

# Центр веб-исследований посредством смарт-статистики

## Акцент на данных из объявлений о работе в Интернете

Большие данные в контексте аналитики рынка  
труда

Training programme ETF – Make it Match Network

Fernando Reis

22, 23, 24 November 2021

# Интернет в качестве источника статистических данных

- Веб-скрапинг – это легко, но...
- Вы хотите, чтобы процесс был:
  - **Автоматизированным**
  - **Методологически правильным**
  - **Устойчивым**
  - **Прозрачным**
  - **Воспроизводимым**
  - **Единообразным**
  - **Эффективным**
  - **Последовательным**

# Интернет в качестве источника статистических данных

- Веб-скрапинг – это легко, но...
- Сбор официальной статистики – это сложный процесс!
- Центр веб-исследований – наш инструмент, который позаботится о сложной части процесса.

# Факты о многопользовательской системе веб-исследований

- Сбор веб-данных в рамках составления статистики – непростой процесс (например, соглашения об обмене данными)
- Необходима инфраструктура с возможностями обработки больших данных
- Необходимы специальные навыки
- Расширение Европейской статистической системы (ЕСС) через возможности веб-исследований займет много времени
- Европейская система по объявлениям о работе в Интернете всё равно продолжит существовать

# Краткая история исследований больших данных официальной статистики

- 2013 год: Схевенингенский меморандум по большим данным
  - **Проверка потенциальных источников Больших данных для официальной статистики**
- 2015 год: План работы с большими данными и перспективы развития
- 2016–2020 годы: проект ESSnet Big Data I
  - **Использование больших данных в тестовом режиме (включая объявления о работе в Интернете)**
- 2018 год: Бухарестский меморандум об официальной статистике в датифицированном обществе (Надежная смарт-статистика)
  - **Акцент на внедрении**



# От Больших данных к Надежной смарт-статистике

- Мы можем рассматривать Смарт-статистику как будущую систему официальной статистики, если **сбор, обработка и анализ данных будут внедрены в саму систему**, начиная с цифровых следов, оставленных в результате активности человека в Интернете
- **предполагается, что внедрение исследований во все стадии жизненного цикла данных** позволит Официальным статистическим учреждениям **сохранить и укрепить свою роль в качестве ключевых поставщиков данных в цифровом мире**



# Принципы Надежной смарт-статистики

- Сбор статистики из множества источников
- Многоцелевые источники данных
- Уровневая организация: модель «песочные часы»
- Методологическая основа по модульному принципу
- Вытеснение вычислений
- Использование данных без передачи

# Надежная смарт-статистика

Надежная смарт-статистика

Управление ЕСС

Центр веб-исследований

Вакансии  
в  
Интернете

Новые  
источник  
и данных  
Европейс  
кого  
реестра  
групп

...

Сообщество  
экспертов

Сообщество  
экспертов

Сообщество  
экспертов

Центр надежных смарт-  
исследований

Иследо  
вание  
времяза  
трат

Обследо  
вание  
бюджетов  
домашних  
хозяйств

...

Сообщество  
экспертов

Сообщество  
экспертов

Сообщество  
экспертов

Центр данных операторов  
мобильной связи

Демографи  
ческая  
статистика

Региональ  
ная  
статистика

...

Сообщество  
экспертов

Сообщество  
экспертов

Сообщество  
экспертов

...

...

...



# Центр веб-исследований

Вариант  
использования 1  
Навыки

Вариант  
использования 2  
SD4MNE

Новый  
вариант  
использования

Уровень  
статистики

Статистика  
по навыкам

Расширенная  
статистика  
по вакансиям

Расширенный  
Европейский  
реестр групп

Расширенн  
ая  
статистика  
цен

Уровень  
приложений

Объявления о работе в  
Интернете

Бизнес-данные  
транснационал  
ьных компаний

Онлайн-  
цены

Уровень  
услуг

Платформа веб-исследований

Уровень  
данных

Веб-данные



# Центр веб-исследований – Услуги

- Обеспечение поддержки партнерам ЕСС в следующем:
  - Сбор данных (веб-скрапинг, API)
  - Транснациональные соглашения об обмене данными
  - Модели партнерства для национальных соглашений об обмене данными
  - ИТ-инфраструктура и инструменты
  - Аналитические услуги (например, обработка естественного языка)
  - Методология
  - Особенности регулирования
  - Навыки (учебные материалы)
  - Сотрудничество в сфере научно-исследовательских работ
  - Управление

# Центр веб-исследований – Принципы

## Некоторые принципы

- Центр ЕСС
- Удовлетворение потребностей на национальном и общеевропейском уровнях
- Модульная структура
- Гарантия определенных процессов и продуктов
- Приоритет совместной работе, возможность индивидуальных действий
- Программы должны быть с открытым исходным кодом
- Максимально возможная прозрачность
- Процессы общего пользования должны быть утверждены и понятны
- Линия данных и процессов
- Промежуточные продукты могут использоваться всеми партнерами

## Перемещение больших данных в сферу внедрения

- ESSnet Big Data I: тестирование
- Необходимость смещения внимания на внедрение самых успешных тестовых проектов
- ESSnet Big Data II: новые тестирования + надежная смарт-статистика + внедрение
  - **Внедрение = разработка спецификаций для внедрения, экспериментальной статистики, рекомендаций по управлению данными / процессом**
  - Объявления о работе в Интернете
  - **Веб-сайты компаний**
  - **Интеллектуальные счетчики электроэнергии**



# Системы сбора данных по объявлениям о работе в Интернете

- Преимущественно национальный подход
  - **ESSnet Big Data**
- Европейский подход
  - **Генеральный директорат по коммуникационным сетям, контенту и технологиям**
  - **Европейский центр по развитию профессионального образования**

# Как создать центр веб-исследований

- Создать **платформу**
- Создать **сообщество**
- Разработать **методологии**
- Спроектировать **образовательные ресурсы**
- Обеспечить надлежащую **нормативную базу**
- Помнить о **коммуникации**
- Разъяснить **принципы управления**

## 6 шагов для создания Платформы веб-исследований

1. Создать ИТ-инфраструктуру для Платформы веб-исследований

**Сделано**

2. Развернуть систему по объявлениям о работе в Интернете Европейского центра по развитию профессионального образования в рамках ИТ-инфраструктуры Платформы веб-исследований

**Сделано**

3. Спроектировать архитектуру для возможности повторного использования Платформы веб-исследований

**Подготовлен проект, обсудить с участниками проекта**

4. Разработка компонентов Платформы веб-исследований (MVP)

**Начинается в текущее время**

5. Запустить новый(-е) вариант(ы) использования в рамках Платформы веб-исследований

**2021 год**

6. Перенос объявлений о работе в Интернете на Платформы веб-исследований

**2022 год**



# Как создать сообщество Платформы веб исследований

- **Сеть веб-исследований**
  - **Продвигать использование Центра веб-исследований в рамках составления официальной статистики**
  - **Поддерживать его методологическую разработку;**
  - **Материализовать роль Национальных статистических институтов в Центре веб-исследований:**
    - ✓ использовать службы Центра веб-исследований;
    - ✓ принимать участие в запуске вариантов использования;
    - ✓ вносить вклад в разработку компонентов;
- **Сеть веб-исследований запущена 1 апреля 2021 года**



# Как разработать методологии Центра веб-исследований

- Внедрить и в дальнейшем настраивать методологии, разработанные в контексте системы по объявлениям о работе в Интернете Европейского центра по развитию профессионального образования
- Собрать вклады сообщества по официальной статистике (Сеть веб-исследований)
- Разработать методологии для новых вариантов использования
- Разработать основу оценки качества в зависимости от ОС в отношении веб-данных



# Как создать образовательные ресурсы

- Обучающие курсы
  - Обучающая программа по европейской статистике – **курсы по веб-скрапину с 2015 года**
  - Продвинутый коучинг – **начало следующего года**
- Учебные материалы – доступны в Интернете
- Интерактивная среда
- Организовать семинары

# Как обеспечить надлежащую нормативную базу

- Внедрить политику ЕСС по извлечению веб-данных
- Определить потребности в сфере регулирования
- Негосударственное регулирование в отношении данных



**Благодарим за внимание**

**Фернандо Райс**

**Евростат**

 [fernando.reis@ec.europa.eu](mailto:fernando.reis@ec.europa.eu)

 <https://github.com/reisfe/>

 <https://twitter.com/reisfe/>

 <https://linkedin.com/in/reisfe/>