

Наполнение государственных реестров и информационных систем

Новые потребности и решения
в обработке документов и данных



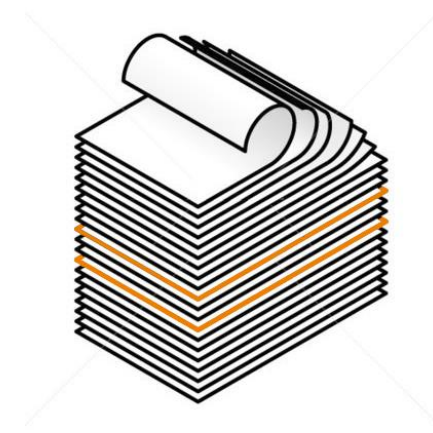
Веригин Илья, Руководитель направления
аналитики и работы с госсектором

Два пути развития



Реестры и регистры

- Подготовка большого объема данных
- Требование достоверности и полноты
- Необходимость интеграции реестров и совместимости данных



Информационные системы

- Много связанных документов и записей БД
- Необходимость поиска не только данных, но и документов по контексту
- Необходимость составления пакетов документации

Наполнение реестров

Общая задача для всех федеральных и региональных регистров и базовых информационных ресурсов

Пример: ФГИС «ЕГР ЗАГС»

Актные записи - сложнейшие рукописные документы с исправлениями и часто в ветхом состоянии

- как извлечь?
- как обеспечить аутентичность?

Сведения регионального уровня.

Регионы уже почти 10 лет активно переводят архивы ЗАГС в электронный вид.



Способы наполнения



Прямой ввод

Плюс – возможность привлечения местных операторов

Недостаток 1 – Ошибки

- Решается разработкой АРМ ввода (справочники, исторические справочники, шаблоны данных, обучение операторов)

Недостаток 2 – Ограниченные объемы

- Только для текущего ввода.



Аутсорсинг

Сканирование и обработка у исполнителя, ввод данных в систему заказчика

Преимущества:

- Обработка больших объемов
- Заданное качество данных

Основная технология аутсорсинга

Экономичный способ обработки больших массивов, а также конфиденциальных документов



Каждый оператор вводит свой пул фрагментов (эффективность конвейера)

Центр индексирования
Программное и ручное извлечение данных
(несколько тысяч операторов)

Итоговая база данных собирается также на территории заказчика

Сотрудники ЭЛАР связанных данных не видят



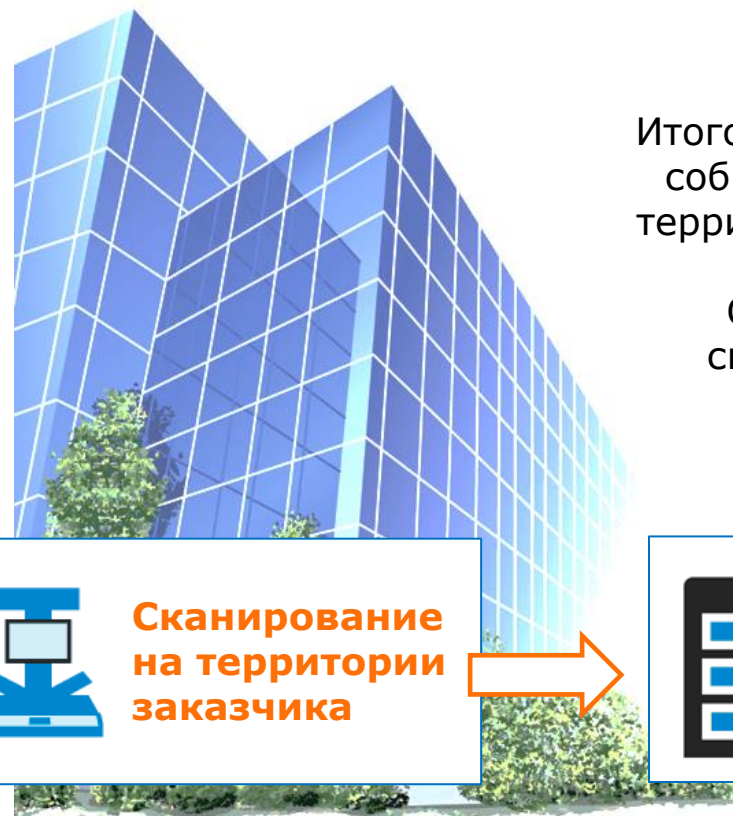
Сканирование на территории заказчика



Автоматическая разрезка изображений на обезличенные фрагменты

фамилия	АЛЕКСАНДРОВА
имя	АЛЕКСАНДРА
отчество	
дата рождения	02.03.1987

ID - ключи фрагментов кодирования



Технология позволяет

Создавать **большие электронные информационные ресурсы:**

- **В заданный срок**
- **С контролем качества**
- **С экономией** (глубокая автоматизация, недорогие операторы ручного ввода)



Проводить **массовую оцифровку конфиденциальных документов** до уровня «СовСекретно» (есть лицензии ФСБ, ФСТЭК)

Примеры:

Оперативно-справочные картотеки МВД

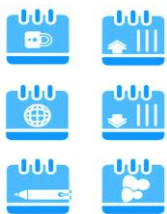
Кадровые дела Минобороны

ЗАГС

Техническая документация Росатома

...

Задачи наполнения государственных ИС



Документы хранятся в разных подразделениях (бумага) и системах (электронные)

Это приводит к невозможности собрать комплект даже вручную



Многие документы регистрируются по реквизитам, а значимые данные – в текстах

Во многих ИС необходимо заполнение расширенных полей в карточках

Для поиска надо извлекать информацию из текстов



Часть документов выпадает из поиска из-за отсутствия заданного реквизита

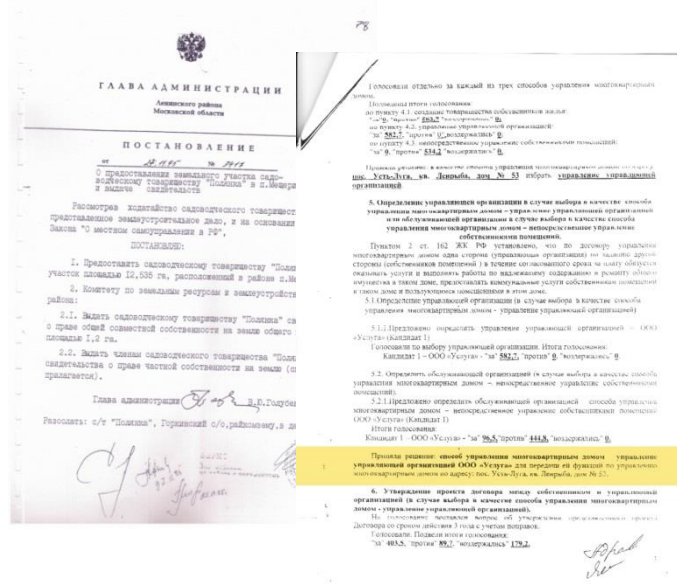
Надо готовить контент – объединять документы в дела или строить гиперссылки



Полнотекстовое распознавание приводит к релевантным спискам (как в Яндексе)

Слишком большая выборка = необходимость вручную искать данные

Пример. Управление землей и имуществом



Обеспечение доходности бюджетов требует качественного и полного учета – наполнения систем

Дела по объектам как таковые не ведутся.

Классическая ситуация – минимум 3 отдельных архива:

- Правоустанавливающие документы, Постановления
- Договоры аренды/соцнайма
- Обращения и переписка

Проблема: ФИО и адреса всегда в текстах (в приложениях)

Поиск работником адреса или персоналии в тексте занимает много времени

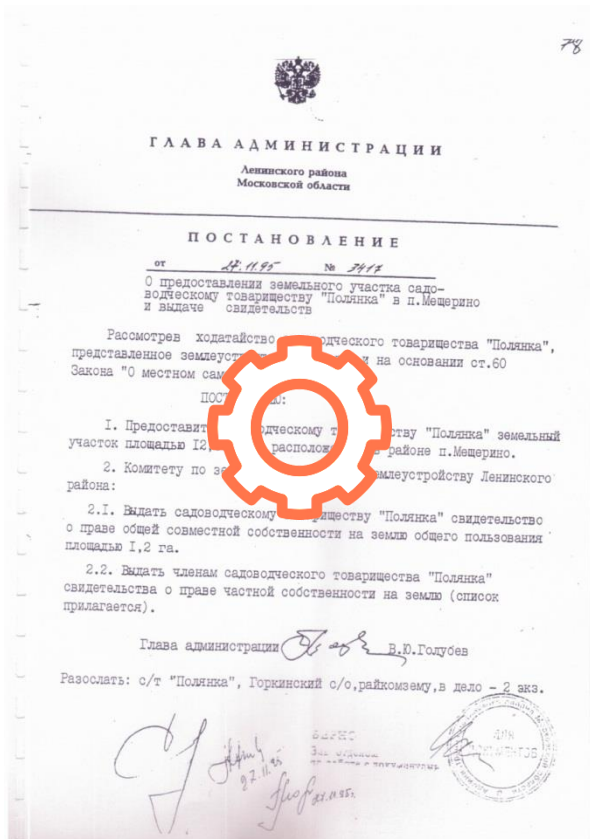
Полный комплект часто так и не находится

Потребность – не наводить порядок в бумаге (дорого и неэффективно), а объединить документы в комплекты в электронном виде:

- Отсканировать архивы
- Связать документы с объектами/субъектами
- Загрузить ресурс в учетную систему (в т.ч. сформировать слой в геоинформационной системе)

Интеллектуальное индексирование

ЭЛАР с 2010 года применяет программные алгоритмы выделения значимых сущностей
Имеются настройки для типовых документных массивов



Наименования юрлиц, поселений и адреса



Выделение всех ФИО собственников
на последующих страницах приложений

**Технология = повышение качества
(программа не пропускает данные)**

Задача нормализации

Адрес – до 22 вариантов написания



ФИО – более 7 вариантов написания

Иванов Александр Юрьевич
Александр Юрьевич Иванов
А.Ю. Иванов
А.Ю.Иванов
А. Ю. Иванов
Иванов А.Ю.
Иванов Александр
Александр Иванов

Качественные связь и поиск информации возможны только после приведения ее к единому формату написания.

Услуги обязательно включают программную и ручную нормализацию.

Сложный проект - КрымРеестр

Для внесения сведений в ГКН были оцифрованы более 900 000 дел по земельным участкам.

При обработке вводились украинские адреса из документов, которые в последствии были приведены к российскому адресу объекта в ФИАС.

Технология в составе систем

Единая информационно-справочная система Правового департамента Министерства обороны РФ

- Встроенная технология интеллектуальной обработки
- > 60 000 приказов уже в 2016 г.

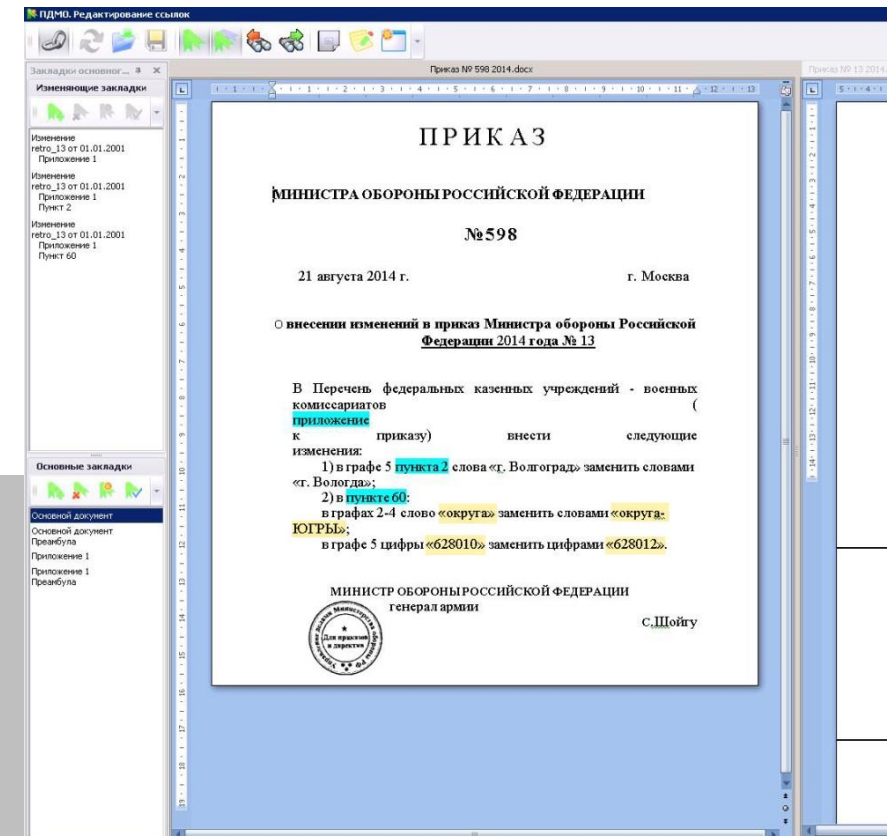
Система умеет: при вводе нового документа самостоятельно вносить изменения в существующие и создавать гиперссылки

Система обеспечивает: работу всех воинских подразделений с актуальными документами на любую дату

Гиперссылки – дополнительный, часто необходимый механизм качественного поиска. Применим в т.ч для графических файлов.



АРМЮРИСТ



Проектный подход ЭЛАР

Большинство задач индивидуальны
(срок, бюджет, объем и условия работы ...)

Мы постоянно разрабатываем новые технологии
и оказываем комплекс услуг, чтобы справиться
с конкретными задачами заказчиков.

Кейс

- **Обработать и связать 150 000 документов гиперссылками – ОК**
- **Хранить и сделать их доступными для нескольких ГИС - ?**



- **Проще/дешевле не дорабатывать систему**
- **Сделать интегрированный электронный архив с поддержкой гиперссылок и контекстного поиска**



СЭА Администрации
Президента и Правительства
Республики Саха (Якутия)

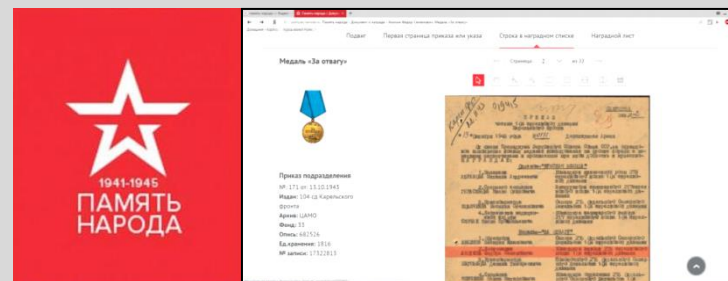
ОРД, архивы, ЗАГС



Электронный архив Министерства труда

Вся входящая/исходящая и
внутриведомственная
корреспонденция.

Интеграция с УКИД, МЭДО, СМЭВ



Более 100 млн записей и документов
Великой Отечественной войны

Спасибо за внимание!



Корпорация ЭЛАР

☎ +7 (495) 274-31-31

✉ office@elar.ru

Посетите наши ресурсы:

www.elar.ru www.wescan.ru

www.saperion.ru