

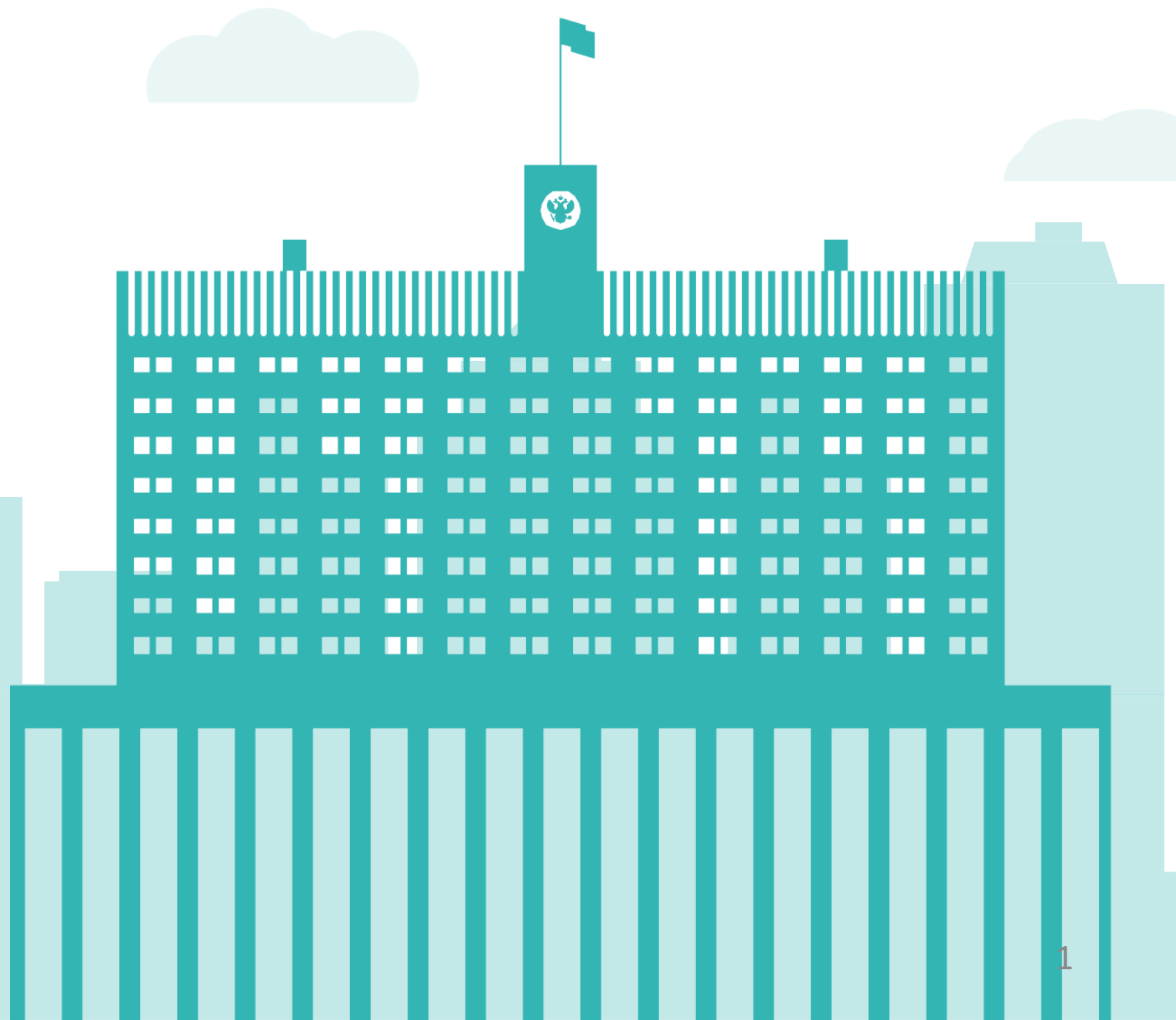
НАЦИОНАЛЬНАЯ СИСТЕМА УПРАВЛЕНИЯ ДАННЫМИ



АНАЛИТИЧЕСКИЙ ЦЕНТР
ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Александр Малахов

*Начальник Управления
методологического сопровождения
работы с данными*



Основная задача НСУД – обеспечить качество данных по всем критериям включая доступность и безопасность

Задачи Правительства РФ

- ✓ Хочу получать непротиворечивые сведения от разных ведомств
- ✓ Хочу дать качественный стимул для развития «цифровой экономики»
- ✓ Хочу иметь дополнительный источник принятия решений

Задачи ОГВ

- ✓ Нет данных, хочу данные
- ✓ Есть «плохие» данные, хочу «хорошие» данные (например для госуслуг)
- ✓ Есть «хорошие» данные, хочу выгоду (аналитика или продажа)
- ✓ Есть данные и куча запросов, хочу уменьшить количество запросов
- ✓ Есть данные, хочу решить другие проблемы при помощи данных

Задачи Минэкономразвития России

- ✓ Хочу, чтобы данные стали новым активом
- ✓ Хочу упорядочить госуправление

Задачи Минкомсвязи России

- ✓ Хочу обеспечить безопасность значимых госданных
- ✓ Хочу обеспечить доступность максимально возможного объема госданных
- ✓ Хочу сформировать условия для формирования рынка данных

НСУД в цифровом госуправлении



АНАЛИТИЧЕСКИЙ ЦЕНТР
ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ



Данные ИР содержат **много ошибок**, требуется регулярно измерять и повышать их качество, а также добиться **высокого качества данных по существенным атрибутам** (напр., **эталонным**)

Сейчас

1. **Проверка** данных в СМЭВ осуществляется **только по правилам ФЛК**, заданным в видах сведений СМЭВ
2. Отсутствует **возможность проверки качества** всего массива данных. Для оценки качества всего массива данных ОГВ требуется провести аудит ведомственного ИР (проект длительностью несколько месяцев)
3. Виды сведений СМЭВ сформированы по технологическим правилам, **отсутствуют методологические правила**, которые позволяли бы **интерпретировать данные**
4. **Не ясно, качество каких данных критично** (существенные атрибуты)

Ограниченный контроль качества данных только при обмене через СМЭВ

Будет с НСУД

1. **Существенные атрибуты установлены в видах данных**, которые уточнят вид сведений СМЭВ, зададут метод. правила формирования данных для межвед. обмена
2. Хранение **правил качества** и управление правилами происходит в «ядре» ЕИП НСУД. Существует **регламент их постоянного пополнения и обновления**
3. **Возможна** регулярная, автоматизированная **проверка всего массива ведомственных данных**, расположенных в витринах данных ЕИП НСУД
4. **Отсутствие жёсткого запрета** на публикацию ошибочных данных при полном контроле за качеством данных в ходе регулярных проверок

Широкий набор инструментов объективного контроля качества всех данных НСУД

Необходимо **кратное ускорение** скорости обработки данных

(например, в ходе внедрения суперсервисов запрос на получение данных о количестве детей в семье должен обрабатываться не 10 минут, а 10 секунд; невозможно заранее определить количество таких ситуаций, которые будет необходимо разрешить в ходе реализации НП «Цифровая Экономика»)

Сейчас

1. Необходимо **быстрое (ASAP) расширение** инфраструктуры ОГВ. Если в ЦОДе ОГВ нет достаточных резервов, то возникнет цикл закупки. Срок расширения инфраструктуры может составить **от 6 месяцев до 1 года**
2. Если ПО ОГВ не поддерживает горизонтальное масштабирование и/или не выдерживает новые показатели нагрузки, потребуется его **модернизация**. Срок разработки ПО может составить **от 1 до 2 лет**

Возрастание нагрузки ведёт к **большим временным и финансовым затратам**, что может стать препятствием на пути реализации национальной программы «Цифровая экономика Российской Федерации»

Будет с НСУД

1. **Расширение инфраструктуры** происходит **автоматически** за счет облачного типового решения НСУД «Витрины данных», масштабируемого с использованием единого облачного ресурса ГЕОП
2. Затраченное время составляет **от 1 часа до 5 дней**, в зависимости от объема данных, хранящихся на витрине данных ЕИП НСУД
3. Использование ГЕОП позволяет **существенно снизить затраты** на резервные мощности для масштабирования, за счет единого пула ресурсов

Возрастание способности обработки растущей нагрузки происходит **оперативно и с меньшими затратами**
Экономия за счёт масштаба (унифицированное решение vs модернизация каждой ИС ОГВ)

Обеспечение доступности данных: юридическая значимость данных



Для перехода на реестровую модель ведения данных* необходимо обеспечить
неизменяемость уже использованных данных

** когда запись в реестре является подтверждением юридического факта*

Сейчас

1. Если данные ОГВ скомпрометированы (есть признаки изменённых «задним числом» данных), то **проверка подлинности данных затруднена**
2. **Нет уверенности** в том, что **данные не были изменены** «задним числом» в принципе
3. Определение факта компрометации – сложная задача, которая в настоящее время **не имеет стандартных решений**

Единой защиты от компрометации данных нет

Будет с НСУД

1. **Работа с данными унифицирована** с помощью разработанных и утвержденных методик, механизма витрин данных ЕИП НСУД и подсистемы доступа к данным СМЭВ
2. Происходит **постоянная проверка подлинности данных**, что делает невозможным изменение данных «задним числом»
3. **Возможна проверка подлинности данных**, поскольку реализован специализированный криптопротокол и подписи журнала транзакций витрин данных хранятся в НСУД

Обеспечивается **неизменяемость уже использованных данных**, есть механизмы соответствующего контроля

Обеспечение доступности данных: поддерживаемые типы запросов (1 из 2)



Ситуация

Требуется **выполнить сложный запрос, включающий выборку данных из источника и анализ данных для принятия решений**

- (1) *выяснить средний показатель здоровья новорожденных по шкале Апгар с группировкой по родильным домам, врачам, регионам;*
(2) *найти всех родственников по связям ребёнок-родитель*

Сейчас

1. Можно выполнить только **предопределённые запросы** к данным, **которые поддерживаются** поставщиком данных
2. **Новый запрос к данным** требует поддержки со стороны поставщика данных, которому нужно **доработать ИС ОГВ**.
3. Время доработки ИС ОГВ **от 1 до 3 месяцев** (при наличии соот. обязательств у разработчика)

Только простые и предопределённые запросы.
Расширение числа запросов – **время и деньги**.

Будет с НСУД

1. Можно выполнить **произвольные запросы** следующих типов:
 - Использование фильтров
 - Аналитические запросы (для кейса №1)
 - Полнотекстовые запросы
 - Графовые запросы (для кейса №2)
2. Подготовка аналитики через модуль BI, которым может выступать ГАС «Управление»
3. Время подготовки запроса **от 1 до 7 дней**

Поддержка сложных запросов при **небольших временных затратах**

Обеспечение доступности данных: поддерживаемые типы запросов (2 из 2)



Требуется выполнить распределённый запрос, включающий объединение данных из разных источников и обезличивание результатов для принятия управленческих решений

(например, необходимо составить реестр недвижимости (кадастровый номер, адрес недвижимости), принадлежащей генеральным директорам предприятий региона. Для этого потребуется получить данные из ЕГРН, ЕГРЮЛ и Реестра паспортов МВД)

Сейчас

1. При выборке данных из разных источников необходимо провести объединение данных, что требует поддержки со стороны поставщиков данных, которому нужно **доработать ПО** или привлечь разработчика для исполнения запроса
2. Для объединения нужно загрузить сведения из паспортов генеральных директоров из МВД, что требует **высокого уровня доступа**
3. Организация процесса потребует **участия руководителей ведомств и больших орг. затрат**
4. Технологически, процесс может занять **до 6 месяцев**

Только простые и predetermined запросы.
Расширение числа запросов – **время и деньги.**

Будет с НСУД

1. При обработке запроса результирующий набор данных будет сформирован на основании правил, заложенных в «ядре» ЕИП НСУД по установленным протоколам или на их основе, и потребитель получит виды данных и порядок сбора. При этом, персональные данные, необходимые для выполнения запроса, потребителю не передаются
2. Организационно **выполняется аналитиком** без участия руководителя
3. Процесс подготовки запроса **от 1 до 7 дней** (при наличии протоколов это занимает **менее 1 часа**)

Поддержка распределённых запросов при **небольших временных и финансовых затратах**

Обеспечение качества данных: сравнение данных из разных источников



АНАЛИТИЧЕСКИЙ ЦЕНТР
ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Ситуация

Требуется **провести сопоставление данных** как один из способов контроля качества
(например, в ЕГРН и в ФИАС записи о новых объектах недвижимости появляются несинхронно и часто с большой задержкой по времени)

Сейчас

1. **Невозможно системно** проводить **сопоставления данных**. Для их проведения требуется выполнение сложного проекта с вовлечением двух и более ОГВ
2. Проведение подобного проекта будет затруднено необходимостью сбора продуктивных данных для проверки, что может быть невозможно из соображений безопасности (**нет доступа к данным** или поставщик не имеет законного механизма предоставить эти данные)
3. Время выполнения подобного исследования – **от 2 месяцев до 1 года** (с учетом изменения НПА и подписания соглашений)

Сверка массивов – **сложный, длительный и дорогой проект**

Будет с НСУД

1. Разработаны **протоколы («стандарты») сверки и объединения данных**
2. Описаны и систематизированы **механизмы доступа к данным**
3. Алгоритм сопоставления исполнен на витринах данных ЕИП НСУД и в «ядре» ЕИП НСУД, а в отчет войдут только статистические данные о выявленных расхождениях
4. Можно выполнить **распределенный запрос** (стандартный механизм ЕИП НСУД) и **выявить расхождения данных** в ИР ОГВ
5. Время выполнения подобного запроса – **от 10 минут до 48 часов**

НСУД позволит **существенно упростить сопоставление данных** в больших массивах

Требуется **исправить ошибку в данных в реестре, переведенном на реестровую модель**

Сейчас

1. **ОГВ не имеет полномочий самостоятельно править ошибки**, запросы к поставщику данных или невозможны (поставщик – гражданин) или не приносят результата (поставщик – ИС не заинтересован)
2. **Не существует порядков** по исправлению ошибок (за редким исключением)

Действующее законодательство **не позволяет вносить изменения** в юридически значимые данные

Будет с НСУД

1. Разработаны порядки исправления **технических ошибок** без привлечения поставщика
2. Разработаны порядки по **уменьшению количества собираемых данных** при отсутствии их использования
3. Владельцы данных наделены дополнительными **полномочиями по запросу корректных сведений** от поставщиков
4. Инструменты ЕИП НСУД позволяют **оперативно осуществлять взаимодействие с поставщиками** с целью исправления ошибки

НСУД **позволит исправлять ошибки** в юридически значимых данных

Потребитель узнает, что он **получил** в прошлом от поставщика **некорректные данные**, на основании которых была оказана услуга

Сейчас

1. Текущая организация взаимодействия **не обеспечивает хранения достаточного объема информации для расследования инцидентов**
2. **Ретроспективные запросы** (когда можно автоматизировано обратиться к данным в том состоянии, в котором они находились на в определенный момент времени в прошлом) **не поддерживаются**

Будет с НСУД

1. Витрина данных ЕИП НСУД поставщика **содержит все версии данных**, которые когда-либо были актуальны, а НСУД содержит контрольные суммы каждой версии данных этой витрине. Можно **получить** детальную **историю изменений** данных
2. Благодаря тому, что витрины данных ЕИП НСУД **поддерживают ретроспективные запросы** потребитель может собрать достаточно информации для формирования инцидента
3. Возможна **совместная работа** Поставщика, Потребителя и Центра компетенций над данным инцидентом

НСУД реализует инструменты **объективного разбора инцидентов** качества данных

Бесконтрольное использование данных при разовом получении доступа

(например, подписка в СМЭВ на данные ЕГРН не позволяет отслеживать, накапливает ли потребитель данные, что нарушает законодательство)

Сейчас

1. Модель взаимодействия через predetermined запросы (виды сведений СМЭВ), позволяет потребителю накапливать данные поставщиков и обрабатывать их локально
2. После того, как права на доступ к виду сведений СМЭВ предоставлены, **отсутствует контроль** со стороны поставщика **за накоплением** данных у потребителя
3. В случае необходимости удаления отдельных записей, процесс удаления необходимо проводить вручную. Автоматизированный **контроль удаления записей** из массива накопленных потребителем данных **невозможен**

Отсутствие **контроля за использованием и копированием данных**

Будет с НСУД

1. Предоставляются **права на доступ и на накопление данных** в реплике (по отдельному реестру инф.взаимодействий, разработанному в составе инфраструктуры электронного правительства)
2. Возможно реализовать **запрет на хранение «чужих» данных** и механизм «забвения» данных вне НСУД (например, в случае использования программы защиты свидетелей)

Контроль за использованием данных, управление их доступностью, «забвение» данных

Приложение

НСУД решает проблему качества данных

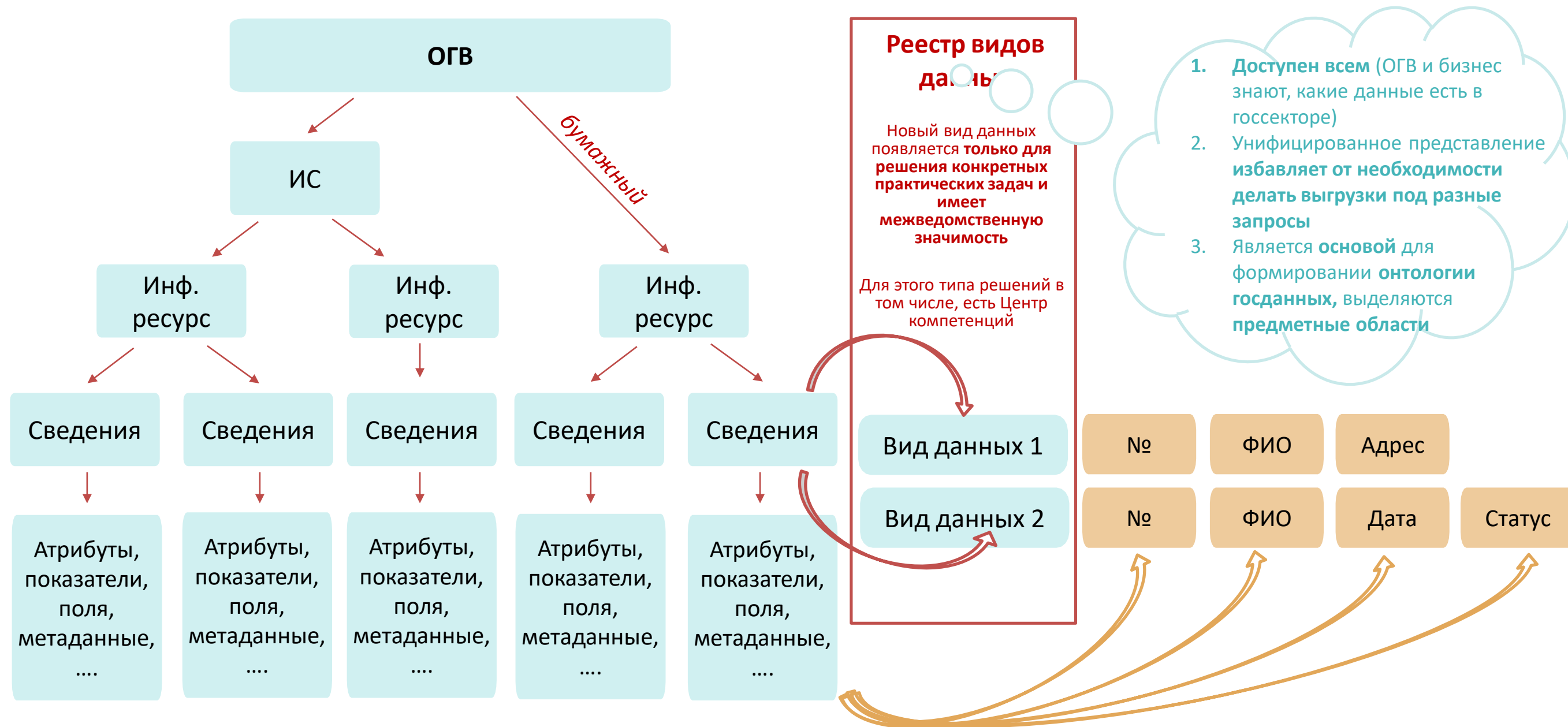


КАЧЕСТВО

<i>Методология</i> Что делаем?	Почему?	Что получаем?
Создаем требования к управлению данными внутри ОГВ	<ul style="list-style-type: none">ОГВ собирает данные в любом формате, структуре, онтологииОГВ не проверяет, есть ли эти данные уже в госсектореОГВ не проектирует использование данных в будущем	Единые подходы к управлению данными внутри ОГВ (управление жизненным циклом) приведут к правильной интерпретации данных
Устанавливаем требования к определению эталонности данных	<ul style="list-style-type: none">ОГВ не может определить, какому источнику доверять	Понятные правила и порядки получения юр.значимых данных, синхронизация данных в различных ИС
Создаем требования по порядку работы с выявленными ошибками в собственном ресурсе	<ul style="list-style-type: none">ОГВ не может поправить данные, которые есть в собственном ресурсе, но источником которых он не является	Регламенты по исправлению тех. ошибок без привлечения поставщика (напр. гражданина) или полномочия по работе с поставщиком
Создаем требования по порядку работы с выявленными ошибками в полученных данных	<ul style="list-style-type: none">ОГВ не может поправить полученные данные, хотя именно на их основе формируются решения	Порядки информационного взаимодействия по выявленным инцидентам
Проверяем качество при информационном обмене	<ul style="list-style-type: none">Данные ОГВ проверяются только при оказании госуслуг на прохождение ФЛК, больше проверок нет	Весь информационный обмен построен по единым принципам и технологиям, состав проверок расширен
Делаем данные более доступными (сл.слайд)	<ul style="list-style-type: none">В данных, которые используются мало, трудно выявить ошибки	Повышение использования данных

+ Технология

НСУД решает проблему доступности данных



НСУД решает проблему интероперабельности и безопасности данных

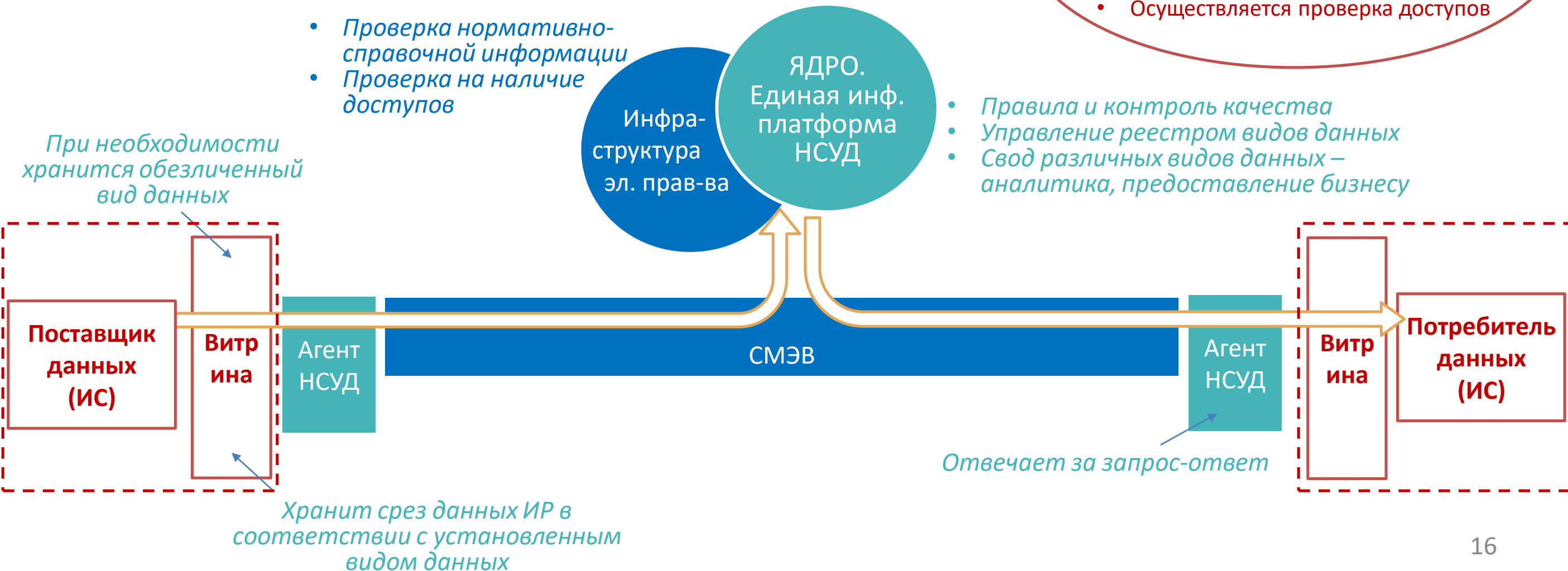


- Данные хранятся в «витринах»
- С помощью ядра осуществляются проверки

- Проверка нормативно-справочной информации
- Проверка на наличие доступов

- Данные в Ядре единой инф. платформы НСУД **НЕ ХРАНЯТСЯ**
- Каналы Инфраструктуры электронного правительства максимально защищены
- Осуществляется проверка доступов

- Правила и контроль качества
- Управление реестром видов данных
- Свод различных видов данных – аналитика, предоставление бизнесу



Участники

1. **ОБСЛЕДОВАНИЕ**
описывают информационные ресурсы
2. **МЕТОДОЛОГИЯ**
определяют требования к структуре, свойствам и связанности данных
3. **НПА**
корректируют / вносят предложения по доработке НПА
4. **ТЕХНОЛОГИЯ**
подключают ИС к ЕИП НСУД
5. **РЕАЛИЗАЦИЯ**
обеспечивают информационное взаимодействие через ЕИП НСУД

Цели эксперимента НСУД

ЦИК
«Регистр избирателей»

Автоматизированное формирование регистра избирателей

ФНС
ФИАС

Повышение качества данных Федеральной информационной адресной системы

Росстат
ЦАП

Создание Цифровой аналитической платформы статистических данных

Минэкономразвития
ГАС «Управление»

Формирование наборов сведений о муниципальных образованиях

Банк России
ЕСИА

Разработка инфраструктуры Цифрового профиля в части банковских услуг

Апробируемые механизмы

Использование доступности и качества данных:

- Работа с типом данных – реестр (мастер-данные)
- Интеграция с большим количеством источников
- Тестирование механизмов сопоставления данных
- Использование существующих механизмов (СМЭВ, качество)

Обеспечение качества данных:

- Точечное изменение НПА
- Внедрение дополнительных проверок (параметров качества)
- Изменение бизнес-процессов передачи данных (автоматизация / передача через регионы)

Обеспечение эффективного управления данными:

- Работа с типом данных – отчетность (агрегаты)
- Унификация данных
- Поиск эталонных источников
- Тестирование ЦАП как часть ЕИП НСУД

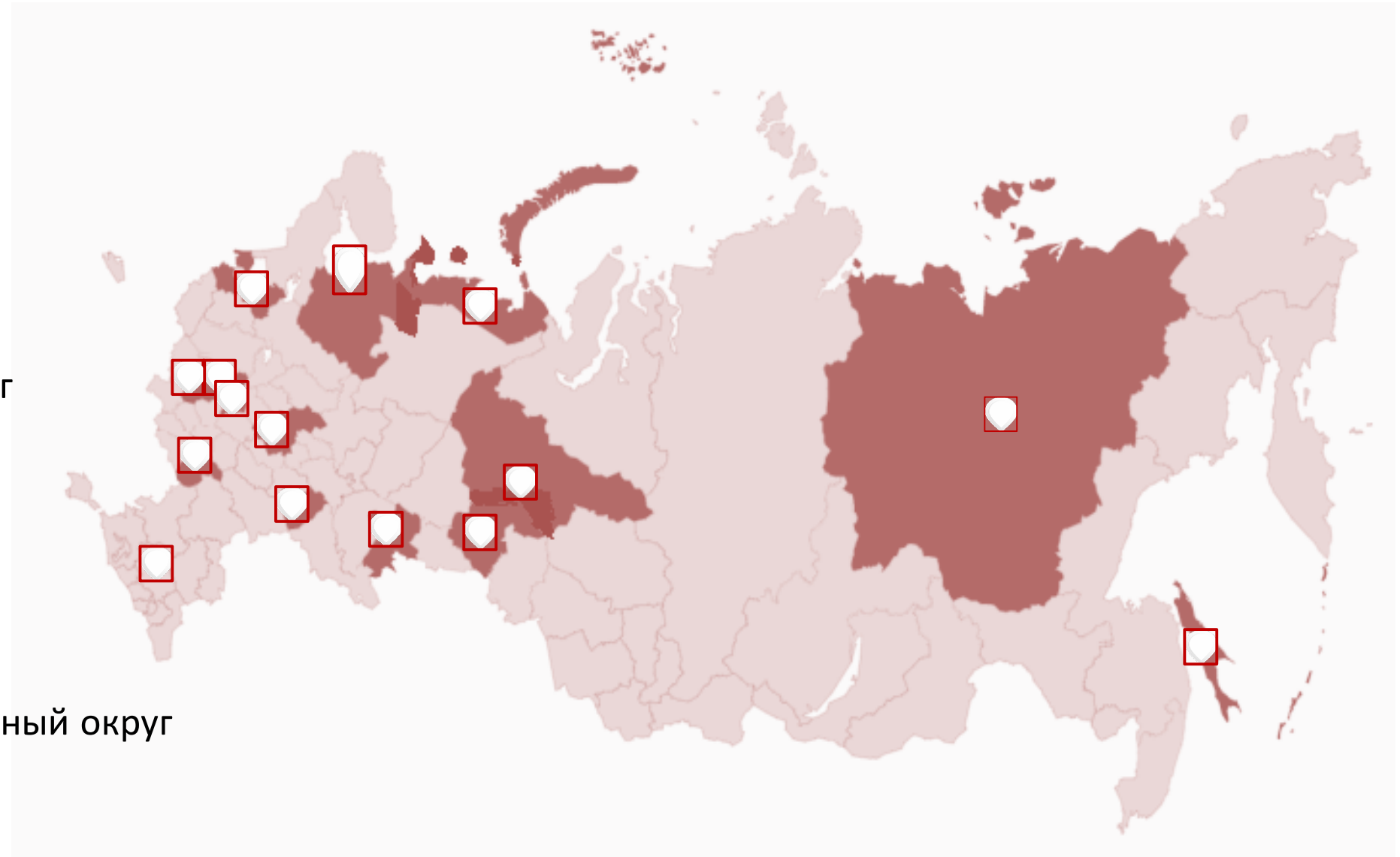
Обеспечение доступности данных:

- Поиск данных в существующих ИР
- Интеграция с большим количеством источников
- Создание нового информационного ресурса
- Тестирование ГАСУ как аналитического инструмента ЕИП НСУД

Обеспечение доступности и безопасности данных:

- Управление инцидентами (работа с гражданами)
- Разработка правовых механизмов доступа к данным через посредника (банки и страховые организации)
- Тестирование взаимодействия с бизнесом

1. Архангельская область
2. Воронежская область
3. Калужская область
4. Ленинградская область
5. Г. Москва
6. Московская область
7. Ненецкий автономный округ
8. Нижегородская область
9. Самарская область
10. Республика Саха (Якутия)
11. Сахалинская область
12. Ставропольский край
13. Тюменская область
14. Ханты-Мансийский автономный округ
15. Челябинская область



Методология (единые требования к управлению данными)



Управление собственными данными

Управление жизненным циклом данных

1. Описание ЖЦ
2. Определение источников данных

Управление качеством данных

1. Методика оценки качества
2. Мероприятия повышения качества данных

Управление изменениями данных

1. Регламентирование правил корректировки (исправление технических ошибок)

Управление ИТ структурой

1. Требование к ИС по ФЛК
2. Требования к хранению
3. Требования к защите

Управление кадрами

1. Организационная структура (ответственных за данные)
2. Обучение
3. Мотивация

Частично
рекомендатель-
ный характер

В обоих блоках
требований к
управлению
данными есть как
тех. требования, так
и методологические
требования

Модель информационного обмена

Требования к данным

1. Ведение в эл.форме
2. Доступность
3. Актуальность
4. Предоставление данных
5. Ответственность поставщика
6. Ответственность потребителя

Требования к ИР

1. Описать (паспортизация)
2. Модель данных ИР
3. Актуализация модели

Требования к видам данных

1. Определение видов данных
2. Определение идентификаторов
3. Гармонизация видов данных

Порядок взаимодействия

1. Подключение к НСУД
2. Предоставление через НСУД (Витрины, форматы и др. техническая составляющая)
3. Запрет на дублирование сбора данных

Разбор инцидентов

1. Право потребителя обратится за уточнением данных
2. Порядок разбора инцидентов
3. Уточнение данных в первоисточнике

Обязательный
характер

Подлежат первичному заполнению Паспорта ИР и включению в РВД



1. Информационные ресурсы, участвующие в эксперименте (ПП РФ от 03 июня 2019 года №710)
2. Информационные ресурсы, данные в которых наиболее часто запрашиваются через СМЭВ
3. Информационные ресурсы подлежащие переводу на реестровую модель
4. Информационные ресурсы необходимые для суперсервисов
5. Информационные ресурсы, обрабатывающие данные участвующие в системе межведомственного электронного взаимодействия (в том числе ПП РФ от 15 августа 2012 года №1471-р)
6. Базовые государственные информационные ресурсы (ПП РФ от 01 июня 2016 года № 487)
7. Информационные ресурсы специальной деятельности ФОИВ (на основе планов информатизации)

А

60 информационных ресурсов содержащих сведения о физ.лицах, юр.лицах, имуществе, правах и обязанностях и пр. (ЕГРЮЛ, ЕГРИП, ЕГР Налогоплательщиков, ЕГРН, ГИС «МИР», Перс.учет ОМС, Перс.учет ПФР, Перс.учет ФСС и др)

В

20 общероссийских справочников и классификаторов (ОКВЭД, ОКТМО, ОКОПФ, ОКОГУ, ОСМ и др.)

Инфраструктура электронного правительства (ИЭП) – это инфраструктура, обеспечивающая информационно-технологическое взаимодействие информационных систем, используемых для предоставления государственных (муниципальных) услуг, исполнения государственных и муниципальных функций.

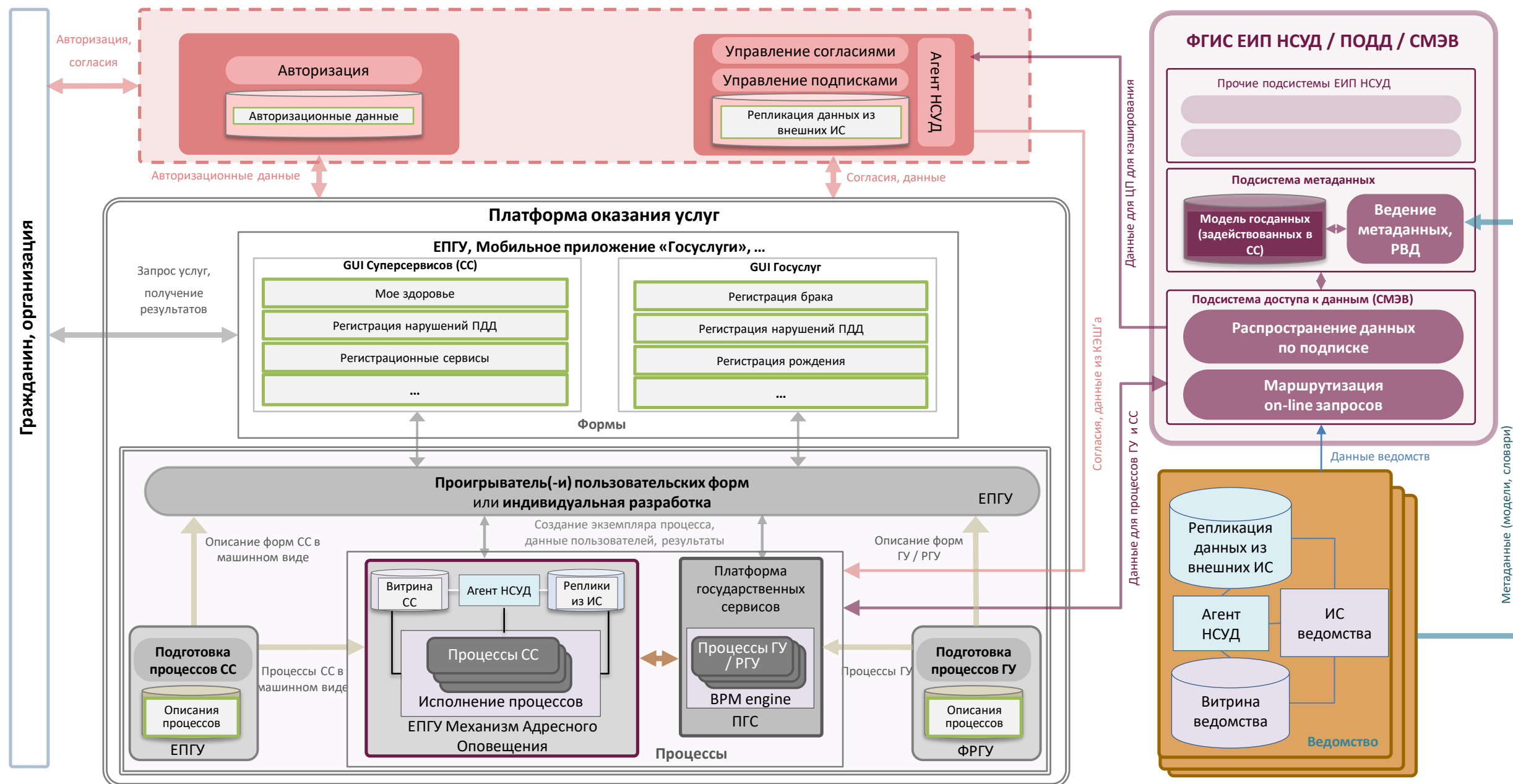
Интеграция НСУД с существующими элементами ИЭП:

- **ЕСИА** – планируется использовать для идентификации и аутентификации пользователей информационных продуктов НСУД
- **ЕПГУ** – интеграция не планируется
- **ФРГУ** – должна стать средой проектирования услуг, основанной на реестре видов данных ЕИП НСУД
- **СМЭВ** – планируется использовать для реализации интеграций с информационными системами ОГВ
- **ЕС НСИ** – должна стать частью НСУД и обеспечивать централизованное ведение и распространение данных общероссийских справочников и классификаторов

Интеграция НСУД с планируемыми в рамках ЦГУ элементами:

- **ЦП** (Цифровой профиль) – должен стать витриной, основанной на реестре видов данных
- **ЦАП** (Цифровая аналитическая платформа) – часть НСУД
- **ЕФИР о населении** – ИС–поставщик эталонных данных
- **ГАС «Управление»** – Аналитическая система, работающая на данных, получаемых через ЕИП НСУД

Архитектура суперсервиса (с использованием НСУД)



Обязательные к использованию:

1. **Реестр видов данных** – узнать, где/какие данные есть для межведомственного обмена
2. **Онтология / модель данных** – узнать, где/какие данные есть в принципе
3. **Реестр правил контроля качества данных (ККД)** – обеспечить исполнение и дополнить правила ККД поставщика
4. **Агрегатор данных** – сформировать простые запросы, провести обезличивание
5. **База знаний (единые требования к управлению данными)** – узнать, как правильно работать с данными
6. **Управление инцидентами**
7. **«Ядро» ЕИП НСУД** – обеспечение ведения и исполнения продуктов (п.1-6)
8. **Витрины данных ЕИП НСУД** – хранение полученных данных, использование в соот. с НПА и методиками
9. **Обучающие курсы**

Предлагаемые к использованию:

1. **Обследование** (ИР, процессов, данных)
2. **Лаборатория данных** – сформировать сложные запросы
3. **Маркетплейс** – поделиться данными с бизнесом

Модели обмена данными в международной практике



АНАЛИТИЧЕСКИЙ ЦЕНТР
ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ

	Наименование	Ядро (домены)	Область применения
1.	National Information Exchange Model (NIEM)	Деятельность; адрес; случай; дата и время; документ; пункт; событие; место; организация; Лицо	США. Биометрия. Химические, биологические, радиационные, ядерные угрозы. Детство, юность, семья. Чрезвычайные ситуации. Сфера услуг. Иммиграция. Защита инфраструктуры. Разведка. Международная торговля. Правосудие. Морской транспорт. Воен. операции. Скрининг. Наземные перевозки
2.	Реестр структур электронных документов и сведений в эл. виде (Модель общих процессов Евразийского экономического союза)	Документ, запрос, сведения, товар справочники, реестры, структура электронного документа	ЕАЭС. Таможенно-тарифное и нетарифное регулирование. Таможенное администрирование. Техническое регулирование. Здравоохранение. Санитарные, ветеринарные и фитосанитарные меры. Транспорт и перевозки. Подакцизная продукция. Интеллектуальная собственность. Фин. политика
3.	Aeronautical Information Exchange Model (Модель обмена аэронавигационной информацией)	Аэродром / Вертолетная площадка, зоны движения, службы, средства и т.д.; Структуры воздушного пространства; Организации и подразделения, услуги; процедуры; Маршруты; Летные ограничения	Международная практика. Аэронавигационная информация
4.	The Weather Information Exchange Models and Schema (Модели и схема обмена метеорологической информацией)	Сообщения о метеорологических данных, координаты	Международная практика. Метеорологическая информация
5.	Health Level 7 (ISO 13606)	Электронная медицинская карта	Международная практика. Здравоохранение