

Интернет-журнал «Отходы и ресурсы» <https://resources.today>
Russian Journal of Resources, Conservation and Recycling

2026, Том 13, № 2 / 2026, Vol. 13, Iss. 2 <https://resources.today/issue-2-2026.html>

URL статьи: <https://resources.today/PDF/15INOR226.pdf>

DOI: 10.15862/15INOR226 (<https://doi.org/10.15862/15INOR226>)

2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки)

Ссылка для цитирования этой статьи:

Диомидов, И. А. Разработка интегрированной системы поддержки инвентаризации лесных ресурсов на основе учета происхождения данных в гетерогенных информационных средах / И. А. Диомидов // Отходы и ресурсы. — 2026. — Т. 13. — № 2. — URL: <https://resources.today/PDF/15INOR226.pdf>. DOI: 10.15862/15INOR226.

For citation:

Diomidov I.A. Developing an integrated support system for forest resource inventory based on data provenance in heterogeneous information environments. *Russian Journal of Resources, Conservation and Recycling*. 2026;13(2): 15INOR226. Available at: <https://resources.today/PDF/15INOR226.pdf>. DOI: 10.15862/15INOR226. (In Russ., abstract in Eng.).

УДК 004.051

Диомидов Иван Андреевич

ФГБОУ ВО «Иркутский государственный университет путей сообщения», Иркутск, Россия
Аспирант

ФГБУ «Рослесинфорг», Иркутск, Россия
Прибайкалеспроект

Ведущий-инженер программист, аспирант
E-mail: pilad0hwts@yandex.ru

РИНЦ: https://elibrary.ru/author_profile.asp?id=1220889

Разработка интегрированной системы поддержки инвентаризации лесных ресурсов на основе учета происхождения данных в гетерогенных информационных средах

Аннотация. Применение методов учета происхождения является актуальным направлением в лесном хозяйстве, в связи с изменчивостью и исторической гетерогенностью лесоустроительных материалов. Целью данного исследования является разработка системы поддержки инвентаризации лесных ресурсов с целью обеспечения воспроизводимости результатов, прослеживаемости преобразований и учета происхождения сведений. Автором дается характеристика проблеме разработки комбинированных систем интеграции для лесного хозяйства и кратко рассматривается вопрос применения методов учета происхождения. В статье рассматриваются взаимодействие ключевых компонентов системы, применяемая в системе логическая модель учета происхождения и основные сценарии работы. На ее основе дается характеристика основным свойствам и функциям программной реализации: системы учета происхождения Casuga и предметно-ориентированное расширения для лесного хозяйства Лес-Sense. В статье дается численная оценка ее свойствам на основе разработанной автором методики путем формирования экспертной многокритериальной оценки системы учета происхождения применительно к специфике лесного хозяйства. Рассмотренная система обеспечивает достаточно высокий уровень работы с информационно гетерогенными данными, совмещая свойства геоинформационных систем с функциями учета происхождения данных. Наиболее высокие оценки получены для задач, где важны скорость доступа, навигация по объектам, отображение подготовленных представлений.

Исследование вносит вклад в развитие подходов в части учета происхождения и интеграции гетерогенных данных и может быть полезным для специалистов и исследователей в области системного анализа и лесного хозяйства.

Ключевые слова: системный анализ; учет происхождения данных; лесное хозяйство; информационные системы; геоинформационные системы; цифровизация; система

Введение

Цифровизация, применительно к задаче совершенствованию государственного лесного реестра, рассматривается преимущественно как переход от бумажных носителей к электронным формам хранения и представления данных. Такой переход действительно расширяет доступ к информации и создает основу для развития профильных лесных систем. Однако сама по себе электронная форма не устраняет противоречия, накопленные в технологиях сбора, хранения и использования лесоустроительных данных. Напротив, сейчас цифровизация делает более заметной неоднородность источников, форматов, классификаторов и процедур обработки сведений, применявшихся в лесном хозяйстве на протяжении длительного времени [1].

В действующем нормативном контуре лесоустройство определяется как система взаимосвязанных мероприятий, направленных на получение актуальных и достоверных сведений о лесах и лесных ресурсах, используемых в том числе для ведения государственного лесного реестра, лесного планирования, оценки лесов и обеспечения устойчивого управления лесами. При этом лесоустроительная документация [2] должна формироваться в электронной форме, включая машиночитаемые форматы, что переводит проблему из области простого хранения документов в область согласования данных, моделей и процедур их применения.

В исследованиях авторы часто указывают на проблемы лесного хозяйства связывая их с низкой актуальностью материалов лесоустройства, пробелами правового регулирования, недостатками планирования и управления лесами [3; 4]. Другое направление связано с развитием принципиально иных методов таксации и учета лесов, позволяющее увеличить стандартизацию технологических процессов. В частности, это касается увеличение роли материалов дистанционного зондирования Земли [5–7]. Рассматриваемые в данных исследованиях причины существенны, но они не исчерпывают проблему. Более глубокое ограничение связано с разнообразием форм представления сведений о лесах и различием задач, для которых эти сведения используются: таксация, планирование, предоставление участков в пользование, лесовосстановление, контроль изменений, оценка состояния лесных ресурсов. Поэтому потребность в использовании разнородных средств хранения и обработки данных следует рассматривать как системное ограничение, требующее формализации связей между источниками данных, процедурами обработки и управленческими решениями.

С этой позиции актуальным становится разработка нового для лесного хозяйства класса информационно-аналитических систем, ориентированных не только на учет лесов как природных объектов, но и на учет сведений о лесах; формирование из них самостоятельного объекта управления. В таких системах ключевое значение приобретает не простое накопление данных, а фиксация их происхождения, условий получения, преобразований, связей с нормативными основаниями и последующего использования в управленческих процедурах.

Цель настоящего исследования — показать возможность применения систем учета происхождения данных в задачах учета сведений о лесах на примере разработанной информационной системы, описать опыт ее использования и дать оценку полученным результатам.

Характеристика предметной области. Постановка проблемы

Согласно лесоустроительной инструкции [2] последовательность выполнения работ при таксации лесов включает: подготовительный этап, полевой этап, камеральный этап. На практике формирование актуального состояния базы (рис. 1) сопровождается поступлением новых источников, уточнением границ и атрибутивных характеристик, пересчётами показателей, выпуском документов и последующей проверкой результатов. Контроль хода лесоустроительных работ и уточнение состояния данных нередко осуществляются с опорой на ручные организационные средства, не образующие единого контура прослеживаемости. К числу наиболее распространённых относятся: ведение табличных реестров, подготовка текстовых журналов и заметок, бумажные полевые записи, обсуждение спорных случаев в мессенджерах и электронной почте, устные консультации со специалистами, рабочие звонки и совещания, а также ручная сверка картографических и таксационных материалов. Указанные способы позволяют решать локальные задачи координации и оперативного уточнения сведений, однако их применение носит фрагментарный характер и, как правило, не обеспечивает строгой фиксации оснований внесённых изменений.

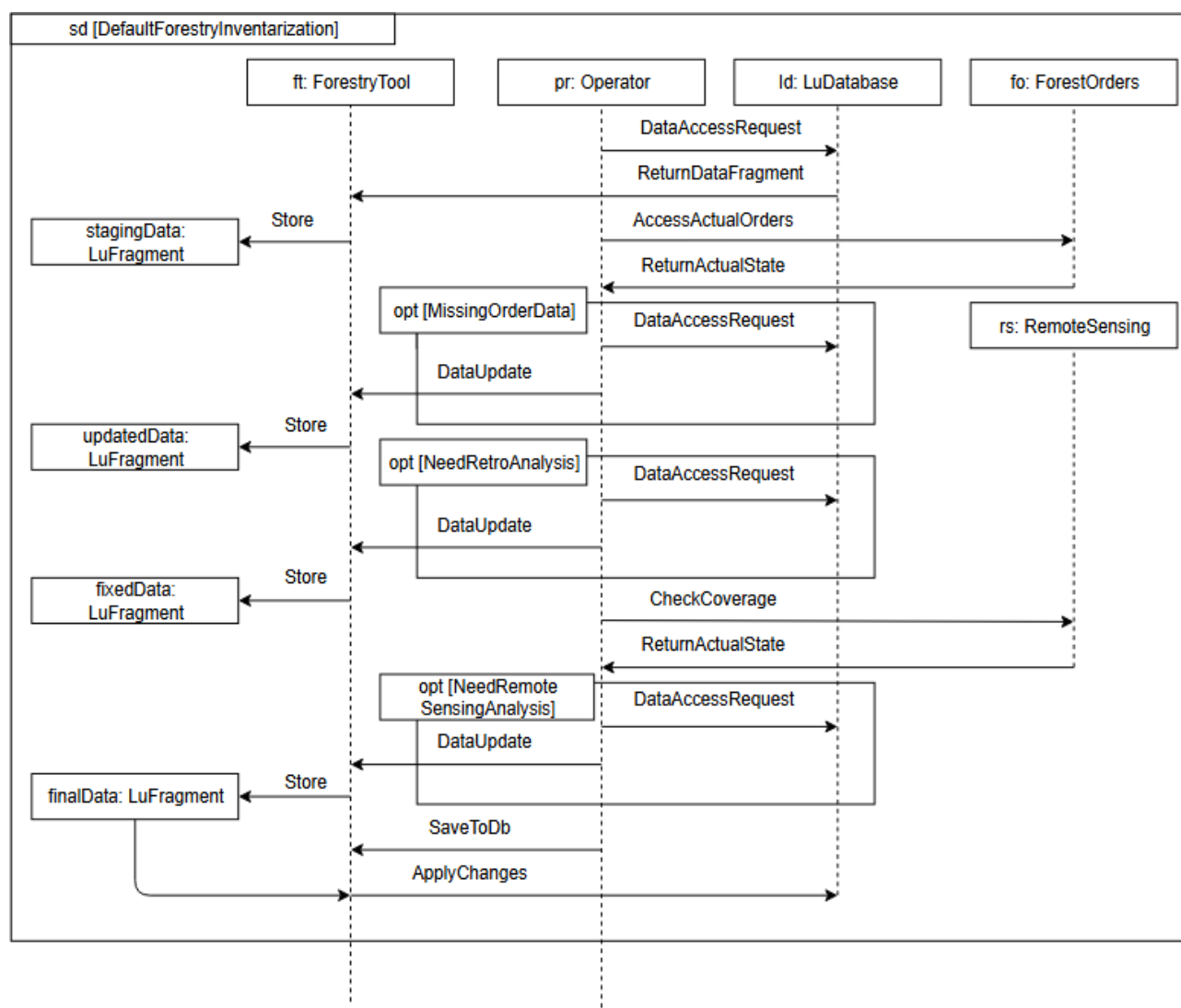


Рисунок 1. Процесс инвентаризации лесных ресурсов
с ручным учетом происхождения (разработано автором)

Один и тот же объект данных проходит через несколько промежуточных состояний, когда внешнему потребителю обычно виден в лучшем случае «вход/выход», в худшем —

только «выход». Всё это остаётся в голове специалиста и частично в разрозненных следах (запросы, файлы, версии). При высокой изменчивости данных и множестве источников это даёт две типовые поломки: результат трудно воспроизвести, доказать и повторить (невозможно однозначно восстановить ход работ), возрастает риск ошибок интеграции и неверной идентификации (перепутанные версии, неучтённые допущения).

Подробно проблематика развития информационных систем в контексте различных задач рассмотрена автором в исследованиях [8; 9]. Как правило, проблема учета происхождения в действующих системах имеет вторичную роль, и может быть, например, учтена в форме создания резервных и рабочих копий материалов. Автор видит дальнейшее развитие систем для лесного хозяйства в создании комбинированных систем, способных осуществлять контроль качества данных, комбинацию ретроспективных и актуальных методов, ведение банков ретроспективных данных, визуализации и способность работать с данными в мультиструктурном формате. Для решения такой задачи была разработана формальная модель на основе интегрированных языковых запросов, рассматривающих процесс интеграции как абстрактные зависимости между наборами данных D и схемами Σ . Предложенная в работе формальная модель позволила рассмотреть проблему интеграции сведений прежде всего, как задачу обеспечения согласованности между типами данных.

Настоящее исследование является логическим продолжением, в связи с чем рассматриваемая в работе [9] нотация и соглашения принимаются и в данной работе. Если подразумевать целью инвентаризации лесов некоторую интегрированную базу данных, то данный процесс может быть записан таким образом $\langle DS^L, Map(DS^L, DS^I), DS^I \rangle$. Мы наблюдаем только два состояния набора данных — до и после шага обработки и фиксируем факт перехода между ними, допускающий существование некоторого преобразования:

Тогда, если опереться на операции или их последовательности, можно представить процесс интеграции пошагово, где каждый шаг определяется входным и выходным типом. Представим конкретный шаг (модификацию данных) в виде некоторого преобразования *Action*.

Для этого обозначим: мы наблюдаем только два состояния набора данных — до и после шага обработки, когда *Action* это функция перехода между ними, допускающий существование некоторого преобразования:

$$\begin{aligned} \langle D_i^t, \Sigma_i^t \rangle, \quad \langle D_j^{t+1}, \Sigma_j^{t+1} \rangle, \\ \langle D_j^{t+1}, \Sigma_j^{t+1} \rangle = Action_t(\langle D_i^t, \Sigma_i^t \rangle). \end{aligned} \quad (1)$$

В рамках введённой модели предполагается, что любое такое преобразование реализуется некоторой последовательностью элементарных операций $S_t \in O^+$, однако при наблюдении перехода эта последовательность может быть неизвестна или неоднозначна. Для формализации наблюдаемого перехода введём отношение *ActionContext*, отражающее изменения между состояниями как пару «вход-выход» с привязкой к некоторой последовательности операций:

$$\begin{aligned} r \in ActionContext, r_{i,j}^{t,t+1} \mid i, j \in I, t \in \mathbb{N}, r_{i,j}^{t,t+1} := \langle \langle D_i^t, \Sigma_i^t \rangle, \langle D_j^{t+1}, \Sigma_j^{t+1} \rangle \rangle, \\ \exists S_t \in O^+: S_t \mapsto Action_t. \end{aligned} \quad (2)$$

Тогда проблема учёта происхождения на данном уровне может быть сформулирована как необходимость превратить $\exists S_t$ в наблюдаемую и проверяемую информацию: обеспечить фиксацию (или однозначную идентификацию) S_t и связанного контекста исполнения (параметры, версии, промежуточные артефакты), чтобы переход $\langle D_i^t, \Sigma_i^t \rangle \rightarrow \langle D_j^{t+1}, \Sigma_j^{t+1} \rangle$ был воспроизводим.

Методы и технологии учета происхождения

Описанные здесь затруднения в научной и инженерной литературе обычно рассматриваются в рамках нескольких близких, но не тождественных направлений: учёта происхождения данных (*data provenance*), прослеживаемости преобразований и зависимостей (*data lineage*), а также трассируемости (*traceability*) связей между артефактами, этапами и решениями [10–15].

Наиболее формализованной основой для описания происхождения является модель W3C PROV [13], в которой данные представляются через взаимосвязь Entity, Activity и Agent: сущности данных создаются и используются действиями, а действия связываются с исполнителями, источниками или программными агентами.

Такая модель задает общий язык для описания того, какие данные были использованы, какая операция была выполнена, какой результат был получен, и кто или что участвовало в этом процессе.

Направление прослеживаемости преобразований акцентирует внимание не столько на общей семантике происхождения, сколько на практическом восстановлении пути данных через источники, преобразования, зависимости и потребителей результата. В инженерных системах обычно используется для анализа последствий изменений, аудита, восстановления цепочек обработки и поиска зависимостей. Понятие трассируемости подразумевает возможность устанавливать проверяемые связи между артефактами, решениями, требованиями, изменениями и результатами проверки.

Учет происхождения данных заключается в сохранении сведений о том, из каких источников получены данные, какие операции над ними выполнялись, какие входные данные использовались и какие результаты были сформированы.

Для обеспечения этой функции создаются модели описания сущностей данных, действий над ними, участников процесса и связей между входными и выходными состояниями. Дополнительно разрабатываются механизмы идентификации, журналирования, адресации, запросов и восстановления цепочек преобразований, позволяющие проверять путь данных от исходного состояния до конечного результата.

Применительно к заданной в текущем исследовании проблеме, ключевыми свойствами системы учета происхождения должны быть интегрируемость, адаптивность и расширяемость, поскольку данные поступают из разных источников, изменяются независимо и могут требовать интерпретации, которая становится очевидной только апостериорно, то есть уже после запуска и накопления данных в системе.

Поэтому особое значение приобретает организация регистрации и хранения данных: система должна фиксировать не только сами сведения, но и их контекст, связи, операции преобразования и возможность последующего расширения модели без перестройки базовой архитектуры.

Архитектура системы

Целью разработанной системы (рис. 2) является обеспечение учета происхождения, прослеживаемости преобразований и трассируемости сведений. Для обеспечения данных функций в основу системы хранения положен подход использования концепции content-addressed storage (CAS).

Он достаточно близок проблематике учета происхождения данных: идентифицирует объекты (артефакты) через их содержимое, а не через внешние имена, пути или изменяемые записи. Это позволяет анализировать похожесть сведений на уровне их физического представления.

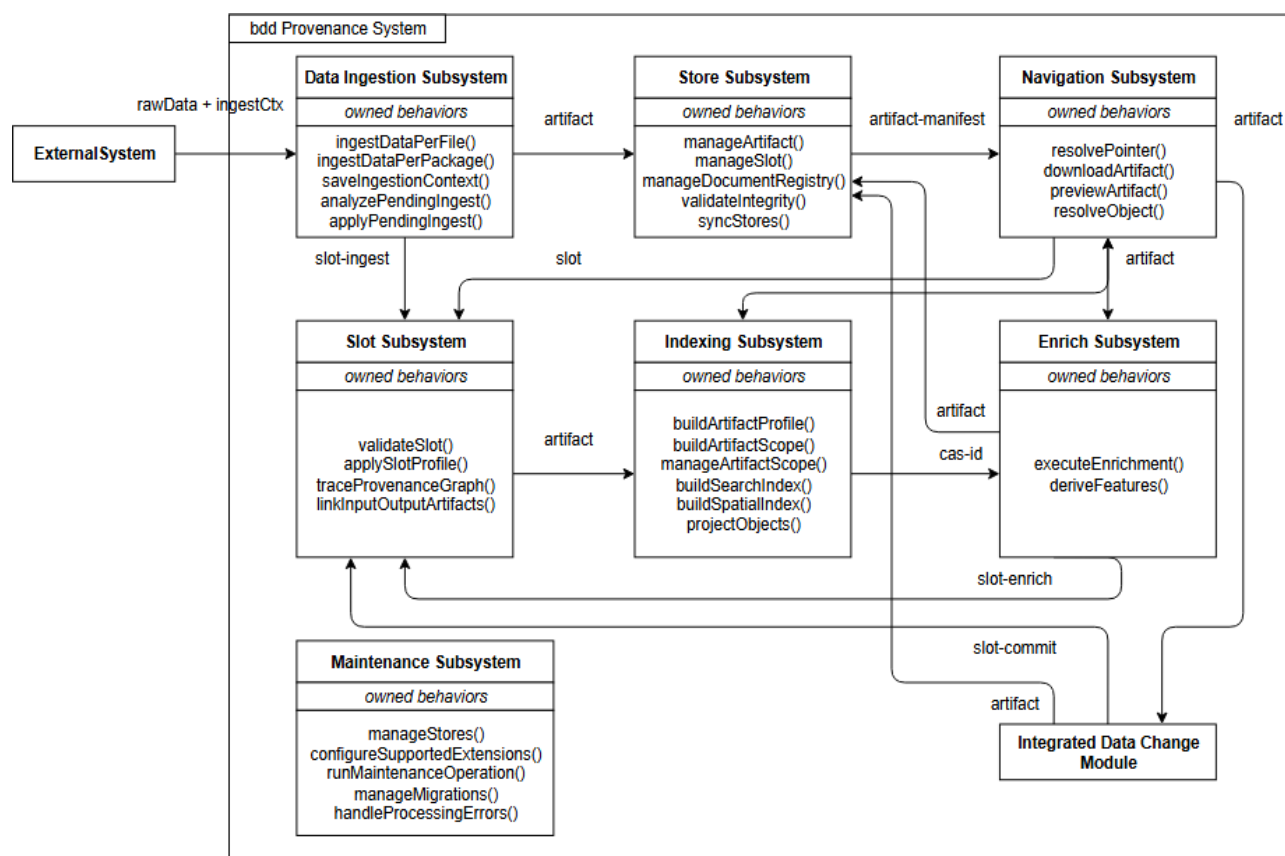


Рисунок 2. Модель взаимодействия подсистем учета происхождения Casura (Разработано автором)

Для реализации функций учета происхождений в системе отдельно от артефактов хранятся их слоты.

Слот — специальный системный объект, обозначающий устойчивый указатель на результат конкретного действия с данными: импорт, изменение состояния и использование с целью обобщения данных. Слот фиксирует контекст появления артефакта: источник, параметры процедуры, связи с входными данными, профиль операции, момент выполнения и набор полученных результатов.

Такое разделение позволяет не смешивать физическую идентичность данных с их функциональным и процедурным значением. В общем смысле, слот выражает собой концепцию *ActionContext*, сохраняя в себе информацию о входном(ых) и выходном(ых) артефактах.

Общий сценарий работы использований с данными следующий (рис. 3). Сначала внешние данные поступают в систему как обычный файл или пакет файлов. На этом этапе они не интерпретируются как конкретная предметная сущность, а помещаются в хранилище и получают уникальный идентификатор на основе своего содержимого.

Затем в реестре слотов создается запись слота «сырой» загрузки (*ingest*). Она связывает полученный артефакт с фактом загрузки. В зависимости от профиля, могут дополнительно записаны источник данных (сервис, сайт), время поступления, информация о пользователе или сервисе, выполнившем импорт, а также дополнительный контекст загрузки.

После этого возможна постобработка сведений в реестре обработки. Для загруженного артефакта формируется его профиль: например, система определяет, что это пространственный слой; фиксирует тип геометрии, систему координат, набор атрибутов и другие признаки.

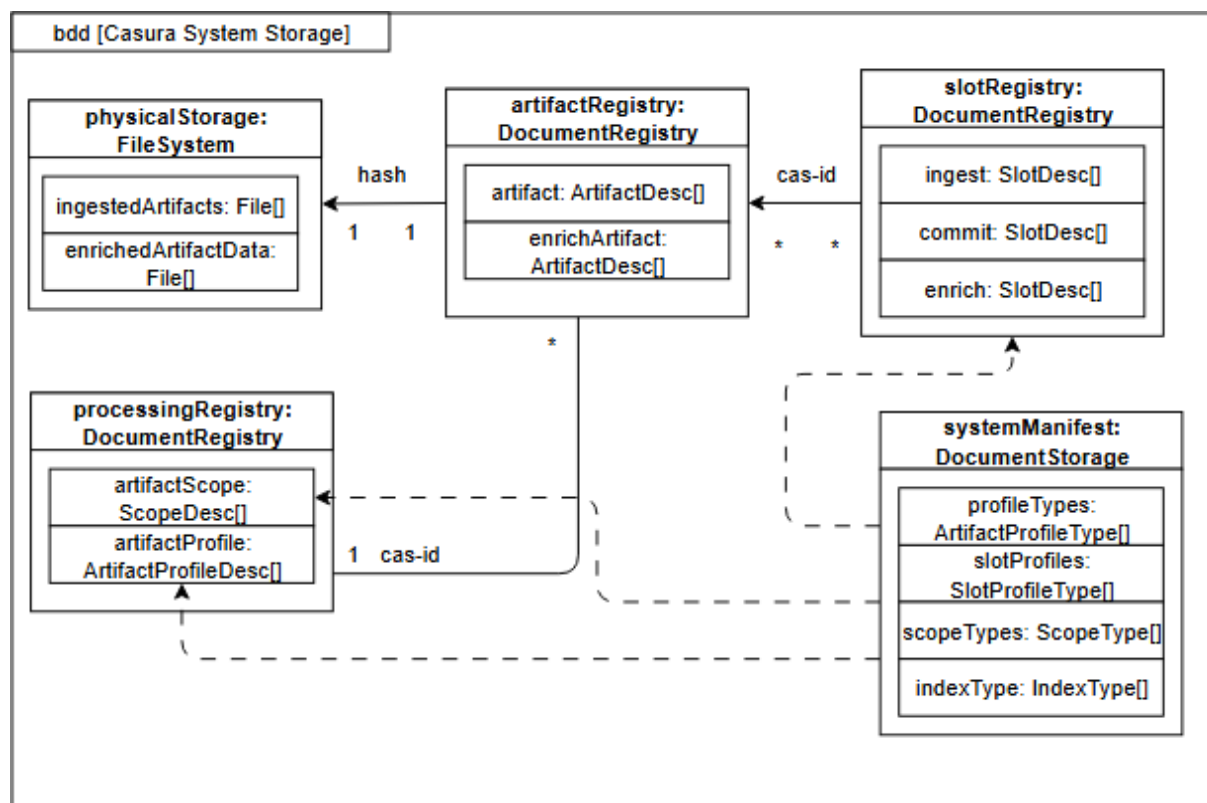


Рисунок 3. Общая модель хранилища системы учета происхождения Casura (Разработано автором)

Для реализации трассируемости сведений любые операции над данными предполагают фиксацию данной операции через слот-фиксатор (`commit`): в нем указываются входные артефакты, параметры операции (профиль слота), исполнитель, время выполнения и выходной артефакт. Если после редактирования появился новый файл слоя, он сохраняется, получает новый хэш и становится новым артефактом. Таким образом, история изменений описывается связью между входным и выходным артефактами через слот-фиксатор.

Затем может быть выполнено с помощью программных алгоритмов обогащение данных. Например, системе дано указание для заданного массива артефактов формировать производный слой на основе заранее заготовленных материалов, например, данными о пожарах, сведениями о космической съемке.

Результат обогащения сохраняется как новый артефакт, а в реестре слотов создается слот-обогащение (`enrich`), где фиксируются использованные входные артефакты, профиль процедуры обогащения, параметры анализа и полученный результат. Отличие данного слота от слота-фиксатора, результаты могут быть пересчитаны и не нуждаются в строгом хранении, и могут быть воспроизведены, фактически выступая описанием команды. В итоге с применением данной модели система позволяет для любого артефакта получить начальные сведения о его происхождении: когда он был загружен, какие и когда с ним выполнялись операции и в каких производных материалах он был использован.

Разработка системы и опыт эксплуатации системы

Из предложенной модели естественным образом следует разделение системы на базовую и предметную части. Универсальные сущности хранения и учета происхождения — артефакты, слоты, профили, области отбора и индексы — не зависят от конкретной предметной области и могут использоваться для фиксации, связывания и обработки разнородных данных.

Поэтому реализованная согласно предложенной модели система разделена на две относительно независимые подсистемы: **Casura System** (от англ. термина CAS), обеспечивающую базовые функции учета происхождения данных, и **Лес-Sense**, предметно-ориентированное расширение, адаптированное к задачам лесного хозяйства. Такое разделение позволяет отделить универсальный слой хранения, регистрации происхождения, индексирования и обогащения данных от прикладного слоя, в котором данные получают предметную интерпретацию.

Casura System реализована в виде API-сервиса и сетевой панели управления Casura Browser (рис. 4). Она обеспечивает настройку хранилищ, работу с артефактами и слотами, управление индексами, областями отбора артефактов и допустимыми профилями артефактов. Данная подсистема выступает инфраструктурной основой для фиксации, поиска и анализа данных независимо от их последующей предметной интерпретации. В самом хранилище при реализации системы для описания сложных структур, таких как слот, контекста операций, профилей, правил, связей и динамически накапливаемых зависимостей используется документная база данных, позволяющая хранить неоднородные и расширяемые записи. Она выступает логической моделью данных и кладется в основу физической модели данных формирования таблиц индексов переводится в физическую реляционную схему. Архитектура хранения сочетает неизменяемое содержимое, гибкие документы происхождения и производные структуры доступа, каждая из которых применяется в соответствии со своей задачей.

Важным условием исходя из исследования [9] возможность расширения системы и реализаций пользовательских сценариев. Для этого система позволяет расширять поведение существующих сущностей за счет реализации пользовательских профилей слотов, профилей артефактов и т. д.

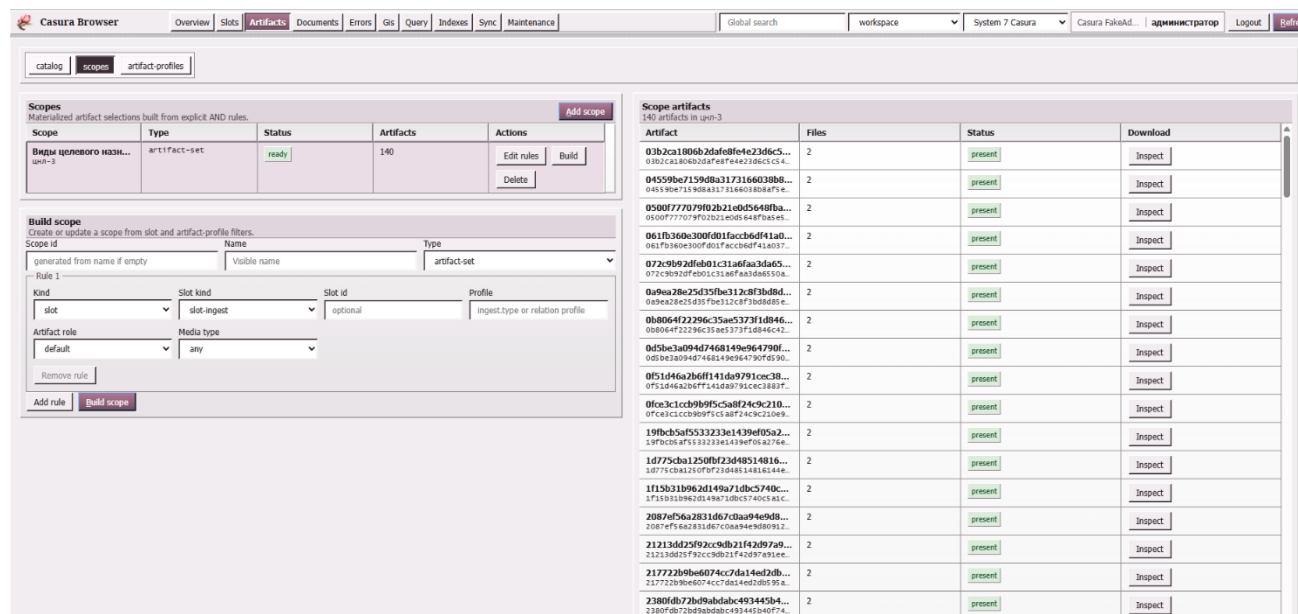


Рисунок 4. Панель управления Casura-Browser,
экран управления выборками артефактов (разработано автором)

Экран **Casura** предназначен для работы с низкоуровневыми функциями системы учета происхождения данных. Он ориентирован не на прикладное отображение слоев, а на сопровождение самого хранилища: просмотр артефактов и слотов, проверку состояния индексов, валидацию связей, удаление «висячих» артефактов, перестроение профилей артефактов и пересчет областей отбора. Через этот экран администратор или разработчик может контролировать, насколько согласованы между собой физическое содержимое, описания артефактов, слоты происхождения, профили и производные представления.

К числу ключевых низкоуровневых функций относятся валидация индексов, выявление и удаление артефактов, не связанных с актуальными слотами или областями отбора, перестроение профилей артефактов после изменения правил распознавания данных, а также пересчет массивов артефактов. Эти операции важны, поскольку система допускает постепенное накопление сведений: данные могут быть загружены раньше, чем для них появятся новые профили, правила фильтрации или прикладные сценарии использования.

При этом система спроектирована так, что частота и качество индексирования определяются пользовательским приложением и его задачами. Например, Лес-Sense обновляет слоты сырой загрузки только в момент загрузки данных через свой интерфейс, а состав артефактов, попадающих на карту, определяется уже административными настройками самой предметной подсистемы. Поэтому в хранилище может находиться условно 100 векторных артефактов, но в область отбора, используемую для отображения на карте, попадут только 50 из них. В реализованном сценарии такая фильтрация может выполняться, например, по наличию требуемой части атрибутов, что позволяет отделить все загруженные данные от данных, пригодных для конкретного прикладного использования.

Лес-Sense представляет собой интерактивную сетевую ГИС (рис. 5), выполняющую задачу сопровождения разнородных лесоустроительных данных и иных связанных сведений на карте. Именно эта подсистема задает предметный смысл универсальным сущностям системы учета происхождения: слот ingest интерпретируется как пакет загрузки данных, слот commit — как запись в истории изменений, а слот enrich на уровне пользовательского интерфейса воспринимается как результирующий слой, содержащий нормированные и интегрированные материалы, на основе всех загрузок.

Таким образом, Лес-Sense не изменяет базовую модель происхождения данных, а прикладным образом проецирует ее элементы на сценарии лесного хозяйства.

Для проверки применимости предложенного подхода в системе была апробирована совокупность материалов (рис. 5) связанных с процессами инвентаризации лесов, в частности, лесоустроительные базы данных в ретроспективных и актуальных форматах, приказы об установлении границ лесничеств, приказы об установлении категорий видов целевого назначения и особо защитных участков лесов, пакеты загрузки Федеральной Государственной информационной системы лесного комплекса, графическое описание местоположения границ и сопровождающие документы.

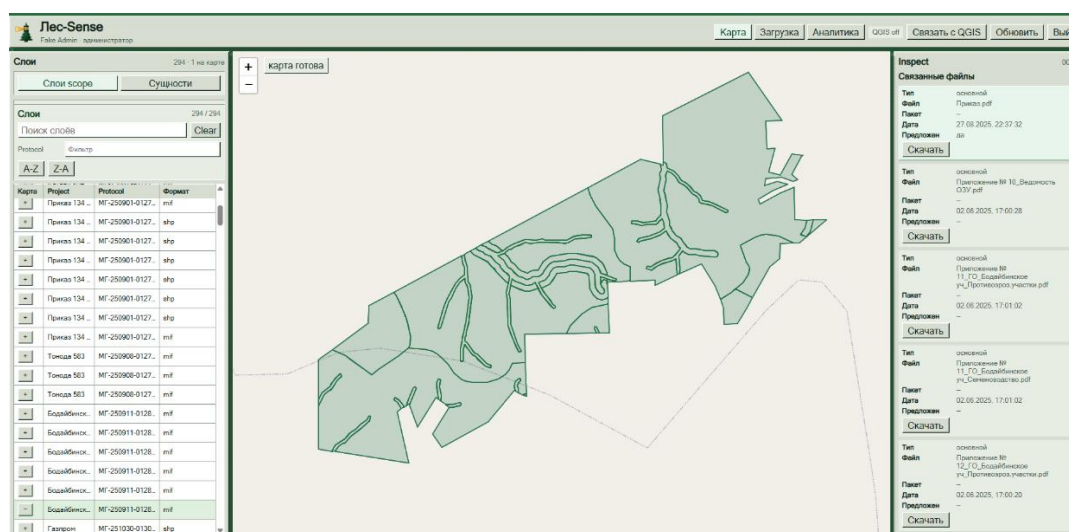


Рисунок 5. Экран панели Лес-Sense с информацией о пакетах загрузки и связанных документах по приказу Федерального агентства лесного хозяйства № 570 об установлении видов целевого назначения, Бодайбинское лесничество, Иркутская область (разработано автором)

Каждый тип материала имеет собственную структуру, формат представления и способ использования.

Нормативные документы задают правовое основание для границ и режимов использования лесов, лесоустроительные базы содержат атрибутивное описание кварталов и выделов, графические материалы фиксируют пространственное положение объектов, а пакеты ФГИС ЛК отражают требования к передаче и формализованному обмену данными.

Для анализа таких материалов в системе используются профили артефактов. Профиль задает, какие свойства материала должны быть извлечены, какие признаки подлежат проверке и каким образом данный материал может участвовать в операциях сопоставления, преобразования и учета происхождения (табл. 1).

Таблица 1

Поддерживаемые профили артефактов в системе Лес-Sense

Профиль артефакта	Что описывает	Для каких материалов применяется
document profile	Реквизиты, даты, номера, тип документа, источник, связи с приложениями	Приказы о границах, приказы о категориях защитности и ОЗУ, пояснительные записки, акты, сопроводительные письма
table/lu-data	Табличную структуру, поля, справочники, кодировки, периоды действия, идентификаторы объектов	Лесоустроительные базы, таблицы из приказов, ведомости, DBF/CSV/XML/XLSX, пакеты миграций ФГИС ЛК
table/spatial-data	Геометрию, CRS, bbox, площадь, топологические признаки, spatial overlap	Графические описания границ, GIS-слои, кварталы, выделы, ОЗУ, границы лесничеств
package profile	Состав комплекта, манифест, вложенные файлы, обязательные элементы, связи между файлами	Комплекты документов с приложениями

Разработано автором

Экран **Лес-Sense** представляет собой интерактивную ГИС-панель для работы с лесоустроительными слоями. Его основные функции включают отображение пространственных слоев на карте, выбор и просмотр объектов, подключение загруженных артефактов, отображение результирующих слоев обогащения, навигацию по данным и получение сведений, связанных с выбранной территорией или объектом.

Экран выступает предметной витриной системы: пользователь работает с понятными объектами лесного хозяйства, слоями, изменениями и результатами анализа. В некотором смысле, артефакты и слоты — это абстракция необходимая прежде всего для разработчиков системы, реализующие конкретные сценарии работы с системой.

Стандартный сценарий применения выглядит следующим образом: пользователь открывает карту, выбирает нужный набор данных или слой, загруженный через систему, после чего слой отображается в пространственном контексте.

Далее пользователь выбирает объект на карте или область интереса, получает связанные сведения, при необходимости подключает дополнительные слои и выполняет анализ или сопоставление.

Если данные были предварительно обогащены, результат отображается как отдельный тематический слой, например, слой участков с выявленными признаками, пересечениями или рисками. При этом техническая цепочка происхождения сохраняется в базовой системе, а Лес-Sense предоставляет прикладной интерфейс для ее использования в задачах лесного хозяйства. В целом для обеспечения работоспособности системы потребовалось реализовать ряд профилей для слотов и способов обогащения данных указанных на диаграмме (рис. 6).

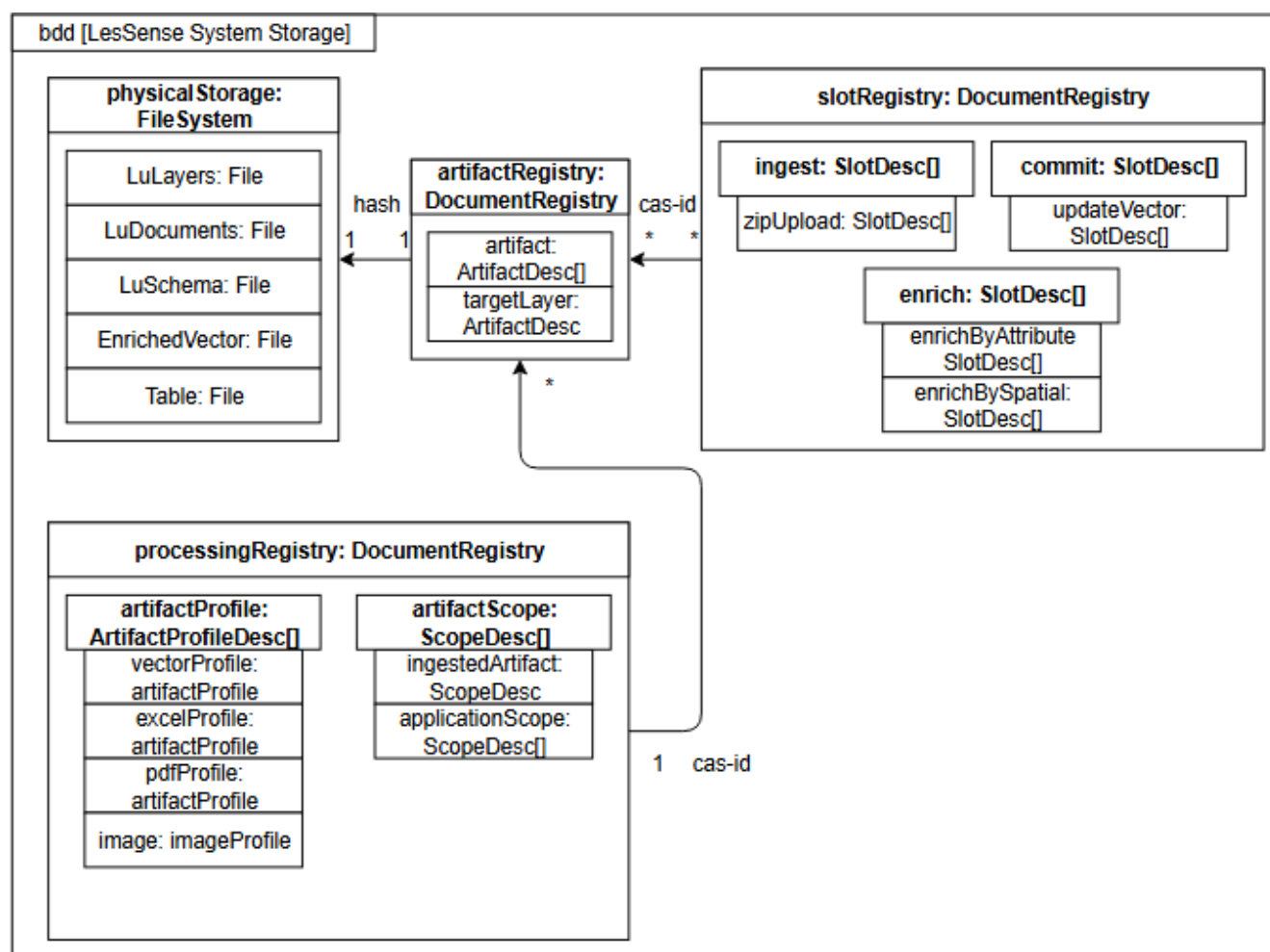


Рисунок 6. Модель хранилища сведений учета происхождения системы Лес-Sense (разработано автором)

Дополнительно для интеграции с пользовательской ГИС был разработан плагин для настольной ГИС QGIS, позволяющий связать экран системы с рабочей средой специалиста. Плагин обеспечивает получение выбранного слоя из системы, его открытие в QGIS, внесение изменений средствами настольной ГИС и последующую передачу измененного результата обратно в хранилище. Это позволяет использовать Лес-Sense как сетевую витрину и средство навигации по данным, а QGIS — как инструмент детального редактирования конкретного слоя.

Качественная оценка системы

Анализ существующих моделей и подходов к оценке качества происхождения показал наличие общих моделей представления происхождения данных и отдельных методик оценки качества учета происхождения. Вместе с тем не выявлена стандартизованная методика, одновременно учитывающая устойчивость системы к информационной неоднородности и ее операционную применимость в рассматриваемой предметной области. В связи с этим предлагается специализированная методика экспертной многокритериальной оценки. Значения частных критериев задаются по пятибалльным шкалам, а интегральный показатель рассчитывается с учетом весов критериев, соответствующих профилю использования системы. Возможные профили использования систем автор выделяет следующие:

1. **Аналитический профиль.** Система используется для задач исследования, сопоставления и анализа разнородных данных.

2. **Профиль интерактивного сопровождения.** Система внедрена для регулярной работы с данными, слоями, историей изменений и результатами проверки.
3. **Архивно-аудиторский профиль.** Система используется для задач долговременного хранения, проверки происхождения и последующего восстановления истории данных.
4. **Интеграционный профиль.** Система используется для решения проблемы интеграции: объединения данных из множества источников и обеспечивающих их последующее использование.
5. **Профиль публичного доступа.** Система используется как витрина, интерактивная карта, где происхождение данных важно, но основной акцент переносится на быстрый и безопасный доступ к уже подготовленным представлениям.

Оценку устойчивости к информационной неоднородности затруднительно представить в виде прямого количественного показателя, поскольку данный критерий во многом характеризует скорее качественную оценку способности системы работать с разнородными, неполными и косвенно связанными данными. Поэтому данный критерий целесообразно рассматривать как экспертную оценку качества работы системы в наборе сложных проверочных случаев, каждый из которых отражает отдельный аспект гетерогенности данных и способов их связывания (табл. 2).

В качестве составляющих критерия могут быть использованы следующие случаи: возможность получения информации о произвольном пространственном объекте без атрибутов; возможность получения лесотаксационных сведений по несвязанному признаку; возможность добавления пользователем собственного слоя без обращения к администратору; возможность ограничения доступа к части сведений при сохранении целостности представления данных.

Таблица 2

Оценка устойчивости к информационной неоднородности системы Лес-Sense

Сценарий	Возможность получения информации о пространственном объекте без атрибутов	Получение лесотаксационных сведений по косвенному признаку	Добавление пользовательского слоя без обращения к администратору	Ограничение доступа к части сведений при сохранении связности данных	Итоговая оценка
Баллы (5 максимум)	3	4	4	2	3.25

Разработано автором

Система может работать через геометрию, профиль и пространственное обогащение, но без атрибутов предметная интерпретация остается неполной (3 балла).

Архитектура хорошо подходит для поиска через слоты, профили и происхождение, но это зависит от качества индексации (4 балла)

Слой после загрузки еще должен пройти профилирование, попасть в нужную выборку и соответствовать правилам прикладной системы (4 балла).

Модель прав на артефакты и слоты не реализована; но имеется ограничение на режим только чтение для неавторизованных пользователей (2 балла).

Оценка операционной применимости характеризует свойства самой системы как программно-информационной среды, обеспечивающей хранение, доступ, сопровождение и развитие механизмов учета происхождения данных.

Он направлен на оценку эксплуатационных свойств системы: поиск, навигацию, контроль целостности, обновление производных представлений, управление доступом и сопровождение изменений (Табл. 3).

В рамках данного критерия необходимо отразить простоту или сложность расширения системы новыми типами данных, профилями, слоями и правилами обработки; возможность отслеживания цепочек происхождения и изменений; согласованность индексов, профилей и областей отбора с фактическим состоянием хранилища; скорость получения нужных объектов и связанных сведений; наличие механизмов проверки, восстановления и очистки ошибок; а также понятность административных действий, необходимых для поддержания системы в рабочем состоянии.

Таблица 3

Оценка операционной применимости системы Лес-Sense

Сценарий	Скорость поиска и навигации по объектам	Согласованность индексов, профилей и областей отбора	Восстановимость цепочек происхождения и изменений	Простота расширения новыми типами данных, слоями и правилами	Обнаружение и исправление нарушений целостности	Итоговая оценка
Баллы (5 максимум)	5	3	4	3	4	3.8

Разработано автором

Система использует индексы, профили, области отбора и объектные представления, поэтому типовые сценарии поиска и перехода к нужным артефактам не требуют полного просмотра всего хранилища (5 баллов).

Согласованность обеспечивается через пересчет и валидацию, но зависит от своевременного запуска этих процедур и правил, заданных пользовательским приложением (3 балла).

Связи между артефактами и слотами позволяют восстановить основные этапы появления и изменения данных, однако полнота описания зависит от того, насколько подробно фиксируются параметры операции и исполнитель (4 балла).

Архитектура допускает добавление новых типов данных, профилей, слоев и правил, но такое расширение требует понимания модели профилей, областей отбора и слотов (3 балла).

Система предусматривает проверки, перестроение профилей и областей отбора, удаление «висячих» артефактов и служебные операции восстановления, но часть действий требует административного контроля (4 балла).

Принимаем 3,25 — устойчивость к информационной неоднородности, 3,8 — операционная применимость. После получения оценок по базовым критериям необходимо определить центральный интегральный показатель — качественную оценку учета происхождения данных (табл. 4). Он отражает не отдельное свойство системы, а ее общую пригодность к решению задач, связанных с идентификацией объектов, сохранением контекста происхождения, восстановлением цепочек преобразования и эксплуатационным сопровождением данных. Значение данного показателя формируется как взвешенное сочетание критерия информационной гетерогенности и критерия эксплуатационной состоятельности, при этом веса задаются в зависимости от профиля использования системы.

Полученная итоговая оценка в 3,53 показывает, что система уже обеспечивает достаточно высокий уровень работы с информационно гетерогенными данными. Полученное значение для рассматриваемой предметной области — это существенный результат, поскольку специализированные системы учета происхождения данных, ориентированные на задачи лесного хозяйства, фактически пока не имеют устоявшихся аналогов и методик оценки.

Таблица 4

Расчет качественной оценки учета происхождения данных в системе Лес-Sense

Профиль оценки	Вес информационной гетерогенности	Вес эксплуатационной состоятельности	Итоговая оценка по профилю
Аналитический профиль	0,7	0,3	$0,7 \cdot 3,25 + 0,3 \cdot 3,8 = 3,42$
Профиль интерактивного сопровождения	0,4	0,6	$0,4 \cdot 3,25 + 0,6 \cdot 3,8 = 3,58$
Архивно-аудиторский профиль	0,5	0,5	$0,5 \cdot 3,25 + 0,5 \cdot 3,8 = 3,53$
Интеграционный профиль	0,6	0,4	$0,6 \cdot 3,25 + 0,4 \cdot 3,8 = 3,47$
Профиль публичного доступа	0,3	0,7	$0,3 \cdot 3,25 + 0,7 \cdot 3,8 = 3,64$
Общая применимость системы	—	—	$\frac{3,42 + 3,58 + 3,53 + 3,47 + 3,64}{5} = 3,53$

Разработано автором

По результатам оценки система в большей степени проявляет себя как интерактивная сетевая ГИС с функциями учета происхождения данных. Наиболее высокие значения получены для задач, где важны скорость доступа, навигация по объектам, отображение подготовленных представлений и эксплуатационная устойчивость: профиль публичного доступа и профиль интерактивного сопровождения. Это показывает, что система наиболее применима в сценариях, где пользователь работает с уже загруженными, проиндексированными и связанными данными через прикладной интерфейс.

Ключевым ограничением остается сложность разграничения доступа. В текущей реализации система ориентирована преимущественно на работу с общедоступными сведениями, поэтому вопросы скрытия части данных, ограничения доступа к отдельным артефактам, слотам, профилям или цепочкам происхождения требуют дальнейшего развития. Основная проблема состоит в противоречии между трассируемостью и секретностью: чем подробнее система фиксирует происхождение данных и связи между ними, тем выше риск раскрытия служебного или ограниченного контекста. Дополнительным существенным фактором является сложность программной реализации такой архитектуры, поскольку необходимо поддерживать согласованность между физическим хранением, документными реестрами, слотами, профилями, индексами, областями отбора и прикладными представлениями. Разрешение этих вопросов является задачей дальнейшего развития системы.

Заключение

В данной работе был рассмотрен вопрос применения методов учета происхождения в лесном хозяйстве путем разработки и оценки специализированной системы на ее эффективность. Показана значимость задачи учета происхождения и версий данных, поставлена формально проблема учета, рассмотрены ключевые компоненты системы, модель данных и ее предметное расширение. Система совмещает в себе сущности свойственные для моделей учета происхождения и подходы к построению программных систем, позволяя формировать предметно-ориентированные расширения на базе профилей артефактов и профилей слотов. Была разработана программная реализация данной системы и на тестовых данных из открытых источников выполнена апробация данной системы. По результатам оценки были сделаны выводы, что система обеспечивает достаточно высокий уровень работы с информационно гетерогенными данными, совмещая свойства геоинформационных систем с функциями учета происхождения данных. Наиболее высокие оценки получены для задач, где важны скорость доступа, навигация по объектам, отображение подготовленных представлений.

Имеются риски безопасности в части противоречия между трассируемостью и секретностью: чем подробнее система фиксирует происхождение данных и связи между ними, тем выше риск раскрытия служебного или ограниченного контекста. Достоинства системы прямо порождают ее недостатки: расширение круга обрабатываемых данных, учет эволюции и использование абстракций повышают гибкость, но одновременно делают систему менее очевидной для восприятия, более трудной для освоения и более требовательной к качеству сопровождения. В этом смысле инновационность подхода действительно оборачивается более высоким порогом внедрения. Область применения разработанной системы определяется характером самой организационной среды. Чем выше доля параллельных, слабо формализованных и плохо отслеживаемых процессов, тем выше практическая значимость системы. Разрешение этих вопросов является задачей дальнейшего развития системы.

ЛИТЕРАТУРА

1. Бородина О. Б. Государственный лесной реестр как информационная база в системе устойчивого управления землями лесного фонда российской федерации / Бородина О. Б., Чуксин И. В., Фомина А. В., Желонкина Е. Э. — DOI 10.55186/2413046X_2022_7_4_204 // Московский экономический журнал — 2022. — Т 7 № 4 — URL: https://elibrary.ru/doi_resolution.asp?doi=10%2E55186%2F2413046X%5F2022%5F7%5F4%5F204 (дата обращения 30.06.2026).
2. Данеев А. В. Разработка геоинформационной системы в сфере лесоустройства / Данеев А. В., Данеев Р. А., Диомидов И. А. — DOI 10.26731/2658-3704.2020.3(8).89-98 // Информационные технологии и математическое моделирование в управление сложными системами — 2020 — № 3(8) — С. 89–98 — URL: <https://elibrary.ru/item.asp?id=44200274> (дата обращения 30.06.2026).
3. Колесникова А. В., Основные проблемы и вызовы в развитии отечественного лесного комплекса / DOI: 10.30680/ECO0131-7652-2013-11-5-25 // ЭКО — 2013 — Т 43 № 11 — С. 5–25. URL: <https://cyberleninka.ru/article/n/osnovnye-problemy-i-vyzovy-v-razvitii-otechestvennogo-lesnogo-kompleksa> (дата обращения 30.06.2026).
4. Колесник В. Г. Система государственного управления лесным комплексом: Текущая ситуация и основные проблемы / Колесник В. Г., Синятуллина Л. Х. // Вопросы государственного и муниципального управления — 2017 — № 1 — С. 129–148 — URL: <https://cyberleninka.ru/article/n/sistema-gosudarstvennogo-upravleniya-lesnym-kompleksom-tekuschaya-situatsiya-i-osnovnye-problemy> (дата обращения 30.06.2026).
5. Архипов В. И. Современная технология таксации лесов дешифровочным способом «От съемки — к проекту / Архипов В. И., Черниковский Д. М., Березин В. И., Белов В. А. // Известия Санкт-Петербургской лесотехнической академии — 2014 — № 208 — С. 22–42 — URL: <https://www.elibrary.ru/item.asp?id=22909975> (дата обращения 30.06.2026).
6. Хлюстов В. К. Технология комплексной оценки древесных ресурсов методами дистанционного зондирования Земли / Хлюстов В. К., Юрчук С. А., Хлюстов Д. М., Ганихин А. М. — DOI 10.26897/1997-6011-2021-4-129-138 // Природообустройство — 2012 — № 4 — С. 129–138 — URL: <http://elib.timacad.ru/dl/full/gmgup-18-2021-4.pdf> (дата обращения 30.06.2026).
7. Barredo J.I., Mapping and assessment of primary and old-growth forests in Europe / Brailescu C., Teller A., Sabatini F.M., Mauri A., Janouskova K. — DOI: 10.2760/797591 // Publications Office of the European Union — 2021 — URL: <https://publications.jrc.ec.europa.eu/repository/handle/JRC124671> (дата обращения 30.06.2026).

8. Диомидов И. А. Развитие систем обработки информации о лесах и лесных ресурсах / И. А. Диомидов — DOI: 10.15862/02INOR224 // Отходы и ресурсы. — 2024. — Т. 11. № 2. — URL: <https://resources.today/PDF/02INOR224.pdf> (дата обращения 30.06.2026).
9. Диомидов И. А. Модель интегрированной системы обработки информации о лесах и лесных ресурсах / И. А. Диомидов — DOI: 10.15862/09INOR424 // Отходы и ресурсы. — 2024. — Т. 11. № 4. — URL: <https://resources.today/PDF/09INOR424.pdf> (дата обращения 30.06.2026).
10. Herschel M., A survey on provenance: What for? What form? What from? / Herschel M., Diestelkämper R., Ben Lahmar H. — DOI 10.1007/s00778-017-0486-1 // The VLDB Journal — 2017 — V. 26, № 6 — pp. 881–906 — URL: <https://dl.acm.org/doi/10.1007/s00778-017-0486-1> (дата обращения 30.06.2026).
11. Simmhan Y.L., A survey of data provenance in e-science / Plale B., Gannon D. — DOI 10.1145/1084805.1084812 // SIGMOD Rec. — 2005 — V.34 №.3 — pp. 31–36 — URL: https://www.researchgate.net/publication/220415414_A_Survey_of_Data_Provenance_in_e-Science (дата обращения 30.06.2026).
12. Cheney J. Provenance in Databases: Why, How, and Where / Cheney J., Chiticariu L., Tan W.C. — DOI 10.1561/19000000006. // Foundations and Trends in Databases — 2009 — V. 1 № 4 — pp. 379–474 — URL: https://www.researchgate.net/publication/220626609_Provenance_in_Databases_Why_How_and_Where (дата обращения 30.06.2026).
13. Missier P. The W3C PROV family of specifications for modelling provenance metadata / Missier P. Belhajjame K., Cheney J., DOI 10.1145/2452376.2452478 // Genoa Italy, in Proceedings of the 16th International Conference on Extending Database Technology, 2013, pp. 773–776. — URL: https://www.researchgate.net/publication/266369089_The_W3C_PROV_family_of_specifications_for_modelling_provenance_metadata (дата обращения 30.06.2026).
14. Papadakis G. A Survey of Blocking and Filtering Techniques for Entity Resolution / Skoutas D., Thanos E., Palpanas T. — <https://doi.org/10.48550/arXiv.1905.06167> // Association for Computing Machinery — 2020 — V. 1, № 1 — URL: <http://arxiv.org/abs/1905.06167> (дата обращения 30.06.2026).
15. Schreiber A., GitLab2PROV Provenance of Software Projects hosted on GitLab / Andreas S., de Boer.C, and von Kurnatowski L. // 13th International Workshop on Theory and Practice of Provenance, July 2021. — URL: https://www.researchgate.net/publication/353346345_GitLab2PROV-Provenance_of_Software_Projects_hosted_on_GitLab (дата обращения 30.06.2026).

Diomidov Ivan Andreevich

Irkutsk State Transport University, Irkutsk, Russia
Roslesinforg, Irkutsk, Russia
Pribaikallesproject

E-mail: pilad0hwts@yandex.ru

RSCI: https://elibrary.ru/author_profile.asp?id=1220889

Developing an integrated support system for forest resource inventory based on data provenance in heterogeneous information environments

Abstract. The application of data provenance methods is a relevant area of research in forestry due to the variability and historical heterogeneity of forest inventory materials. The aim of this study is to develop a system for supporting forest resource inventory in order to ensure the reproducibility of results, the traceability of transformations, and the accounting of data provenance. The author characterizes the problem of developing combined integration systems for forestry and briefly considers the application of data provenance methods. The article examines the interaction of the key system components, the logical provenance model used in the system, and the main operational scenarios. On this basis, the main properties and functions of the software implementation are characterized: the Casura provenance accounting system and the domain-oriented extension for forestry, Les-Sense. The article provides a numerical assessment of the system's properties based on a methodology developed by the author through the formation of an expert multi-criteria evaluation of the provenance accounting system as applied to the specific features of forestry. The system under consideration provides a sufficiently high level of support for informationally heterogeneous data, combining the properties of geographic information systems with data provenance functions. The highest scores were obtained for tasks in which access speed, object-based navigation, and the display of prepared views are of primary importance. The study contributes to the development of approaches to provenance accounting and the integration of heterogeneous data and may be useful for specialists and researchers in the fields of systems analysis and forestry.

Keywords: system analysis; data provenance; forestry; information systems; geographic information systems; digitalization; system