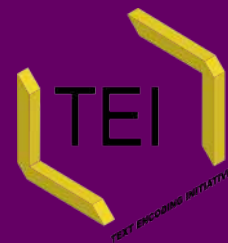


# Разметка документов XML-TEI с помощью Oxygen



А.М. Лаврентьев

Лаборатория ICAR – Национальный центр  
научных исследований Франции и Лионский  
университет

EI'Manuscript 2012

Петрозаводск, 4.09.2012

# Цель занятия

- Подготовить несколько файлов XML-TEI для последующей работы с платформой TXM
- Вы можете
  - ☐ использовать подготовленные файлы из папки
    - txmSandbox\tei
  - ☐ использовать ваш собственный текст
    - .doc, .odt, .rtf, .txt
  - ☐ скачать текст из Интернета
    - вышеперечисленные форматы или .html, .xhtml

# Работа с подготовленными файлами

- Запускаем Oхугеn
- Открываем файлы
  - Pushkin\_KapitanskayaDochka.xml
  - Bunin\_Baybaki.xml
- Изучаем структуру каждого документа
- Для комфортного чтения текста пользуемся Author mode
  - кнопка внизу закладки с документом

# Работа с подготовленными файлами

- Чем отличается структура одного документа от другого?
- Как можно улучшить разметку?
  - Байбаки
    - разметить номера страниц `<pb n="xx"/>`
    - убрать лишние абзацы
    - структурировать разделы
  - Капитанская дочка
    - разметить стихотворные строки `<lg> / </>`
  - Везде
    - прямая речь `<q>`, имена собственные `<name>`

# Работа с подготовленными файлами

## ■ Заполняем teiHeader

- добавляем информацию об ответственности

```
<respStmt>  
  <resp>Разметка XML-TEI</resp>  
  <name xml:id="AL">Алексей Лаврентьев</name>  
</respStmt>
```

- добавляем запись об изменении

```
<revisionDesc>  
  <change who="#AL" when="2012-09-04">Обогащение разметки</change>  
</revisionDesc>
```

# Работа с собственными файлами

- Для первичной конвертации документа воспользуемся сервисом
  - <http://www.tei-c.org/oxgarage/>
- Конвертируем в
  - TEI P5 XML Document
- Смотрим, что получилось
- Улучшаем разметку, как в предыдущих файлах



# Спасибо!