

# **Классические** библиотечные направления для цифровой трансформации

Вадим Дуда  
Российская государственная библиотека  
Рязань, март 2023 г.

# Направления

- **Каталогизация, СБА/СБО.** Искусственный интеллект (ИИ), обработка и раскрытие фонда в цифровой среде.
- **Комплектование.** Цифровые СМИ.
- **Консервация и реставрация.** Большие данные и НЭБ.

# **Каталогизация и СБА/СБО.**

ИИ, обработка и раскрытие  
фонда в цифровой среде

# Ленинка сегодня

- Фонд около 120 млн + 75 млн документов
- Поглощение Книжной палаты в 2022 году
- Оператор национального проекта “Культура”
- 1 млн посетителей в год “ногами” и 30 млн “пальцами”
- 134 тыс квадратных метров на Воздвиженке
- Оператор Национальной электронной библиотеки (НЭБ)

# Масштаб производства

- **Поступления/обработка (77-фз, включая ЭЭ)**
  - 100 тыс книг
  - 250 000 газет / 60 000 журналов
  - 10 000 диссертаций
- **Выдача документов**
  - 2-3 млн в читальные залы
  - 35-40 млн в электронном виде

# Обработка документов

- **Описательная каталогизация** (формальные поля)
- **Содержательная каталогизация**  
(систематизация, предметизация, ключевые слова и события, ... онтология ББК / УДК /...)
- На выходе - **библиографическая запись** и справочно-библиографический аппарат для **навигации и поиска**

# Будущее

- Необходим новый уровень обработки текущих поступлений и ретроспективного фонда Ленинки. **Содержательная каталогизация** (индексирование?) на основе нейросети / ИИ с использованием полных текстов документов, автоматическая обработка.
- Расширение типов документов (**цифровые СМИ**)
- Поиск и навигация по фонду - на основе автоматической **содержательной каталогизации** полных текстов с использованием лучшего опыта ИТ лидеров
- Выдача документов в результате поиска **в читальных залах** (Ленинка + библиотечная сеть) **и через цифровые платформы** (привычные паттерны цифрового поведения)

# Национальное пространство знаний

- **Технологический партнер (ИИ компетенции)**
  - Технологии автоматизированной содержательной обработки потока цифровых документов
  - Поисковый механизм и “библиографическая справка” оффлайн и онлайн
- **Ленинка / НЭБ / Библиотечная сеть**
  - Фонды и оцифровка
  - Текущие поступления (фиды достоверного контента)
  - Добавленная интеллектуальная стоимость обработки документов
  - Государственная политика доступа в бумажном и цифровом виде



# ЧТО ПОЛУЧИМ В ИТОГЕ

- **Государство**

- Введение в оборот широкого массива достоверных документов, создание российского пространства знаний, развитие “человеческого капитала”

- **Люди**

- Доступ к широкому массиву документов в уже привычной цифровой среде для интеллектуальной работы

- **Экономика**

- Большие данные и цифровые массивы информации
- Книжная и интеллектуальная собственность
- Развитие экономики знаний
- Человеческий капитал

# **Комплектование** Цифровые СМИ

# Суть проекта

- Библиотека “будущих поколений”
- Системная работа по отбору, систематизации, хранению и организации доступа к достоверному массиву цифровых документов (веб-сайты, новостные СМИ, соцсети и т.д.)

# Отличия и приоритеты

- **Главный приоритет - полнота, достоверность, долгосрочное хранение важной для ГОСУДАРСТВА цифровой информации**
- Политика комплектования (на основе “ФЗ-77”) - **ГОС**
- Отбор (+ЯНДЕКС, ВК, МТС, другие игроки)
- Хранение - **ГОС**
- Навигация и продвижение (+ЯНДЕКС, ВК, МТС, другие игроки)
- Государственная политика доступа - **ГОС**
- Нормативно-правовое обеспечение - **ГОС**

# Отбор

- Автоматический
- Ручной
- Библиотечная/Корпоративная сеть
- Краудсорсинг

# Комплектование

- **Долгосрочное** (ядро, постоянный репертуар)
- **Краткосрочное** (события)
- **Приоритетное** (горячие темы)

# **Консервация и реставрация.**

Большие данные и НЭБ