

Применение опыта других стран в вопросе цифровизации и гармонизации данных контрольно- надзорной деятельности

УПРАВЛЕНИЕ ДАННЫМИ В ГОСУДАРСТВЕННЫХ И ВЕДОМСТВЕННЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ



МНОГИЕ ГИС И ВИС ИМЕЛИ ОДНУ И ТУ ЖЕ ПРОБЛЕМУ

- ⊙ Исходные данные частично классифицированы и не структурированы
- ⊙ Разработка информационной системы часто подразумевает научно-исследовательскую работу по разработке информационной модели данных, неструктурированных изначально
- ⊙ Исходные данные динамичны, структура Баз Данных нет, поэтому данные устаревают на момент их попадания в информационную систему
- ⊙ История изменения структуры данных не ведется
- ⊙ Отсутствие непротиворечивой информационной модели порождает сложности при обработке, фильтрации данных и их интерпретации

В итоге системы оперируют как правило устаревшими и плохо структурированными данными, что «сводит на нет» их эффективность и не позволяет решать поставленные перед ними задачи.

МИРОВОЙ ОПЫТ



Translational Medicine Ontology (TMO) и Translational Medicine Knowledge Base (TMKB)

Отказ от стандартного подхода, заключенного в предварительной классификации всех данных и реализации на основе классификаторов реляционной (табличной) модели объектов, в сторону семантических технологий и не реляционных (объектных) Баз Данных. Разработка онтологии (набора семантических моделей) в области медицины (TMO) и Базы Знаний на ее основе (TKMB).

Medical Ontology Research

"The LHNBCB Medical Ontology Research project encompasses basic research on biomedical terminologies and ontologies and their applications to natural language processing, clinical decision support, translational medicine, data integration and interoperability."

Tools & Resources

- **RxNav**: Navigating drug information sources
- **SemNav**: UMLS[®] Semantic Navigator (UMLS Terminology Services [UTS] account required)
- Map to specific UMLS vocabularies
- SNOMED CT Status API
- LOINC Mapto API
- RxImage API

Publications & Presentations

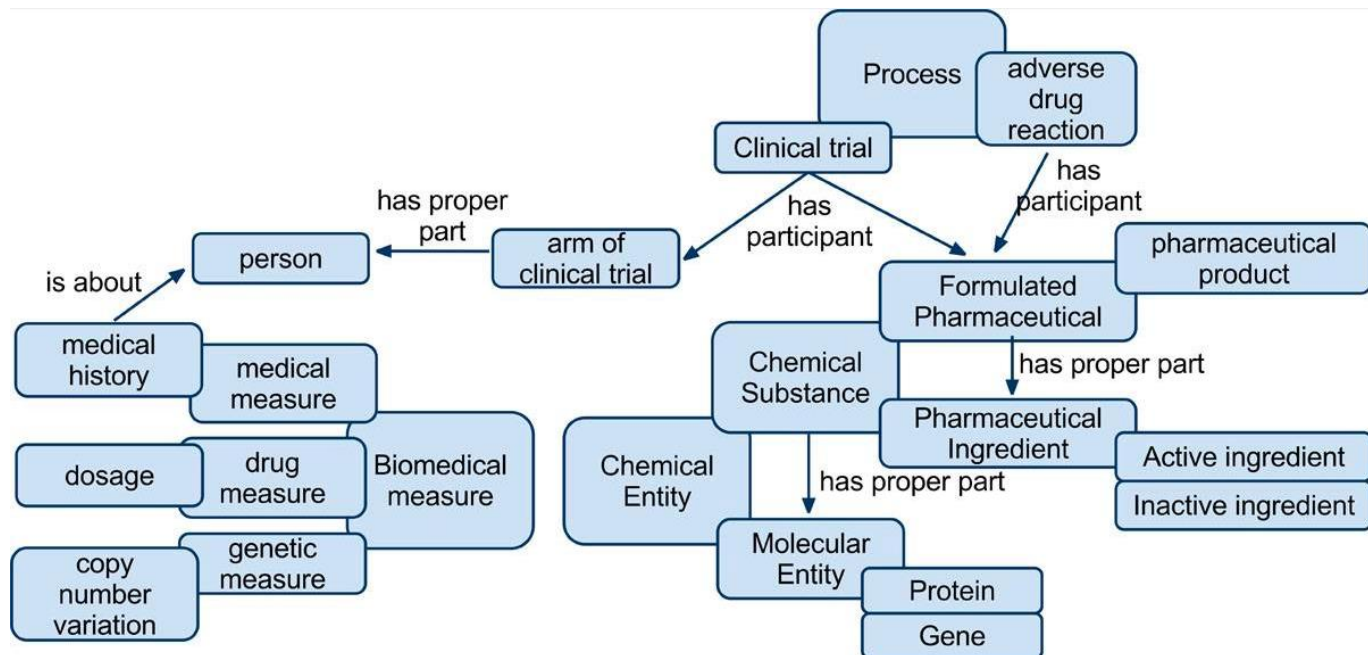
- Publications
- Submissions
- Presentations

People

- **Olivier Bodenreider**, MD, PhD: Group leader
- **Lee Peters**, MS: Senior software engineer
- Alumni and former collaborators

Contact

Email: Olivier.Bodenreider



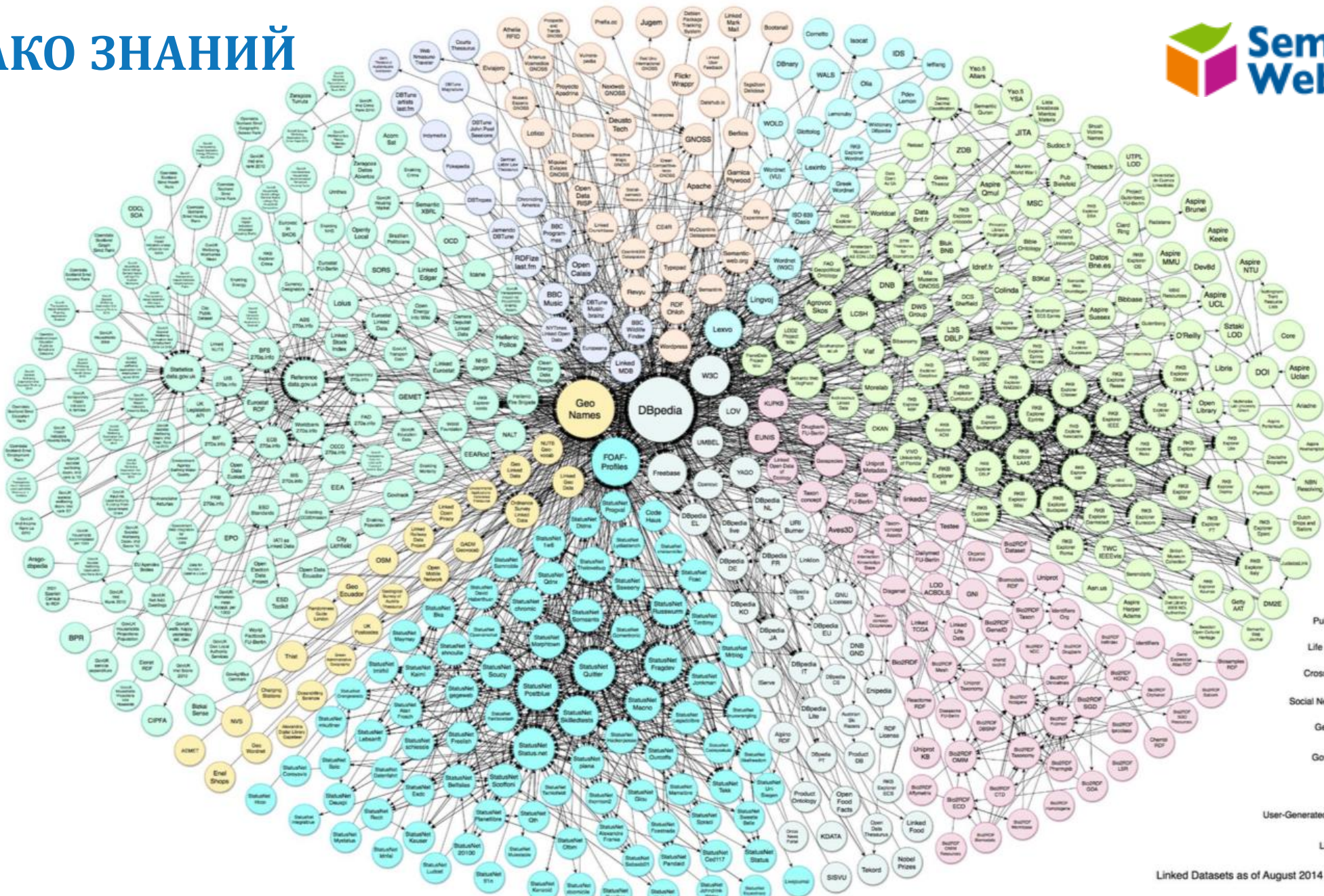
СЕМАНТИЧЕСКИЙ ПОДХОД

ОТКАЗ ОТ СТАНДАРТНЫХ РЕЛЯЦИОННЫХ НСИ И КЛАССИФИКАЦИИ
ОБЪЕКТОВ В СТОРОНУ СОЗДАНИЯ ИХ СЕМАНТИЧЕСКИХ ИНФОРМАЦИОННЫХ
МОДЕЛЕЙ



- В Базе Данных хранятся записи в виде триплетов: субъект-предикат-объект
- Запись триплета осуществляется согласно формату RDF, созданному консорциумом W3C и являющемуся платформо-независимым стандартом
- Отсутствие требований по предварительной классификации всей информации, можно использовать по необходимости разработанные ранее классификаторы
- Формирование информационной модели объектов на основе выявления и записи системы связей объекта с другими объектами; в семантике хранится не столько описание характеристик объекта, сколько описание его связей с иными объектами, по которым можно производить поиск данного объекта

ОБЛАКО ЗНАНИЙ

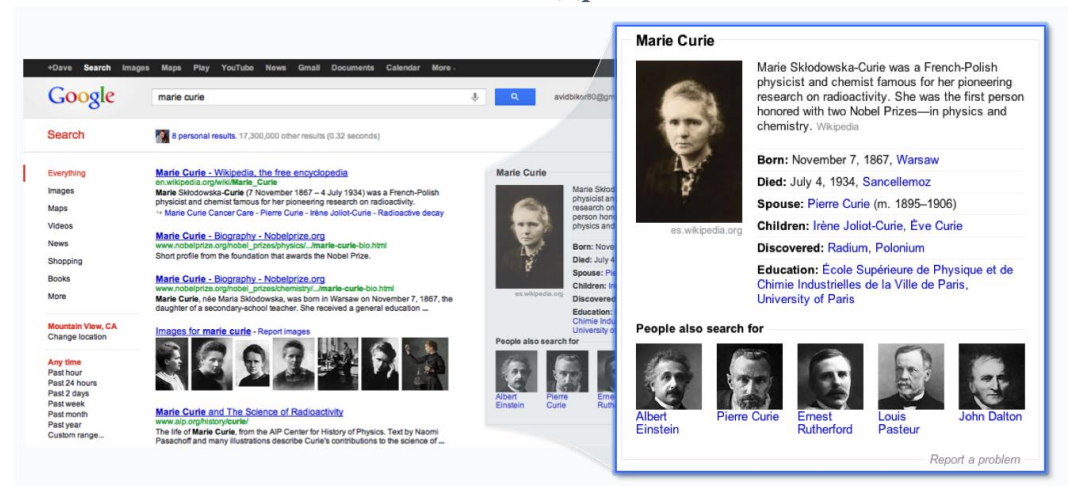


ПРИМЕРЫ БАЗ ЗНАНИЙ НА СЕМАНТИКЕ

ОБЛАКА СВЯЗАННЫХ ДАННЫХ В SEMANTIC WEB



- Google, обслуживающий около 9 млрд. запросов в день, создал в 2012 году Google Knowledge Graph, который насчитывает более 18 млрд. фактов, и составляет основу семантического поиска
- Аналогичные графы знаний используют для семантического поиска Microsoft, Yandex и др.
- Готовые облака связанных данных:
 - Кросс-доменные: DBPedia и Wikidata
 - Географические: Geonames
 - Библиотечные: Europeana, Worldcat, VIAF
 - Науки о жизни: Bio2RDF, Uniprot
 - Лингвистические: BabelNet
 - Средства массовой информации: BBC, New York Times
 - Промышленный Интернет IIoT: Semantic Sensor Network Ontology
- С момента своего создания в 2007 году, Облако связанных данных значительно расширилось, и в настоящее время, его элементы рассредоточены по нескольким каталогам, таким как Datahub.io, publicdata.eu, data.gov, open.canada.ca.



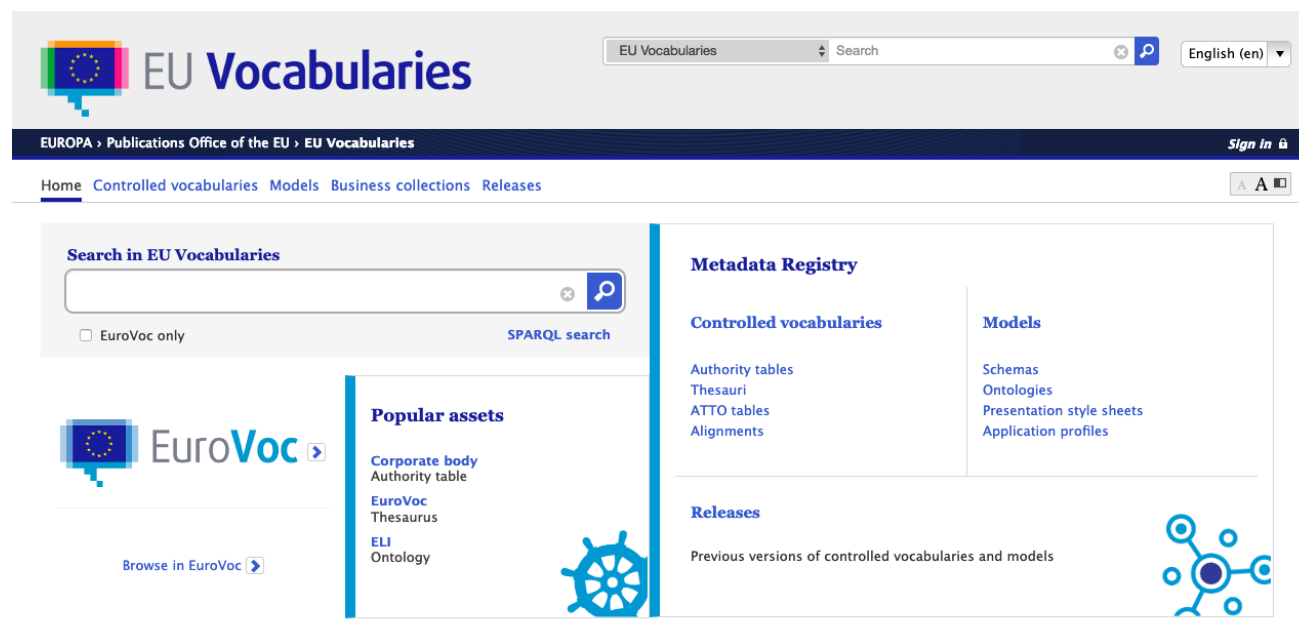
ЦИФРОВИЗАЦИЯ НОРМОТВОРЧЕСКОЙ ДЕЯТЕЛЬНОСТИ



В Европейских странах выход нового нормативного документа сопровождается электронной копией его контента, что значительно облегчает его интерпретацию и последующую имплементацию в автоматизированных системах.

Пример: сайт EuroVoc – место публикации всех справочников, имеющих хождение на территории европейских стран.

Как правило публикуются таксономии и семантические эталонные модели данных в различных отраслях, от финансового менеджмента и бухгалтерии до экологии и пространственных данных.



МЕЖОТРАСЛЕВОЙ РЕЕСТР СУБЪЕКТОВ И ОБЪЕКТОВ НА ОСНОВЕ СЕМАНТИКИ



ЦИФРОВАЯ ПЛАТФОРМА КОНТРОЛЬНО-НАДЗОРНОЙ ДЕЯТЕЛЬНОСТИ

ОНТОЛОГИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Семейство семантических информационных моделей (онтология предметной области) разрабатывается соотв. ведомствами вместе со специалистами-онтологами.

МОДЕЛЬНО-ОРИЕНТИРОВАННАЯ АРХИТЕКТУРА (MDA)

Проектирование всей системы на основе модели данных позволит гибко менять логику ее работы при изменении характера данных

ЗАЩИТА ДАННЫХ

Смена архитектуры (отказ от реляционных Баз Данных в структуре НСИ) позволит снизить уязвимость системы.

СКОРОСТЬ ПОИСКА

Поисковые алгоритмы оптимизированы под структуру данных. Существует готовая технология и язык запросов данных – SPARQL. Технология готова для внедрения Big Data.

ОТКРЫТЫЙ КОД

Существуют готовые решения по созданию Баз Знаний на основе лицензии Apache License 2.0, что существенно сократит время разработки Реестра.

НЕЗАВИСИМОСТЬ

Технология является промышленным стандартом и не зависит от конкретной реализации или производителя.

Спасибо за внимание

<http://borlas.ru>

Рагозин Олег Владимирович

oragozin@borlas.ru