

Сервис контроля и архивации материалов портала открытых данных г. Москвы

[Обсудить в форуме](#) Комментариев — 1

Эта страница опубликована в основном списке статей сайта по адресу <http://gis-lab.info/qa/data-mos-control.html>

Описание сервиса контроля данных

Этот сервис не является сервисом Правительства Москвы или любых его подразделений.

Содержание

- [1 Проблема](#)
- [2 Идея](#)
- [3 Реализация](#)
- [4 Функции сервиса](#)
- [5 Планы на будущее](#)
- [6 Результаты](#)
 - [6.1 Код](#)

Проблема

Первым городом раскрывшим более-менее существенное количество открытых геоданных является Москва. Данные по городу выкладываются на [Портал открытых данных г. Москва](#). Портал неплохо справляется с функцией доставки данных потребителю, нет привязки к API, для получения наборов данных не нужна регистрация. Однако, некоторые другие функции выполняются не очень хорошо: нет четкой системы оповещений об обновлениях, непонятно что же вообще обновилось. Дополнительно, нет уверенности, что отдельные наборы данных завтра не пропадут с портала окончательно.

Поскольку данные открыты, задача общественности взять на себя часть функций, с которыми неоптимально справляется сам портал.

Идея

Создать независимый сервис архивации и проверки изменений в данных портала.

Реализация

Регулярно скачиваем перечень наборов данных и все существующие и новые наборы данных, сравниваем с предыдущими версиями, находим изменения, обо всех значительных - докладываем пользователям.

Функции сервиса

- Ежедневная проверка
- Обнаружение изменений комплектации данных:
 - Новые наборы данных
 - Скрытие из [общего перечня](#)) наборов данных, при сохранении доступности
 - Обнаружение восстановленных, после скрытия, наборов данных
 - Недоступность наборов данных (возможно удаление)

- Обнаружение восстановления после недоступности
- Обнаружение изменений данных:
 - Добавление/удаление атрибутивных полей
 - Добавление/удаление записей
 - Другие изменения содержания (не ведущие к изменению количества полей и/или строк)
 - Создание html-представлений построчных diff-файлов для наглядной оценки изменений
- Сохранение версий измененных наборов данных
- Компрессия данных
- Оповещение о главных событиях в [специальном твиттере](#) и [лог-файле](#).
- Регистрация изменений в специальных лог-файлах для каждого набора ([пример](#))

Планы на будущее

1. Интеграция с [процессом нормализации и конвертации данных](#) с портала в более приемлемый для ГИС и разработчиков вид. Вместо пакетной конвертации, как сейчас, будут конвертироваться только измененные наборы данных.
2. ~~Генерация красивых diff-ов для отображения изменений в данных.~~ Сделано

[Тикеты](#) в трекере проекта.

Результаты

Твиттер для оповещений об изменениях: <https://twitter.com/datamosru>

Архивированные данные: <http://gis-lab.info/data/mos.ru/data>

Код

Весь исходный код на языке Python доступен на [Github](#).

Использованы:

- [python-twitter](#)
- [bitly-api-python](#)
- [diff-match-patch](#) с [line-by-line](#)

[Обсудить в форуме](#) Комментариев — 1

Последнее обновление: 2014-07-20 13:36

Дата создания: 12.05.2013

Автор(ы): [Максим Дубинин](#)