

CLEAN DATA



Российский разработчик программных продуктов для очистки, стандартизации и формирования эталонного профиля физических и юридических лиц в режиме реального времени.

Содержание

³ О компании

Гражданский фактор, модули:

^{4—9} Единая запись

^{10—12} Предиктивный ввод

^{13—14} Качество данных

^{15—20} Маскировщик

О компании Clean Data

ООО «Клин Дейта» создано в феврале 2022 года «АДС-Холдингом» и Дмитрием Журавлёвым (основателем HFLabs и DaData).

Основатели, руководство и сотрудники компании обладают многолетней экспертизой в сегменте Data Quality и опытом реализации крупных государственных и бизнес-проектов.

Текущие работы компании связаны с государственными проектами по витринам данных и информационным системам, запросами по качеству данных от федеральных и региональных государственных структур.

Clean Data входит в число компаний, аккредитованных Минцифры России. Продукты компании зарегистрированы в Реестре российского ПО.


Единая запись

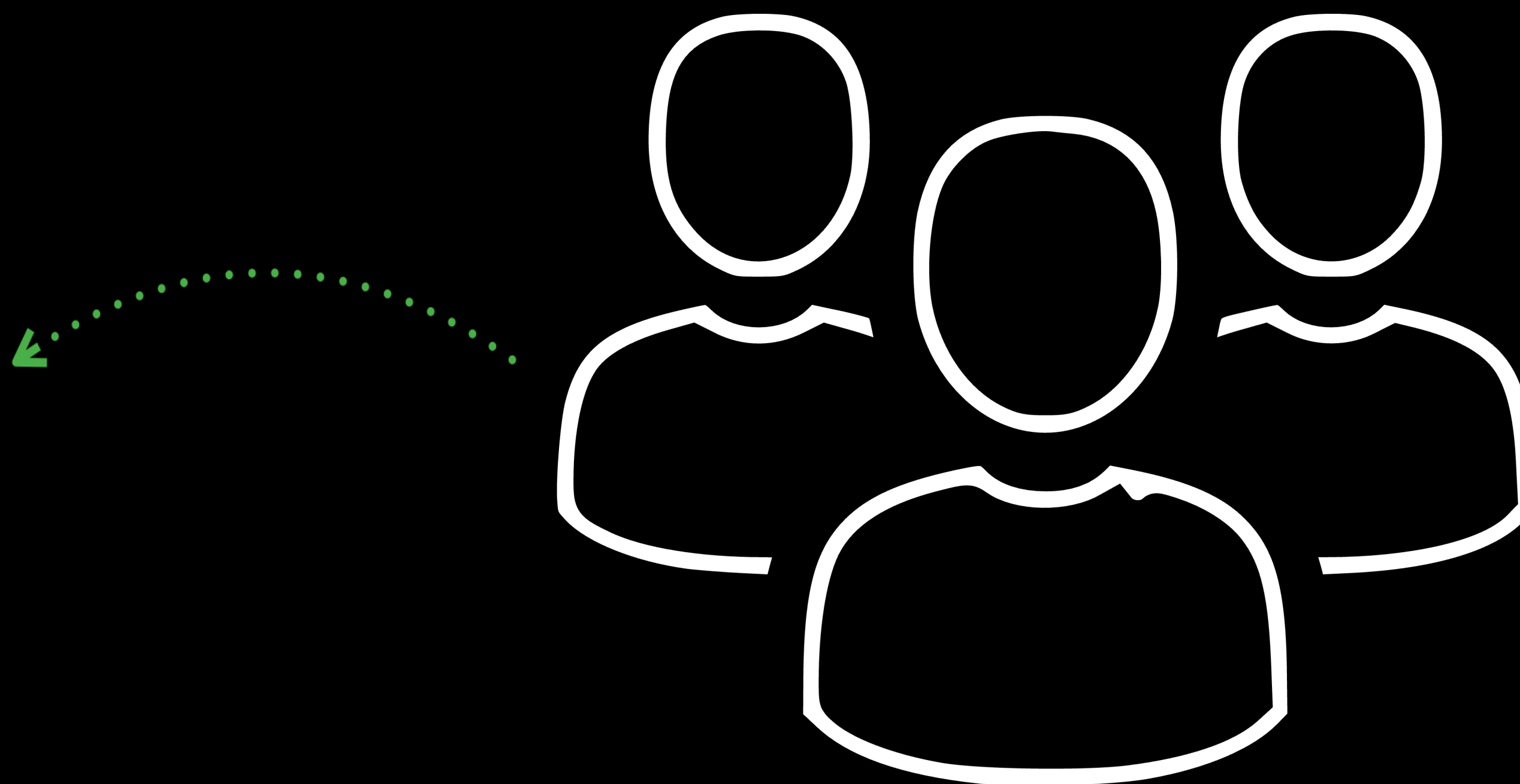
Преимущества решения:

- 85 сценариев доступны «из коробки»
- Учитывает похожесть
- Знает специфику отрасли
- Объединяет 100% дублей
- Решает конфликты
- Собирает всё лучшее
- За 0.2 секунды находит граждан в 50-миллионной базе
- За 1 секунду сохраняет все изменения, включая очистку, проверку на дубликаты и слияние
- Повышает качество сервиса за счёт моментальной идентификации гражданина во всех каналах обслуживания
- Происходит мгновенное оповещение всех систем об изменении данных, предотвращается создание дубликатов
- Real-time marketing, появляется возможность моментально собрать продуктовый профиль онлайн
- Защита от мошенничества без потери скорости обслуживания

Единая запись: основные функции

Создаёт эталонную запись с данными гражданина из всех доступных источников:

Сергей 
Владимирович
Иванов
31.07.1983
141983, Московская обл.
г. Дубна,
проезд Театральный 2-й
д. 6, кв. 17
+7 (49621) 3-56-03
+7 (916) 151-07-23
+7 (495) 737-12-93*101
ivanov_s@mycorp.ru



Единая запись: основные функции

Обозначает источники информации и присваивает коды качества:

Сергей	★
Владимирович	●●●
Иванов	●●●
31.07.1983	●
141983, Московская обл.	●●
г. Дубна,	●●
проезд Театральный 2-й	●
д. 6, кв. 17	●
+7 (49621) 3-56-03	●●
+7 (916) 151-07-23	●●
+7 (495) 737-12-93*101	●
ivanov_s@mycorp.ru	●



	Сергей В. Иванов 31.07.1983 Дубна 2-й Театральный 8(916)151-07-23 3-56-03
	ИВАНОВ СЕРГЕЙ 1-1-2016 096213-56-03 ivanov_s@mycorp.ru 3-56-03
	seregey vlodimirovich ivanov Дубна, Татральный 2, 617 0079161510723-моб 4957371293доб101

Единая запись: основные функции

Распознаёт пол и приводит информацию к единому виду:

Фамилия	★
Иванов	
Имя	
Сергей	
Отчество	
Владимирович	
Пол	
мужской	



**seregey
vlodimirovich
ivanov**

Единая запись: основные функции

Распознаёт адрес и обогащает запись дополнительными данными из справочников:

Индекс	Регион	Тип региона
141983	Свердловская	область
Тип города	Город	
город	Екатеринбург	
Тип улицы	Улица	
проспект	Орджоникидзе	
Дом	Квартира	Площадь
2	18	75,5 м ²

ЕКБ
Орджоникидзе,
12, кв 18



Единая запись: основные функции

Распознаёт и стандартизирует номера телефонов:

Домашний

+7 (49621) 3-56-03

Мобильный

+7 (916) 151-07-23


Рабочий

+7 (495) 737-12-93*101

Часовой пояс

Московское время (МСК)

**3-56-03,
0079161510723-
моб,495737129
Здоб101**



Предиктивный ввод

При ручном вводе данных предлагает корректные варианты написания и обогащает введённые данные:

- ФИО
- E-mail-адресов
- Почтовых адресов
- ИНН, ОГРН, КПП и других реквизитов

Предиктивный ввод также исправляет опечатки, определяет, если информация вносится транслитом или используется некорректная раскладка клавиатуры

Предиктивный ввод

Примеры работы:

Предиктивный ввод

Адрес Адрес по полям Организация Банк ФИО

Город, улица, дом, квартира

Choose a suggestion or continue typing

- Russia, Moscow city, ulitsa **Udaltsova**
- Russia, Moscow city, poselenie Sosenskoe, derevnya Bachurino, ulitsa **U**
- Russia, Saint-Petersburg city, prospekt **Udarnikov**
- Russia, respublika Sakha /Yakutiya/, ulus Mirninsky, gorod **Udachny**
- Russia, gorod Novosibirsk, ulitsa **Udarnaya**
- Russia, gorod Novosibirsk, ulitsa **Udarnikov**

Подсказка верных значений

Адрес Адрес по полям Организация Банк ФИО

Индекс	определяется автоматически
Регион / район	Респ Саха /Якутия/, Мирнинский улус
Город / н.п.	
Улица	
Дом	13
Квартира	г Удачный, мкр Новый город, д 13 г Мирный, пер. Звездная, стр 13

Предиктивный ввод

Результаты:

- Ввод данных осуществляется быстрее на 40%
- Сокращается количество ошибок, допускаемых пользователями
- Ввод адресов становится простой задачей

Качество данных

Распознаёт пол и приводит информацию к единому виду:

Фамилия	★
Иванов	
Имя	
Сергей	
Отчество	
Владимирович	
Пол	
мужской	



**seregey
vlodimirovich
ivanov**

Качество данных

Распознаёт адрес и обогащает запись дополнительными данными из справочников:

Индекс	Регион	Тип региона
141983	Свердловская	область
Тип города	Город	
город	Екатеринбург	
Тип улицы	Улица	
проспект	Орджоникидзе	
Дом	Квартира	Площадь
2	18	75,5 м ²

ЕКБ
Орджоникидзе,
12, кв 18



Маскировщик

Маскирует исходные данные без потери смысла

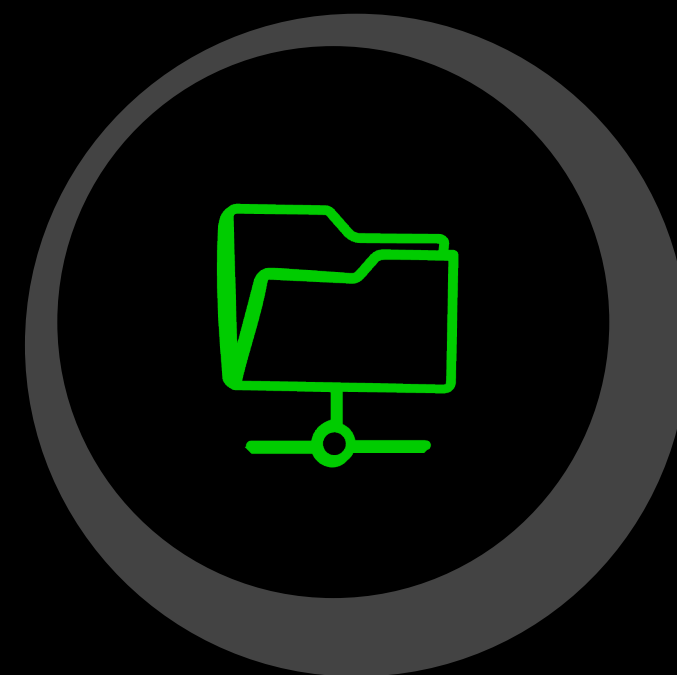


Ведомства хотят избежать утечек персональных данных граждан, поэтому ограничивают доступ сотрудников к персональным данным

Маскировщик

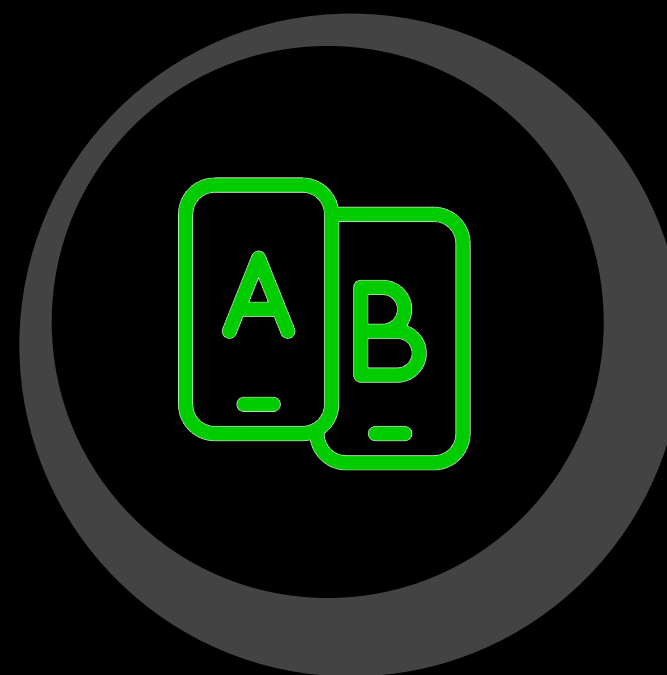
Позволяет решать типовые задачи:

Доступ к данным



Доступ к данным на боевых средах находится под строгим контролем, тогда как доступ к тестовым средам открыт большому количеству сотрудников

Тестирование



Реальные системы требуют тщательного тестирования. Чтобы делать его качественно, необходимо проверять на очень похожих данных

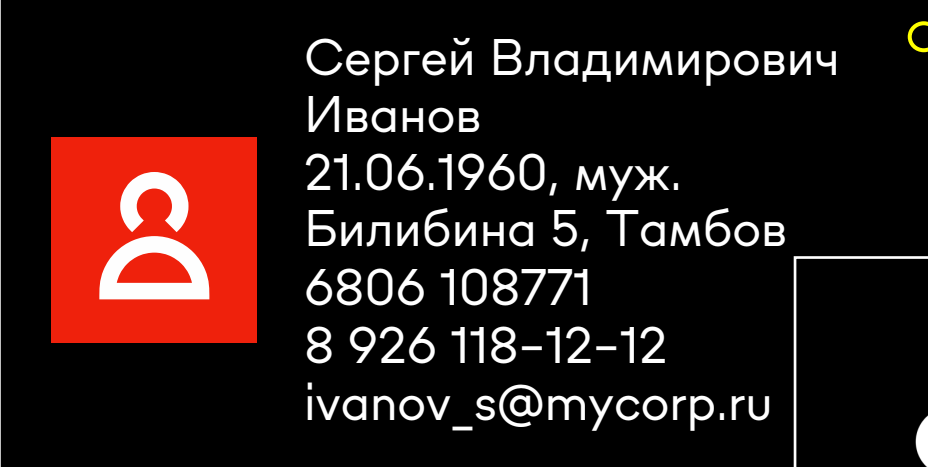
Построение моделей



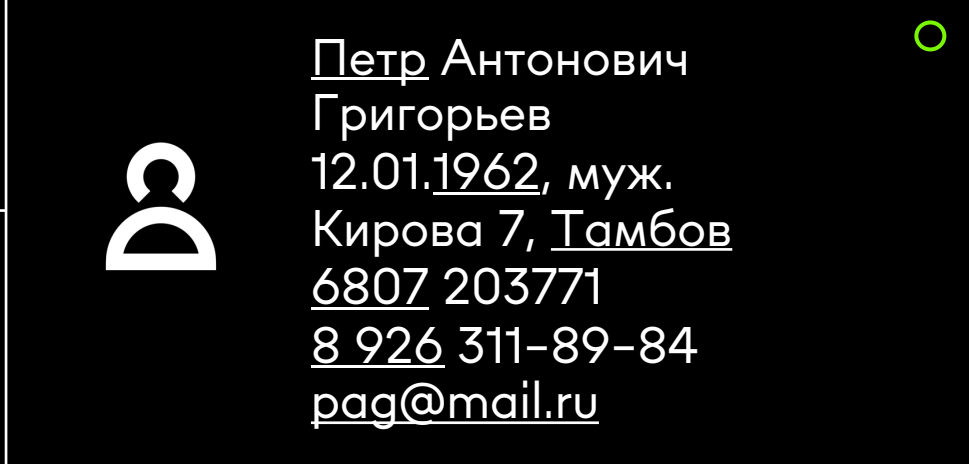
Для построения аналитических моделей (в том числе, с привлечением подрядчиков), также необходим доступ к данным

Как работает Маскировщик?

- Сохранит родственные связи. Если Сергея превратил в Петра, то Сергеевичей — в Петровичей
- Сохранит гендерный баланс
- Сохранит социально-демографические группы
- В адресе сохранит страну, город и район
- Понимает домены e-mail: рабочие, личные, одноразовые такими и останутся
- Оставит оператора и страну для номера телефона
- Серию паспорта привяжет к году рождения, чтобы она прошла форматно-логический контроль



Сергей Владимирович
Иванов
21.06.1960, муж.
Билибина 5, Тамбов
6806 108771
8 926 118-12-12
ivanov_s@mycorp.ru



Петр Антонович
Григорьев
12.01.1962, муж.
Кирова 7, Тамбов
6807 203771
8 926 311-89-84
pag@mail.ru

Как работает Маскировщик?



«Маскировщик» понимает, что Ивановы Сергей Владимирович и Николай Сергеевич, живущие по адресу город Тамбов, улица Билибина, 5 — родственники. И учитывает эту информацию, когда маскирует их ФИО и адрес.

Как работает «Маскировщик»?

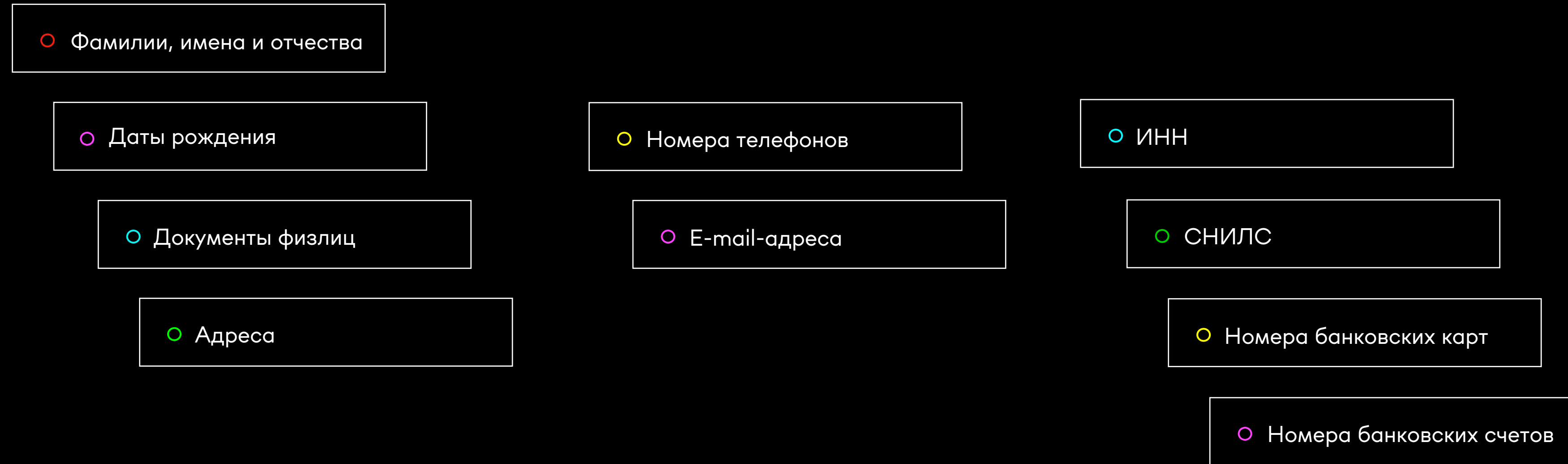
Из базы данных, содержащих записи о миллионах реальных людей, «Маскировщик» делает ровно такую же, неотличимую для человека базу данных людей, только нереальных

Сохраняет:

- Родственные связи
- Качество данных
- Социально-демографические признаки
- Географическое распределение

В каждой сессии «Маскировщик» подбирает замены случайным образом. Данные в файле замен хранятся в виде хеша, набора цифр и букв. Не зная алгоритма, восстановить исходные персональные данные невозможно. По умолчанию кэш сессии удаляется по её завершению.

Как работает «Маскировщик»?



- Отличает типы данных и понимает их структуру, качество данных не портится
- Под капотом уникальные алгоритмы
- Перед маскированием данные разбираются из строки, понимаем, где и что должно быть

Основание для использования модулей ПО Гражданский фактор

- Использование модулей ПО Гражданский фактор обеспечит единую технологию, стандарты, справочники и правила в области управления качеством данных для всех типов государственных организаций, что сократит издержки Государства
- Качество данных имеет высокий приоритет, так как напрямую влияет на адресность оказываемых государственных услуг и сервисов, повышает скорость выполнения запросов граждан и бизнеса
- ПО Гражданский фактор уже применяется в составе сервисов ЕПГУ, ФГИС ЕГРН, ИС КНД для проверки качества данных о гражданах и организациях, решения задач стандартизации и консолидации данных, формирования единых записей

Спасибо за внимание



cleandata.ru