методикам и с разной полнотой. Непрерывное поступательное развитие археологической науки, методики рас-

¹ По объективным причинам нам оказались недоступны археологические материалы, хранящиеся в Государственном Эрмитаже и Смоленском государственном музее-заповеднике. В то же самое время мы приносим искреннюю благодарность сотрудникам Отдела археологических памятников ГИМ, архива ИА РАН и архива ИИМК РАН за понимание и оказанную помощь в сборе материалов. В работу вошли неопубликованные материалы раскопок Центрального селища, любезно предоставленные автором раскопок В.В. Мурашевой. Эти материалы имеют большое значение, поскольку именно они позволили обратиться к трехмерному моделированию.

копок, правил сбора археологической информации послужило причиной неравнозначного состава и появления существенных пробелов в различных типах археологических данных о Гнёздове. Часть информации до сих пор остается необработанной. В настоящий момент остро стоит проблема взаимной сверки, унификации, интеграции и всестороннего анализа всего объема собранного фактического материала с использованием единого набора правил представления и кодирования данных.

5. Необходимость подготовки к публикации и введения в научный оборот неизданных материалов исследований памятника, а также распространения, презентации и широкого использования в учебном процессе всего массива научных знаний о Гнёздове.

6. Необходимость обеспечения заинтересованных исследователей, аспирантов и студентов комплексом современных средств, позволяющих осуществлять быстрый параллельный доступ и многоплановый анализ информации, накопленной за весь период научного изучения Гнёздова.

Было выделено восемь основных стадий разработки проекта.

1. *Каталогизация, классификация, определение состояния, места хранения и доступности информации.* На этой стадии проводилась предварительная каталогизация и общая оценка объема данных. За 130 лет полевых работ было извлечено более 21700 индивидуальных находок, вычерчено более 4500 чертежей, сделано около 7000 фотографий. Предварительный анализ карт и планов, охватывающих территорию памятника, показал большую неоднородность этого материала и наличие в нем, порой, существенных пробелов и ошибок. Это связано с широким хронологическим диапазоном создания картографических материалов, начиная с карты генерального межевания XVIII в. и заканчивая топографическим планом, выполненным в 1970-х годах. Дополнительные трудности связаны с использованием авторами картографического материала различных систем измерения и масштабов.

2. *Предварительный анализ и подготовка информации, выявление недостающих данных и их восполнение, насколько это возможно; пополнение и расширение массива данных.* На этом этапе проводилась подробная оценка и инвентаризация каждой единицы хранения. В итоге было установлено, что фактический состав, состояние и наличие архивных данных в целом ряде случаев не соответствует информации, приведенной в каталогах и описях, имеющихся в официальных источниках. Индивидуальное изучение каждой доступной единицы хранения позволило уточнить созданные ранее сводные каталоги. Некоторые материалы, например часть полевых чертежей,

хранятся в нескольких неравнозначных копиях. Для оцифровки были отобраны оптимальные экземпляры, с точки зрения соотношения качество/доступность. Благодаря копиям удалось также в ряде случаев восполнить недостающие оригинальные экземпляры.

Для решения задачи корректной взаимной ректификации картографических данных в 2001 г. был снят самый детальный на настоящий момент топографический план Центрального поселения памятника в масштабе 1 : 500, принятый в качестве основы для верификации и исправления остального материала крупного масштаба. Для уточнения имеющихся крупномасштабных планов и выявления элементов земной поверхности, в первую очередь антропогенного происхождения, не видимых на местности было привлечено несколько листов зимней аэрофотосъемки (К-1703, 1707, 1709-1713, 1717 15.02.1998). Блок тематических карт был дополнен картами четвертичных отложений и инженерно-геологического районирования.

Массив источников был расширен за счет создания изначально цифровых данных – фотографий, круговых фотопанорам и видео, отснятых на памятниках комплекса в 2000–2003 гг.

Одной из наиболее масштабных задач на втором этапе проекта стало натурное обследование курганных групп Гнёздовского археологического комплекса, в результате которого в 2002 г. была проведена инвентаризация сохранившихся на настоящий момент курганных насыпей. Подобной сверки не производилось с 1950-х годов, когда М.Н. Кисловым была выполнена съемка планов курганных групп. Большая часть паспортов обследованных археологических объектов была снабжена цифровыми фотографиями.

3. *Разработка структуры ИС и ее компонентов, определение форматов хранения и представления информации.* Для создания системы использовался целый ряд современных технологий компьютерного хранения, обработки и анализа информации. Выбор технологических средств для построения ИС обусловлен рядом требований²: а) предоставление “контейнера” для хранения информации, адекватного данным источника. “Контейнеры” должны обеспечивать наименьшие потери информации при оцифровке и гибкие возможности постобработки и трансформации данных;

² В литературе существует мнение, что для эффективной обработки археологической информации нельзя использовать “коммерческие” информационные технологии (Матехин, Богуславский, 1997). Позволим себе принципиально не согласиться с этой точкой зрения. Современные “коммерческие” информационные технологии, в том числе СУБД, предоставляют полнофункциональный набор средств обработки, хранения и организации данных независимо от их происхождения. Акцент должен делаться не на разработку собственных специфических археологических приложений, а на программную настройку наиболее распространенных и мощных “коммерческих” инструментов.

б) обеспечение надежного хранения и целостности данных; в) технические средства должны обладать адекватным инструментарием для анализа данных и решения поставленных проектных задач; г) обеспечение совместимости блоков информации разных типов; д) обеспечение совместимости технических компонентов на уровне архитектуры системы; е) обеспечение свободного доступа и извлечения информации.

Рассмотрим "контейнеры" и форматы хранения данных, использованные в нашей ИС.

По данным американского Института изучения окружающей среды, (ESRI) около 80% всей информации в мире имеет пространственные признаки (*spatial attributes*). В нашем случае, применительно к археологической информации эта доля, очевидно, еще больше. Действительно, помимо собственно картографического материала пространственными признаками насыщены все типы материалов полевой документации, полевые отчеты, паспорта коллекций материальных находок. Оптимальной средой для организации неоднородной информации и построения взаимосвязей между ее блоками на основании пространственных признаков являются геоинформационные системы (ГИС). ГИС работают с информацией на уровне графических векторных объектов, которые могут логически объединяться в группы и слои. Благодаря этому, ГИС обладает мощным набором инструментов как для традиционной аналитической, так и пространственной аналитической и визуальной обработки данных.

Существует два типа геоинформационных систем. К первому относятся специально созданные системы, использующие собственное ядро (*engine*) кодирования и обработки графических объектов. Наиболее известные ГИС первого типа – программные продукты семейства *ArcInfo/ArcView* и *MapInfo*³. Программные продукты второго типа имеют совершенно другую архитектуру и представляют собой надстройки для универсальной системы автоматизированного проектирования *AutoCAD*. Эти системы, к которым в первую очередь относятся *Land Development Desktop* и *AutoCAD Map*, используют ядро и инструментальную среду *AutoCAD*. Функционально ГИС обоих типов схожи, однако имеется и ряд отличий. На инструментальном уровне системы на основе *AutoCAD*, используя всю мощь и разнообразие средств проек-

тирования родовой оболочки, обладают гораздо более развитым блоком координатной геометрии (*COGO engine*), предоставляющим отличный набор функций для сложного 2D- и 3D- моделирования, не заменимых при работе с археологической информацией, расширенным набором инструментов трехмерного пространственного анализа и моделирования. С другой стороны, *ArcView* располагает средствами статистического пространственного анализа геоданных, недоступными в *AutoCAD*-продуктах и, в целом, шире распространены, и предоставляет конечному пользователю более удобный и интуитивно-понятный интерфейс взаимодействия с информацией⁴.

В целом, вопрос выбора той или иной геоинформационной системы определяется множеством факторов. Скажем лишь, что если планируется использовать ГИС, например, для нанесения на карту раскопов или анализа распределения находок то стоит остановиться на *ArcView*. Для работы в полноценном трехмерном пространстве, трехмерного моделирования, построения моделей рельефа и т.д. лучше подойдет ГИС на основе *AutoCAD*. Существует и третий вариант, который мы использовали в нашей ИС – это совместное использование программных продуктов обоих типов. Такой путь создает дополнительные сложности на стадии проектирования системы, но и расширяет ее возможности.

"Контейнерами" для хранения геоинформационных данных системы служат *mxd*-файлы документов *ArcMap 8.2* и *dwg*-файлы рисунков *AutoCAD 2000i*.

В системе используется и другой "контейнер" для информации – реляционная база данных (БД), контролируемая с помощью системы управления реляционными базами данных (СУБД – RDBMS) *Microsoft SQL Server 2000*. Основу БД составляют таблицы (сущности), каждой записи которых соответствует набор значений полей-признаков (атрибутов). В зависимости от содержимого все таблицы нашей БД делятся на три типа: таблицы данных, таблицы метаданных и таблицы связей. В число таблиц данных в первую очередь входят таблицы индивидуальных находок на поселении и в могильнике, таблица раскопов и шурфов, таблицы раскопанных и нераскопанных курганов и другие. Таблицы метаданных содержат в себе исчерпывающую информацию о каждой единице источников, вошедших в состав ИС: фотографиях, документах, полевых отчетах, рисунках, картах, чертежах и т.д. Помимо этого таблицы мета-

³ ГИС *ArcInfo* – сложный многокомпонентный программный продукт корпоративного уровня, в то время как *ArcView* является упрощенной "настольной" версией системы с ограниченными возможностями. Мы не рассматриваем среду *ArcInfo*, поскольку ее использование для решения археологических задач в силу высокой стоимости системы и необходимости специального обучения маловероятно. Последней версией "настольной" системы является *ArcView 8.2*, получившая название *ArcMap 8.2*, однако чаще всего используются предыдущие версии *ArcView 3.1* и *3.2a*.

⁴ Функциональные возможности *ArcView* расширяются за счет подключения так называемых расширений – внешних модулей, таких как, например, *3D Analyst* – для трехмерного пространственного анализа и простого моделирования, *Spatial Analyst* – для расширенного пространственного анализа, а также модулей для импорта файлов *AutoCAD* или поддержки растровых изображений.

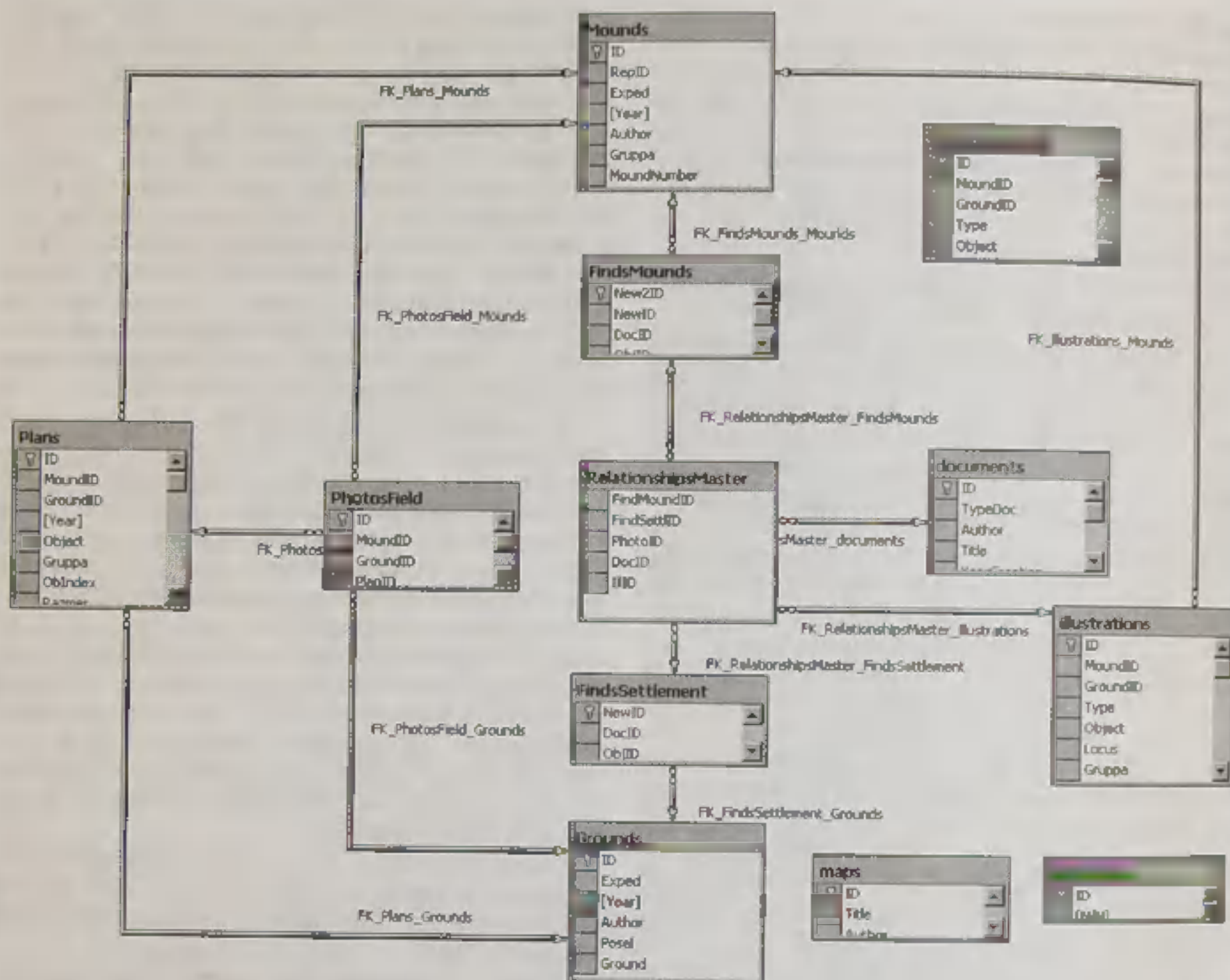


Рис. 2. Схема взаимосвязей таблиц в БД информационной системы.

данных содержат в себе ссылки на описываемые объекты. Таблицы связей имеют служебный характер и позволяют установить соответствия между данными и метаданными, хранящимися в БД. С их помощью, например, система «знает», что индивидуальным находкам № 5670–5770 соответствуют документы № 51, 89 и 123 (полевой отчет, дневник, диссертация), фотографии № 1234–1421, полевой чертеж № 2056, раскоп ЦГ-XX и т.д. В БД используются связи двух типов: один-ко-многим (например, перечню индивидуальных находок соответствует один раскоп, на котором они были обнаружены) и много-ко-многим (например, индивидуальной находке соответствует несколько фотографий, содержащих в себе другие индивидуальные находки). Связи между данными в таблицах выстраиваются с помощью ключей: первичных и внешних. Первичный ключ (*primary key*) уникально идентифицирует каждую запись в таблице, внешний ключ (*foreign key*) содержит значения первичного ключа в связываемой таб-

лице. *SQL server* оформляет отношения один-ко-многим напрямую, а для отношений много-ко-многим требует использования узловых таблиц, функцию которых в нашей системе выполняют таблицы связей (рис. 2).

При разработке таблицы для хранения паспортов индивидуальных находок было невозможно учесть все признаки описания для каждой категории или каждого типа предметов. Эта проблема решается за счет привязывания к базовой таблице развернутых БД на отдельные категории находок, например мечей, монет или торгового инвентаря, разработанных сотрудниками Смоленской археологической экспедиции в их исследовательской работе. В этом случае используется третий вид связей – один-к-одному, который позволяет однозначно сопоставить описания одного и того же предмета в разных таблицах.

Извлечение данных, хранящихся в БД, осуществляется с помощью синтаксиса языка запросов

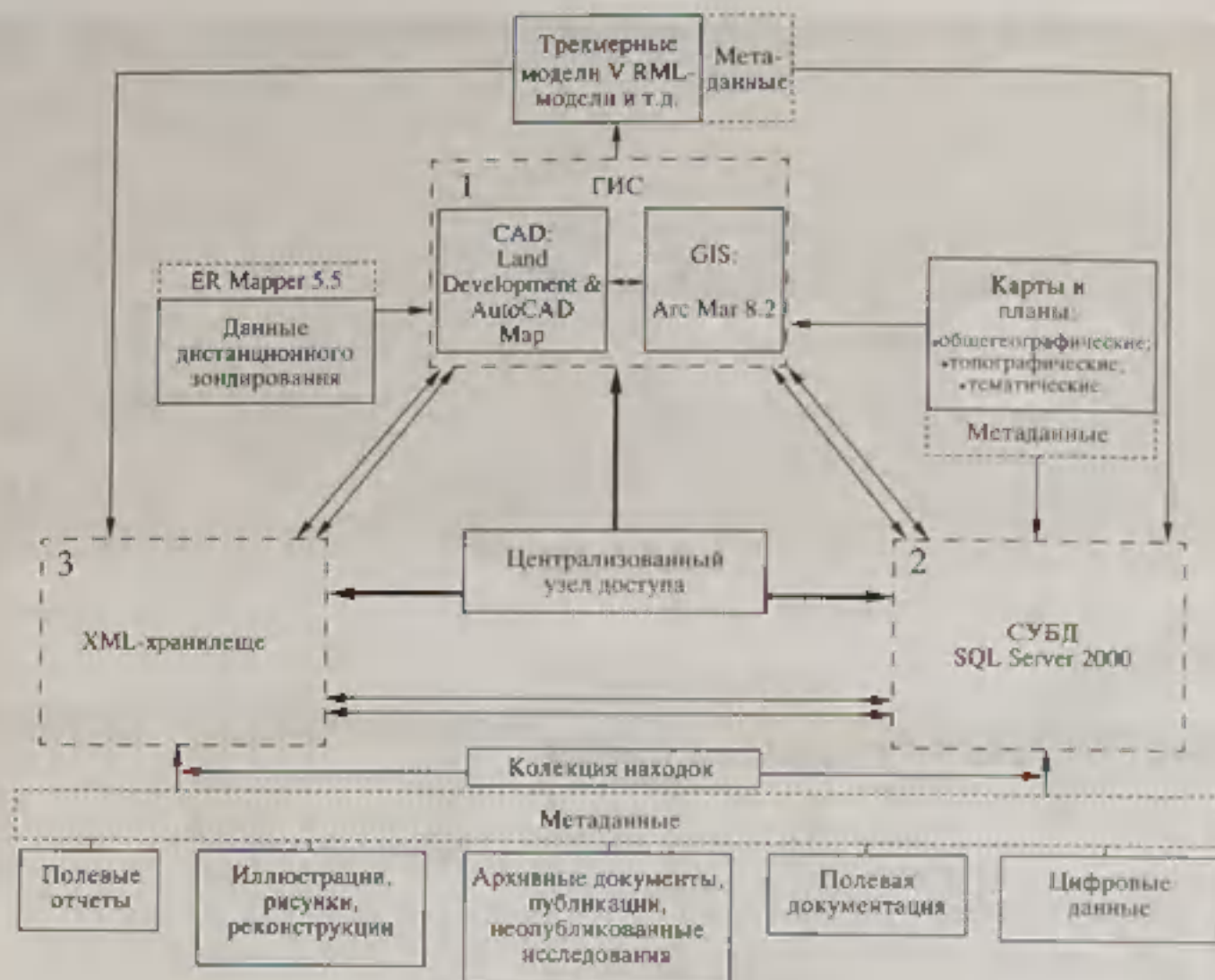


Рис. 4. Обобщенная схема информационной системы.

стью поток данных разбивается на структурные элементы, организованные в объектно-ориентированное иерархическое дерево. Объектно-ориентированное иерархическое структурирование информации гораздо лучше соответствует характеру большинства археологических источников и позволяет корректно организовать их внутри ИС. Структурная разбивка и создание объектной модели документа (DOM) позволяет системе обращаться к информации на уровне логических блоков, а не документа в целом. На практике это означает, что системе, например, может быть направлен запрос отобрать из всех полевых отчетов описания всех пластов, лежащих на одной глубине, независимо от их номеров или принадлежности раскопам и объединить в один сгенерированный XML-файл. Помимо текстовых документов, XML-структуры используются ИС и для организации растровых изображений.

Средства SQL Server 2000 позволяют связать информацию, хранящуюся в XML-документах и таблицах БД. Поток XML трансформируется в реляционные таблицы с помощью OPENXML – расширения SQL Server 2000 для языка запросов Transact-SQL. Созданные таблицы динамически включаются в массив БД. По такой схеме, например, осуществляется подключение фотографий индивидуальных находок к таблице их паспортов. Обратная трансформация данных также возможна. В этом случае таблицы и представления SQL Server 2000 преобразуются в XML-поток либо с по-

мощью параметра FOR XML оператора SELECT языка Transact-SQL, либо с помощью запроса, записанного в HTTP-заголовок обращения к SQL Server. В разработанной ИС в основном используется второй способ извлечения табличных данных, причем типовые запросы хранятся в специальных шаблонах. Шаблоны позволяют не только упростить использование типовых запросов, но и объединять несколько запросов вместе.

В большинстве случаев системе требуется получить XML-документ не целиком, а лишь некоторую его часть. Для этих целей используется технология Xpath 1.0, с помощью которой формируется результирующий XML-поток.

XML является не единственным средством организации текстовых и графических данных в разработанной ИС. Все архивные документы, отчеты, дневники и т.д. хранятся также в PDF (Portable Document Format) файлах. Использование PDF позволило нам включить отсканированные текстовые материалы в ИС еще до того, как они были распознаны и преобразованы в машиночитаемый текст.

Обобщенная структура ИС представлена на рис. 4. Ее ядро составляют три основных компонента, рассмотренных выше: ГИС (1), СУБД SQL Server 2000 (2) и XML-хранилище (3). Компонент ГИС образован двумя модулями: CAD-модулем, включающим в себя AutoCAD 2000i, CAD Overlay 2000i, AutoCAD Land Development Desktop R2i и

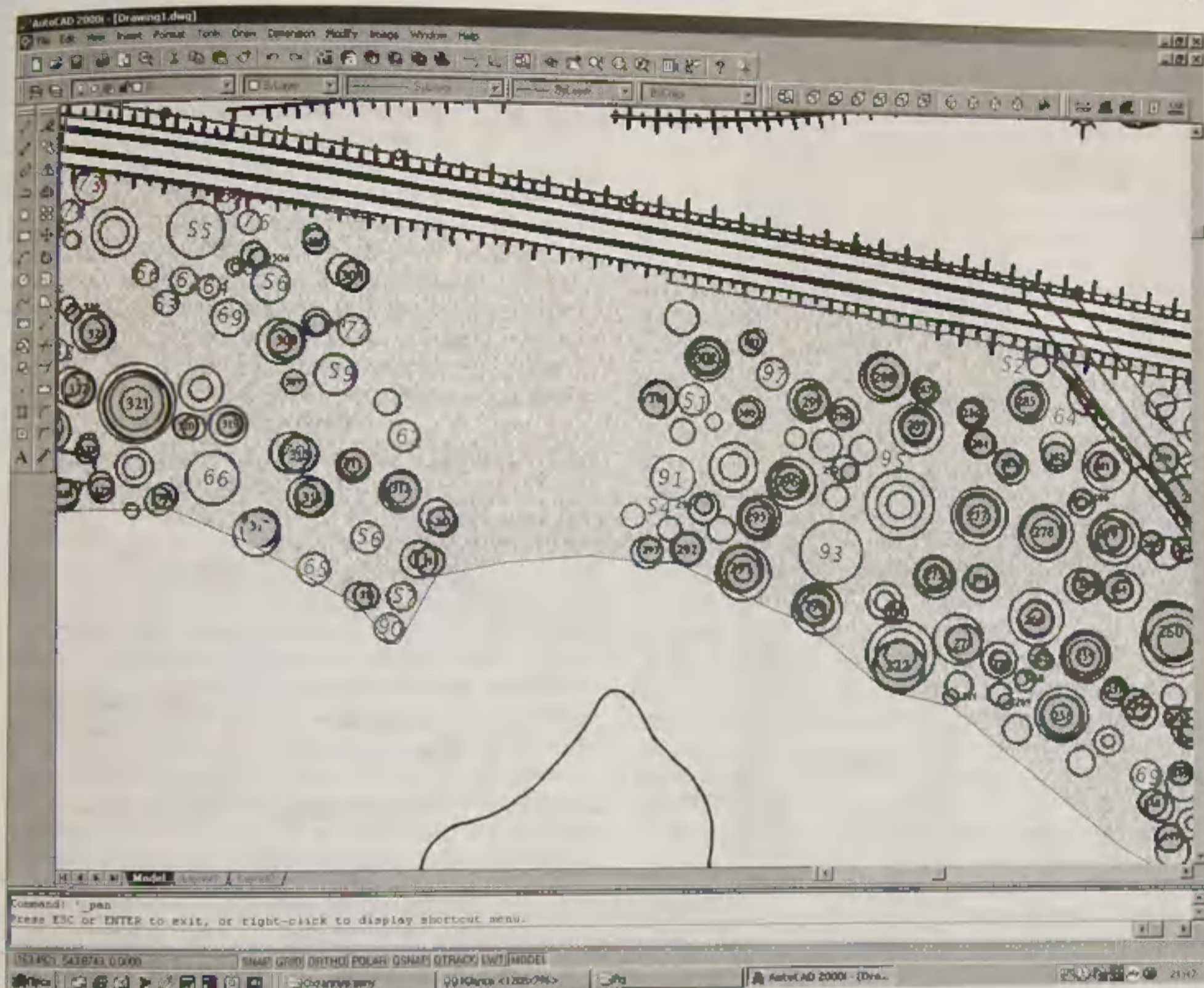


Рис. 5. CAD-модуль ГИС-компонента отображает планы Центральной курганной группы, снятые в 1953 г. (пустые круги) и 1976 г. (залитые круги). Хорошо видна динамика изменения состояния памятника.

AutoCAD Map 3, и *GIS*-модулем, содержащим *ESRI ArcMap 8.2*. Все компоненты взаимосвязаны и могут взаимодействовать, обмениваясь данными, формируемыми по внутрисистемным запросам. Принцип совместной работы *XML*-хранилища и *SQL Server 2000* уже обсуждался, поэтому остановимся на технологии интеграции компонента ГИС. Одним из важнейших функциональных средств программных продуктов как *CAD*-, так и *GIS*-модулей, является возможность подключения внешних БД и связывания графических данных с данными, содержащимися во внешних таблицах. Таким образом, подключенные данные *SQL Server 2000* или трансформированный в реляционные таблицы поток *XML* связываются и обрабатываются вместе с графической информацией. В обратном направлении какой-либо из компонентов или сама система осуществляют доступ к данным и объектам ГИС с использованием *ADO 2.6*

(*ActiveX Data Objects*), входящего в состав *MDAC 2.6* (*Microsoft Data Access Components*). Примеры работы ГИС-компонента приведены на рис. 5, 6.

При описании архитектуры и технологической базы разработанной ИС мы нередко использовали более короткий термин – система. Под системой подразумевается четвертый из основных компонентов – узел централизованного доступа к данным. Архитектура ИС реализована в топологии клиент/сервер. Основу централизованного узла доступа составляют связанные динамические *HTML*-страницы, созданные с помощью технологии *ASP 3.0* (*Active Server Pages*) и объединенные на одной стартовой странице, запускаемой в любом современном *Web*-браузере. Для того чтобы сервер мог обрабатывать *ASP*-страницы, на нем должен быть установлен *IIS 5.0* (*Internet Information Services*). *ASP*-страницы хранятся на сервере, и при обращении к ним клиента (пользователя, открывшего

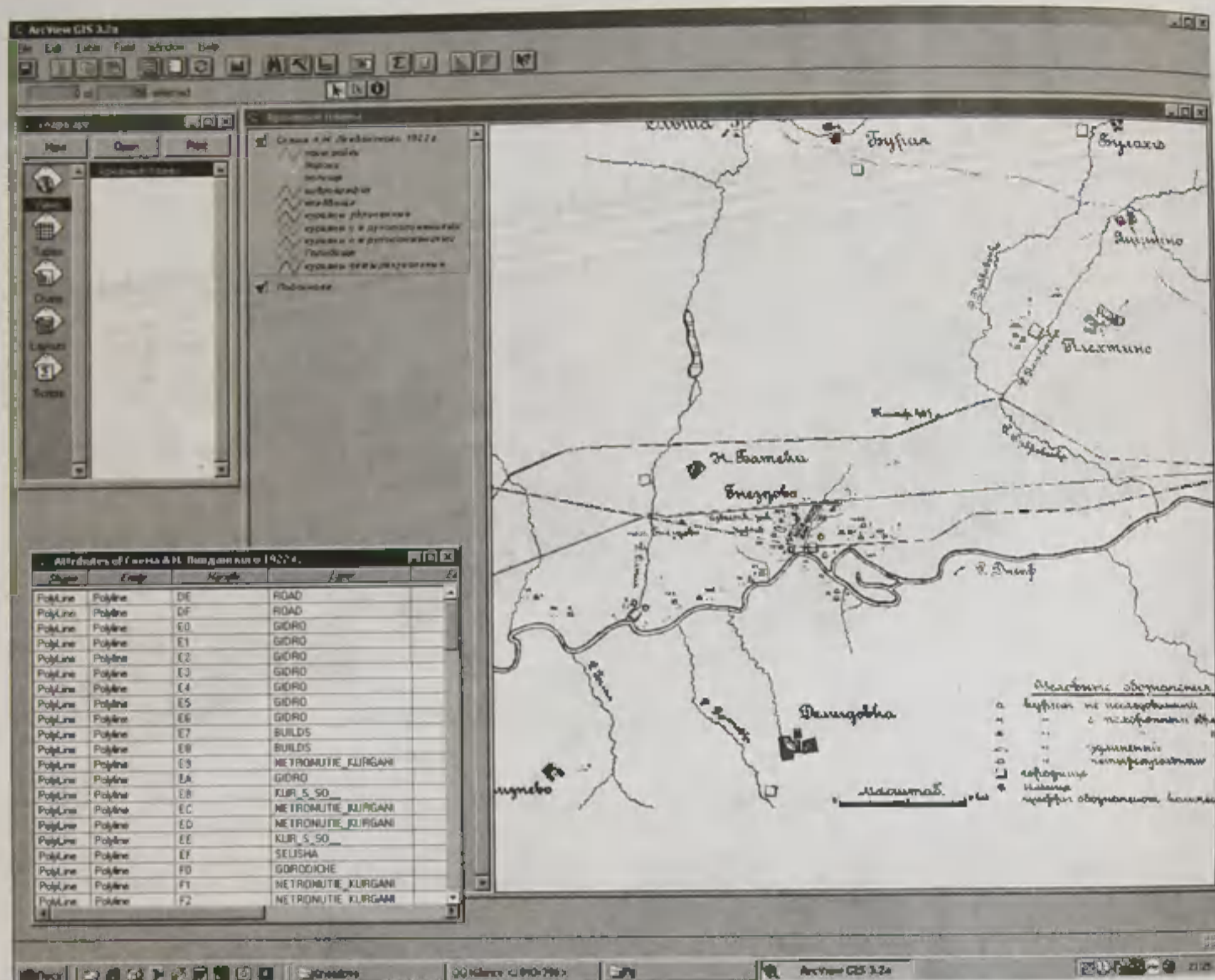


Рис. 6. Фрагмент карты Гнездо, снятой А.Н. Лявданским в 1922 г., отображаемый в GIS-модуле системы.

стартовую страницу, например в *Internet Explorer*, и задавшего некоторые параметры) формируют HTML-код в соответствии с запросом, передаваемый обратно клиенту и отображаемый в его браузере. На практике это означает, что доступ к ресурсам ИС не требует установки никакого специального программного обеспечения, за исключением стандартного Web-браузера, и не налагает ограничений на аппаратную или системную конфигурацию компьютера пользователя. Например, если пользователю требуется отобрать находки, происходящие из одного раскопа, и просмотреть их фотографии, то он задает параметры запроса на стартовой странице централизованного узла доступа. Эти параметры передаются серверу, который сам обращается к БД- и XML-хранилищу, производит необходимые операции по извлечению и обработке информации, затем преобразует данные в HTML-поток и отправляет его пользователю. Браузер пользователя отображает полученные

данные. Тем самым пользователю абсолютно неважно, в каком формате хранятся данные на сервере и как с ними работать. Просмотр картографических данных требует дополнительной несложной установки специальных объектов просмотра в браузер (*ArcExplorer* или *ArcReader* для GIS-модуля, *VoloView* для CAD-модуля, *CostoPlayer* или любой аналог – для VRML-контента).

Обеспечение функционирования ИС и взаимодействия ее компонентов потребовало большого объема программирования. Для создания ASP-страниц и централизованного узла доступа к данным использовался *Visual Basic Script (VBScript)*, для программирования работы *SQL Server 2000* – язык запросов *Transact-SQL* и *Visual Basic 6.0 for Applications (VBA)*, для работы с XML-хранилищем – язык разметки XML, *Xpath* и *VBA*, для работы с объектами *ADO* и ГИС компонентом – *VBA*.

4. *Оцифровка и кодирование информации.* Это наиболее трудоемкий этап проекта, на котором разработанная структура ИС наполняется данными. Карты и схемы Гнёздовского археологического комплекса представлены листами разного формата от фотокопий размером А4 до ватманских листов длиной более 3 м. Все они были отсканированы с цветностью, соответствующей источнику с разрешением 400 и 600 dpi. В ряде случаев при наличии нескольких копий оригинала в качестве источника для сканирования были использованы профессиональные фотокопии. Увеличение разрешения при сканировании позволяет "растянуть" полученные изображения до размеров оригинала без существенной потери качества.

Полевые чертежи составляют огромный массив данных и представлены в подавляющем числе случаев листами большого формата А1-А0. Все они были отсканированы в формате 8-bit grey (256 оттенков серого) с разрешением 400 dpi. 8-битное кодирование цвета при сканировании полевых чертежей обычно не используется, так как на порядок увеличивает объем отсканированных изображений. Однако это единственное приемлемое решение для чертежей, выполненных на миллиметровках, а для чертежей на кальке или ватманских листах позволяет существенно увеличить ресурс их последующей компьютерной обработки. При наличии нескольких копий чертежа предпочтение отдавалось перебеленным экземплярам.

Следующим после сканирования этапом обработки чертежей, карт и схем является коррекция растрового изображения: "чистка", "выравнивание", "кадрирование". Немаловажным аспектом, достойным упоминания, является коррекция размеров изображения. При сканировании, вне зависимости от модели сканера происходит искажение реальных размеров оригинала в среднем на 0,3–0,5%. Таким образом, при размере чертежа 100 × 100 см искажение в среднем составляет 0,3–0,5 см, или в масштабе полевого чертежа 1 : 20 – 6–10 см. Очевидно, что при масштабе карты, например 1 : 5000, ошибка может достигать 25 м и более. Исправление искажений производится с помощью контрольных точек, реальное расстояние между которыми достоверно известно.

Печатные документы, в первую очередь полевые отчеты, сканировались и распознавались в OCR *FineReader* 6.0. Начиная с пятой версии этот программный продукт позволяет достичь вполне приемлемого качества распознавания напечатанного на пишущей машинке текста, за исключением текста, отпечатанного под копирку. Текст сканировался как черно-белое изображение с разрешением 300 dpi с автоматическим подбором яркости средствами самой программы. Листы альбомов, входящих в состав полевых отчетов, сканировались как обычные фотографии с фиксирован-

ном размером, соответствующим размеру листа, и добавлялись впоследствии в "пакет" *FineReader*.

Целый массив архивных документов, в первую очередь дневников раскопок, существует лишь в рукописном виде и не может быть преобразован в электронный текст. Такие материалы сканировались постранично как изображения с параметрами цветности в зависимости от оригинала и разрешением 400 dpi, после чего полученные изображения страниц собирались в универсальном формате *PDF* и помещались в *XML*-хранилище ИС.

Прочие иллюстративные материалы, в подавляющем большинстве случаев черно-белые, были отсканированы со стандартным разрешением 400 dpi с 8-битным кодированием оттенков серого цвета. Среди нюансов отметим, что в ряде случаев сканирование черно-белых фотографий как цветных изображений (24-битное *RGB* кодирование цвета) обеспечивает более корректную цветопередачу, что немаловажно для полевых снимков. Приемлемые результаты также дает применение фильтра резкости на стадии сканирования.

Важнейшей составляющей стадии ввода данных в ИС стало наполнение разработанных таблиц БД информацией. Особое внимание было уделено созданию максимально удобного и упрощенного интерфейса ввода данных в БД посредством формы.

В нашем случае основной принцип функционирования формы БД основан на системе контекстно-зависимых выпадающих списков, позволяющих не впечатывать, а выбирать нужное значение из предлагаемого списка, формируемого "на лету". Например, для БД на индивидуальные находки группа макросов управляет динамическим изменением списка предлагаемых значений в зависимости от типа предмета, его названия, сохранности, функциональной группы, к которой он принадлежит, и большого числа других факторов. Для каждого поля определены жесткие условия на вводимое значение (например, запрет отрицательных или дробных значений для поля "год раскопок" и заданный интервал от 1867 до 2003 г. – время изучения Гнёздова). Тем самым сводится к минимуму процент опечаток и ошибок, а введенная информация делается легко проверяемой.

После заполнения общего паспорта вещи пользователь вводит название одной из функциональных групп предметов. В зависимости от выбранной функциональной группы формируется список названий входящих в нее материальных находок. После выбора названия предмета система позволяет ввести описание вещи, формализованное по ряду полей (таких как, материал, цвет, конструктивные части и пр.), причем доступными являются лишь те поля, что нуждаются в заполнении для каждого конкретного предмета. Список вариантов значений, предлагаемый для каждого из полей,

строга зависит от выбранного названия находки. Например, **Функциональная группа** -> Предметы вооружения -> **Название** -> Наконечник стрелы -> **Материал** -> Железо и т.д. Создание такой системы ввода данных потребовало проведения специальной работы по созданию полного списка возможных значений параметров описания для каждого наименования материальных находок и постепенного внесения в него корректив в течение работы над проектом. Помимо этого, с помощью специально созданных пользовательских кнопок пользователь имеет возможность для ускорения и удобства ввода установить ряд значений полей значениями по умолчанию, которые автоматически заполняются для каждой новой записи, или сдублировать запись, стереть или восстановить ее. Работа автоматизированных форм ввода для всех БД ИС – курганов, картографического материала, полевых чертежей, всех типов документов (полевые дневники, отчеты, публикации и т.д.), цифровых и архивных фотографий, цифровых фотопанорам – обслуживается выполнением около 100 макросов и программ на языке VBA.

В специально разработанную БД были внесены данные проведенной в 2002 г. инвентаризации всех курганных групп Гнёздовского могильника.

К данной стадии выполнения проекта относится также пополнение существующей источниковой базы цифровыми фотографиями материальных находок из Гнёздовской коллекции. Произведена сплошная съемка всех предметов из хранения Государственного исторического музея и кафедры археологии МГУ им. М.В. Ломоносова, содержащих в общей сложности около 84% единиц хранения всей коллекции. Для съемки использовался цифровой фотоаппарат *Olympus 2500L* с зеркальным объективом. Выдающиеся находки были сняты с использованием технологии создания трехмерных фотографий. Эта технология аналогична технологии создания круговых фотопанорам с тем исключением, что в данном случае объект помещается на вращающуюся подставку и фотографируется с 12 различных ракурсов. Компьютерная обработка фотографий позволяет создать псевдо-трехмерное изображение предмета.

5. *Обработка информации, ее унификация и интеграция в ИС.* Отсканированные и откорректированные карты и схемы были переведены в векторный формат. Векторизация растровых изображений карт, схем и чертежей производилась в ручном режиме с использованием программных продуктов *Rx Spotlight 97* компании *Consistent Software* и *CAD Overlay 2000* компании *AutoDesk*. На настоящий момент векторизован весь картографический материал и часть полевых чертежей.

Оцифрованные изображения были интегрированы в ИС. Первым этапом стало цветовое и послойное структурирование информации (выделе-

ние дискретных групп однотипных объектов, таких как изолинии, высотные отметки, элементы гидрографии, дороги, строения, линии ЛЭП и т.д. – для карт и планов и слои, прослойки, пятна, остатки сооружений, камни и т.д. – для полевых чертежей). Затем были унифицированы система условных обозначений, масштаб, единицы измерения, ориентация и прочие формальные признаки. На последнем этапе полученная однородная информация была организована в едином трехмерном пространстве ГИС компонента (*CAD-модуль*) ИС с четким соблюдением относительного масштаба, единой системы цветового структурирования и единым принципом наименования и организации слоев. Картографическая составляющая ИС прошла дополнительную обработку, в ходе которой была выполнена уточняющая ректификация различных элементов многослойной растрово-векторной подосновы памятника на основе листов аэрофотосъемки. В качестве опорных точек использовались объекты гидрографии, дороги, железная дорога, узлы триангуляционной сети, характерные элементы рельефа.

Дифференцированные графические объекты были связаны с соответствующими им записями БД. Например, значки, обозначающие курганные насыпи на плане могильников, соотнесены с описаниями в базе или точки съемки круговых панорам привязаны к соответствующим записям. Помимо подобных “простых” связей организованы и более “сложные”: к примеру, все полевые фотографии, относящиеся к определенному пласту раскопа, сопоставлены с его векторным чертежом, хранящимся в ГИС-компоненте системы.

Следующим этапом обработки картографических материалов стало построение трехмерных моделей рельефа (*3D surfaces*) для территории Гнёздовского комплекса археологических памятников (рис. 7). Трехмерные модели предоставляют широчайшие возможности для визуального и пространственного анализа топографии археологических памятников, создания трехмерных реконструкций и презентаций. К числу аналитических средств относятся: построение изолиний и высотных отметок любой плотности, моделирование геологических и стратиграфических разрезов и профилей любой конфигурации, вычисление площади, размеров и объема любых объектов, расчет высот и склонов, построение моделей водосбора и водостока, виртуальное нивелирование и сглаживание поверхностей, моделирование угла падения теней и затененности объектов (уникальный инструмент виртуального выявления неразличимых на местности археологических объектов, например, распаханных курганов) (Zhukovsky, 2001).

После обработки в среде *3D Studio MAX* фирмы *Kinetix* трехмерные модели становятся доступными в качестве *VRML*-контента. *3D Studio* позволя-



Рис. 7. Трехмерная модель рельефа части памятника.

ет определить местоположение источников света, точек наблюдения, наложить на трехмерную поверхность любые растровые текстуры, в том числе изображения карт и аэрофотосъемки. Наблюдатель имеет доступ к VRML-модели как в активном режиме свободной навигации в трехмерном пространстве (облет, "пешая прогулка" и т.д.), так и пассивно – просматривая анимационные ролики (рис. 8). Теоретические вопросы, связанные с проблемой восприятия VRML-моделей, а также возможностями их визуального анализа, изложены в статье одного из авторов (Zhukovsky, 2002).

Дополнительная обработка потребовалась для данных, хранящихся в БД насыпей курганных могильников Гнёздовского археологического комплекса. При вводе в БД параметры паспортов были формализованы, что крайне удобно для их хранения, обработки и статистического анализа, но малоприспособно, если требуется связное текстовое

описание каждого кургана. Задача была решена с помощью специально созданной программы на VBA, автоматически генерирующей связный текст в формате *Microsoft Word* из формализованных значений полей БД.

Цифровые фотографии индивидуальных находок после небольшой обработки были сопоставлены с записями БД Гнёздовской коллекции. Все введенные тексты обработаны и помещены в XML-хранилище.

6. Организация доступа к ресурсам ИС. Информационная система построена в топологии клиент/сервер. В стандартном случае сервером является удаленный компьютер, доступ к которому осуществляется через локальную сеть или сеть Интернет. Аппаратные требования к серверу определяются его загрузкой, т.е. количеством пользователей и обращений к данным, и зависят от каждой конкретной ситуации. Программная часть



Рис. 8. Интерактивная VRML-модель рельефа части памятника.

сервера более определена: он должен работать под управлением *Microsoft Windows 2000 Server* и содержать в себе *IIS 5.0 (Internet Information Services)* и *SQL Server 2000* с поддержкой XML. Клиентский компьютер пользователя должен иметь установленную операционную систему семейства *Microsoft Windows* и стандартный браузер (например, *Internet Explorer 5* и выше или *Mozilla/4.0* и выше).

Однако и клиентская, и серверная части системы могут располагаться и на одном компьютере с параметрами сервера. Такая архитектура используется для доступа к ресурсам ИС в настоящий момент, пока производится ее тестирование.

7. *Тестирование целостности и оптимальности структуры ИС, проверка работы всех ее компонентов, выявление ошибок.* Тестирование ИС проводилось на двух уровнях: уровне связей и взаимодействия компонентов и уровне данных. Архитектура и работоспособность системы проверялась по следующим параметрам: корректная работа отдельных компонентов, корректное взаимодействие компонентов, ввод контрольных значений и соответствие результатов проектной задаче, устойчивость системы к вводу некорректных значений или некорректным действиям пользователя, работоспособность клиентской части системы в различных операционных системах семейства *Windows*.

Значительная часть тестирования проводилась уже на этапе ввода данных, что позволило опера-

тивно выявлять и устранять ошибки в структуре системы и баз данных.

Целостность данных проверялась вручную на наличие систематических и случайных ошибок в БД (опечатки, неточная атрибуция, занесение значений параметров в несоответствующие поля и т.д.), пропусков в оцифрованных материалах (пропущенные страницы или отдельные чертежи и т.д.), некачественное сканирование и аналогичные ошибки.

8. *Внесение исправлений, устранение выявленных ошибок, оптимизация и финальное тестирование.* Работы на этой стадии продолжаются в настоящий момент.

Несмотря на это, уже сейчас функциональная мощь ИС используется авторами для решения прикладных и исследовательских проблем изучения Гнёздовского комплекса археологических памятников.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект № 01-07-90076).

СПИСОК ЛИТЕРАТУРЫ

- Авдусин Д.А. Краткая история полевых исследований // Гнёздовский могильник. Археологические раскопки 1874–1901 гг. (по материалам ГИМ). Ч. 1. Тр. ГИМ. Памятники культуры. Вып. XXXVI. М., 1999.

Археологическая карта России (АКР). Смоленская область. Ч. 1. М., 1997.

Базы данных в археологии. М., 1995.

Ениосова Н.В., Митоян Р.А., Сарачева Т.Г. Информационная система "Цветные металлы и их сплавы на территории Восточной Европы в эпоху средневековья" // ВМУ. Сер. 8. История. 2003. № 1.

Матехин К.А., Богуславский О.И. Базы данных для археолога (современное состояние и возможности решения) // Памятники старины. Концепции. Открытия. Версии. Памяти В.Д. Белецкого (1919–1997). Т. II. СПб.; Псков, 1997.

Методы естественных наук в археологических реконструкциях. Новосибирск, 1995.

Мурашева В.В., Нефедов В.С., Пушкина Т.А. Изучение Гнёздова и его округи (Смоленская область, 1999–2000 гг.) // Вестн. РГНФ. 2001. № 1.

Zhukovsky M.O. Virtual 3-D Reconstruction of the Kiafar Site, North Caucasus, Russia, Computing Archaeology for Understanding the Past // CAA 2000. Proceedings of the 28th Conference. Ljubljana, 2001.

Zhukovsky M.O. Handling Digital 3-D Record of Archaeological Data in Situ, Archaeological Informatics – Pushing the Envelope // CAA 2001. Proceedings of the 29th Conference. Visby. 2002.

Digital archive of the materials from the Gnezdovo archaeological complex

T. A. Pushkina, M. O. Zhukovsky

Summary

The article discusses the project worked out the Chair of archaeology, History faculty, Lomonosov State University, Moscow. The work was aimed at creation of unified multifunctional informational system for compiling the whole corps of scientific information concerning the Gnezdovo archaeological complex. The system must provide reliable data preservation, free access, analytical processing, presentation and distribution of the information with regard to scientific and educational programs. The informational system includes textual descriptions, excavation reports, drawings of excavation trenches (layer plans and stratigraphic columns), field records, register books, photos and drawings of finds, archaeological and topographic layouts and schemes executed in different scales, geographic and topographic maps, archive maps and drawings, the data of land surface remote testing, collection of material finds, and other data accumulated by the excavators during 130-year long period of investigation conducted at Gnezdovo. In the paper all stages of the project accomplishment are considered in detail, from formulation of the task to organisation of the system resource access, the authors specify the used technologies and the technical details of their application.