



# **Открытые связанные данные (Linked open data) и открытый доступ (Open access)**

**в процессах создания и использования  
информационных научных ресурсов**

Клинцов В.П., Огороднийчук Д.Л.

Институт информационных технологий НИУ ВШЭ;

Жебель В.В.

ООО «НЦР Руконт».

# COMPETITIVENESS AND CORRUPTION

© V J ROTT Zurich

very high / high  
competitiveness  
of...

Usa  
Singapore  
Hong Kong  
Netherlands  
Finland  
Norway  
**Switzerland**  
Denmark  
Luxembourg  
Canada  
Ireland  
U.K.  
New Zealand

medium / low  
competitiveness  
of...

Germany  
Australia  
Taiwan  
Sweden  
Japan  
Iceland  
Malaysia  
France  
Austria  
Belgium  
China  
Israel  
Chile  
Spain  
Hungary  
Portugal  
Italy  
Argentina  
Philippines  
Turkey  
Mexico  
Korea

low / very low  
competitiveness  
of...

...

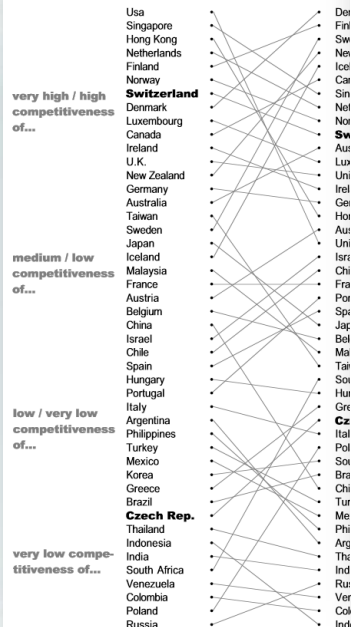
Denmark  
Finland  
Sweden  
New Zealand  
Iceland  
Canada  
Singapore  
Netherlands  
Norway  
**Switzerland**  
Australia  
Luxemburg  
United Kingdom  
Ireland  
Germany  
Hong Kong  
Austria  
United States

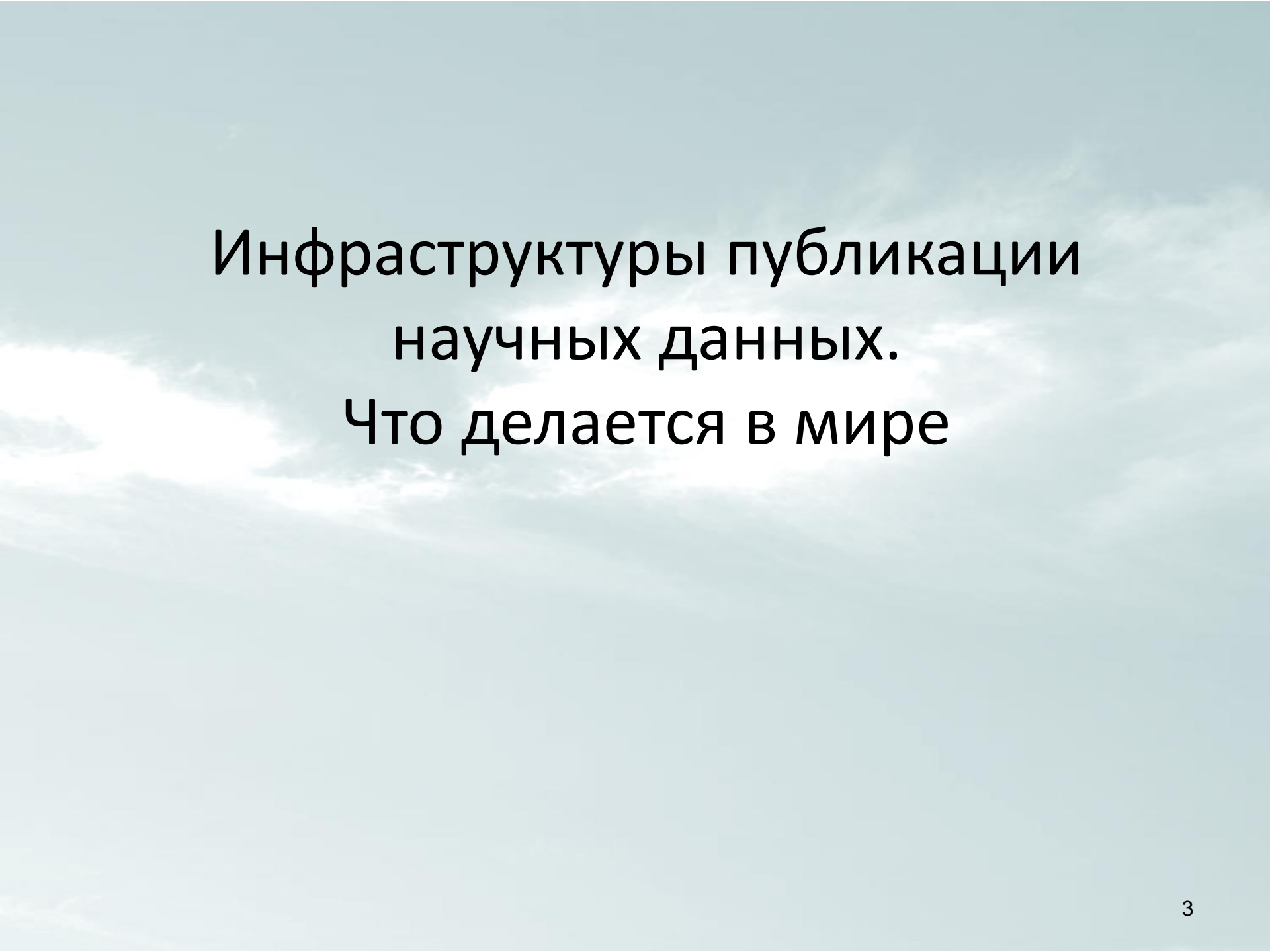
...totally clean  
and clean coun-  
tries

...less clean and  
rather corrupt  
countries

...rather corrupt  
and corrupt coun-  
tries

## COMPETITIVENESS AND C





# Инфраструктуры публикации научных данных. Что делается в мире

# Стратегия открытых данных Евросоюза

## *Открытые данные для науки*

- **Информационное сообщение и Рекомендации** государствам-членам по научной информации, начало 2012 г.;
- **Расширение пилотного проекта** с открытым доступом по научным публикациям до Horizon 2020 в полном объеме + пилотный проект с открытым доступом к данным научных исследований.
- **Расширение области действия Директивы** для включения библиотек, архивов, музейных и университетских библиотек,

## *Поддержка инфраструктур данных научных исследований*

- Рамочной программой FP7, а в дальнейшем - Программой Horizon 2020 предусмотрено оказание Комиссией поддержки работ по созданию **стабильной и надежной сервисной инфраструктуры для научных данных в Европе**, отвечающей потребностям в переработке больших объемов научной информации, прогнозируемых к 2020 г., (доклад 'Riding the Wave' - «Оседлав волну»). Инфраструктура обеспечит интерактивный доступ к информационной среде от первичных данных наблюдений и экспериментов до публикаций во всех областях науки.

## *Направления научных исследований и инноваций*

- **Научно-исследовательские и инновационные проекты**, относящиеся к открытым данным, особенно реализуемые через FP7, CIP и Horizon 2020, с финансированием исследовательских инфраструктур, поддерживающих открытый доступ к научным статьям и данным;
- **Конкурсы по открытым данным (2012-2013 гг.) + улучшение доступа к капиталу** для предпринимателей, работающих в этой области.

# 2013. На каких направлениях сосредоточены исследования и разработки

**API:** Стандартные API позволяют приложениям интегрировать данные в режиме реального времени при помощи новейших способов управления версиями и форматов открытых данных.

**Облачные вычисления:** Брокеры данных, располагающие большим количеством хранимых наборов данных, получают существенную выгоду с помощью облачных вычислений посредством использования инфраструктуры, которая может быть легко установлена и расширена.

**Linked Data, Big Data:** Принцип Семантического Web находит отклик среди лидирующих Web-компаний. С помощью schema.org Google, Yahoo! и Bing обращаются к сценарию Семантического Web с целью получения большего количества данных, улучшенной функциональной совместимости. В рамках работ по унификации моделей данных начаты разработка унифицированных базовых словарей для “выравнивания” и согласования данных. В этих работах (февраль – ноябрь 2012 г. приняли участие 67 экспертов из 21 стран-участников, представляющих университеты, международные стандартизирующие организации и агентства Евросоюза.

**Данные-как-Сервис (Data-as-a-Service):** Активное международное финансирование НИОКР и инноваций обусловлено тем, что научная поддержка рассматривается как обязательная составляющая при преодолении барьеров по открытию государственных данных. Университеты Европы и Америки при помощи финансовых программ реализуют научно-исследовательские проекты по открытым данным, охватывающих широкий спектр исследований и приложений..

Источник: [https://joinup.ec.europa.eu/sites/default/files/D1.1\\_20\\_UNCEFACT.pdf](https://joinup.ec.europa.eu/sites/default/files/D1.1_20_UNCEFACT.pdf)

# Инвестиции в Евросоюзе в R&D: семантика и открытые данные

Online Content, Interactive & Social Media M€		Reasoning & Information Exploitation M€	
<a href="#">APIDIS</a>	€2,62	<a href="#">PlanetData</a>	€3,72
<a href="#">IMP</a>	€4,09	<a href="#">LARKC</a>	€10,69
<a href="#">CASAM</a>	€4,24	<a href="#">Cubist4,36</a>	€4,36
<a href="#">PuppyIR</a>	€4,18	<a href="#">PLUGIT</a>	€3,66
<a href="#">SMARTMUSEUM</a>	€1,93	<a href="#">IKS</a>	8,55
<a href="#">3D POST</a>	€4,25	<a href="#">Advance</a>	€2,79
<a href="#">ANSWER</a>	€4,65	<a href="#">CALBC</a>	€2,20
<a href="#">IRIS</a>	€3,20	<a href="#">OKKAM</a>	€7,36
<a href="#">FOCUS K3D</a>	€1,87	<a href="#">ONTORULE</a>	€8,03
		<a href="#">LATC</a>	€1,19
"Интеграционные" проекты M€		Knowledge Discovery & Management M€	
<a href="#">INSEMTIVES</a>	€ 5,43	<a href="#">ACTIVE</a>	€11,94
<a href="#">NoTube</a>	€ 9,25	<a href="#">Dicode</a>	€3,51
<a href="#">SYNC3</a>	€ 4,48	<a href="#">KYOTO</a>	€3,32
<a href="#">SmartProducts</a>	€ 10,55	<a href="#">eLICO</a>	€4,47
<a href="#">SmartVortex</a>	€ 10,49	<a href="#">Promise</a>	€4,27
<a href="#">iProd</a>	€ 4,66	<a href="#">WEKNOWIT</a>	€7,50
<a href="#">Tridec</a>	€ 8,90	<a href="#">SERVICE FINDER</a>	€1,69
<a href="#">Teleios</a>	€ 3,77	<a href="#">JUMAS</a>	€3,70
<a href="#">LOD2</a>	€ 8,58	<a href="#">VALUE IT</a>	€1,93
<a href="#">PRONTO</a>	€ 3,95		
<a href="#">First</a>	€ 4,57	<a href="#">ISA</a>	€164,60
<a href="#">KIWI</a>	€ 3,80		
<a href="#">Robust</a>	€ 10,32		
<a href="#">Render</a>	€ 4,43		
<a href="#">Khresmoi</a>	€ 10,82		
<a href="#">Digital.me</a>	€ 4,35		
<a href="#">Fish4Knowledge</a>	€ 2,54		



## Связанные датасеты Университета статистики Италии

<http://www.w3.org/wiki/HCLSIG/LODD>

**Сегмент  
связанных баз  
по лекарствам**



Рис. 4 – Наборы данных, входящих в Linking Open Drug Data – LODD

Media  
Geographic  
Publications  
User-generated content  
Government  
Cross-domain  
Life sciences

As of September 2011

## Сегмент государственных данных Евросоюза



# «5 звезд» - модель оценки публикуемых в Вебе данных



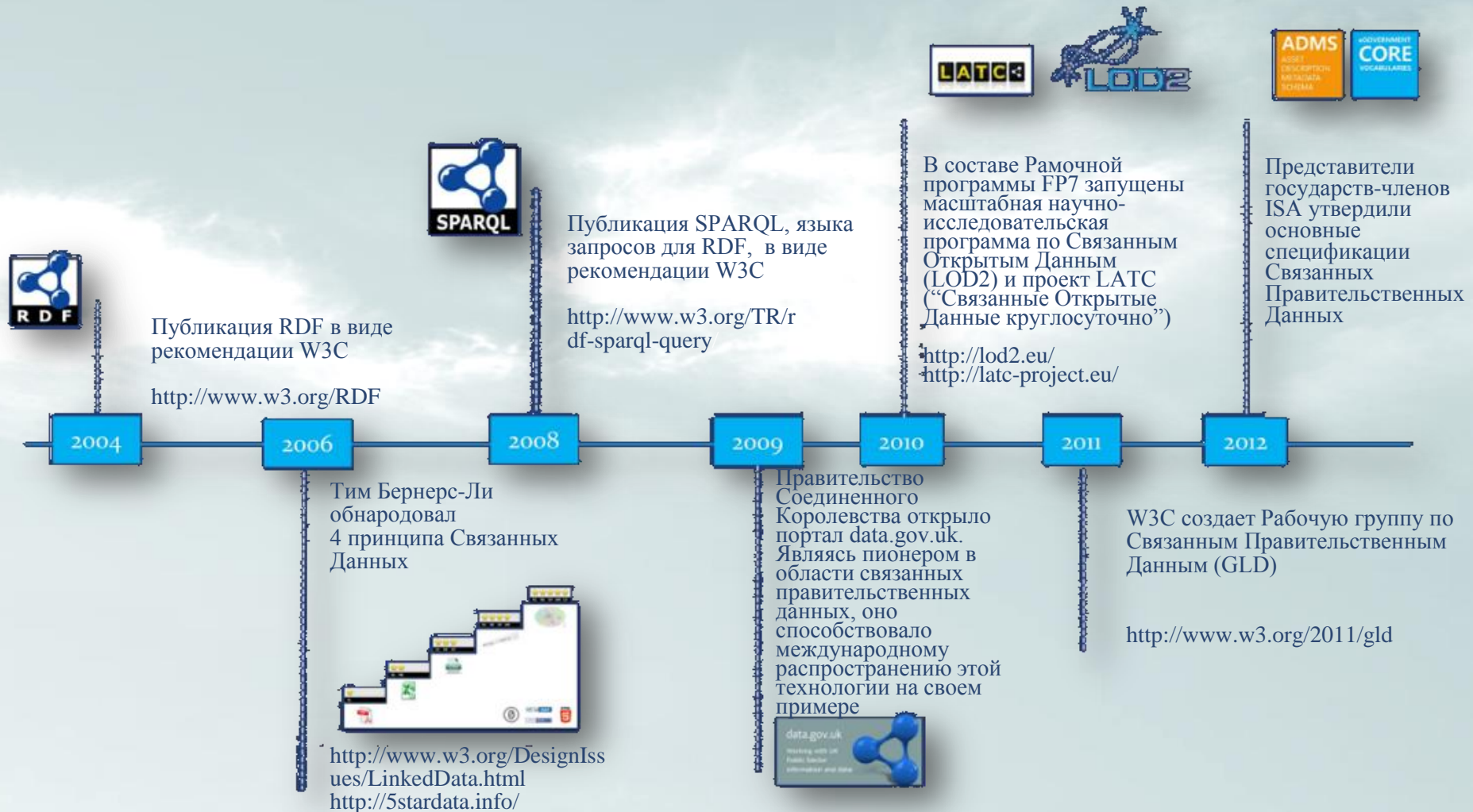
Данные	Оценки	Метаданные
Доступны в Сети по открытой лицензии	★	Пренебрежение метаданными
Доступны в машиночитаемом и структурированном виде	★★	Рассредоточенные и/или закрытые метаданные
плюс несобственнический формат	★★★	Открытые метаданные для людей (не для машин)
плюс RDF и SPARQL	★★★★	Открытые повторно используемые метаданные
Связанные данные	★★★★★	Связанные открытые метаданные

Тим Бернерс-Ли (Tim Berners-Lee)

Вассилиос Перистерас (Vassilios Peristeras)



# Этапы развития концепции Связанных Данных eGovernment в Европе



# Европа. Архитектура Связанных Данных для интеграции данных

Связанные Данные применяются для интеграции данных, размещенных в различных типах источников, имеющих различные форматы, как структурированных, так и неструктурированных.

Первый шаг состоит в разработке постоянных URI, назначения их данным и описания этих данных в виде RDF с использованием общепринятых метаданных. В случае структурированных данных (напр., хранящихся в реляционных базах данных, файлах CSV, Excel или XML), RDF-оболочки, модули экспорта и API адаптируются для преобразования этих данных в RDF с последующей публикацией в Веб, или любых других источниках, допускающих обмен этим типом данных. Для неструктурированных данных (напр., новостные статьи или бизнес-отчеты), то для их публикации в виде Связанных Данных необходимо прежде задействовать средства извлечения сущностей и углубленного анализа текстов с целью обнаружения информационных объектов.

Когда данные представлены в RDF-формате, их следует связать с другими данными, поступающими из надежных источников, чтобы сформировать их контекстное окружение и обогатить смысловое содержание.

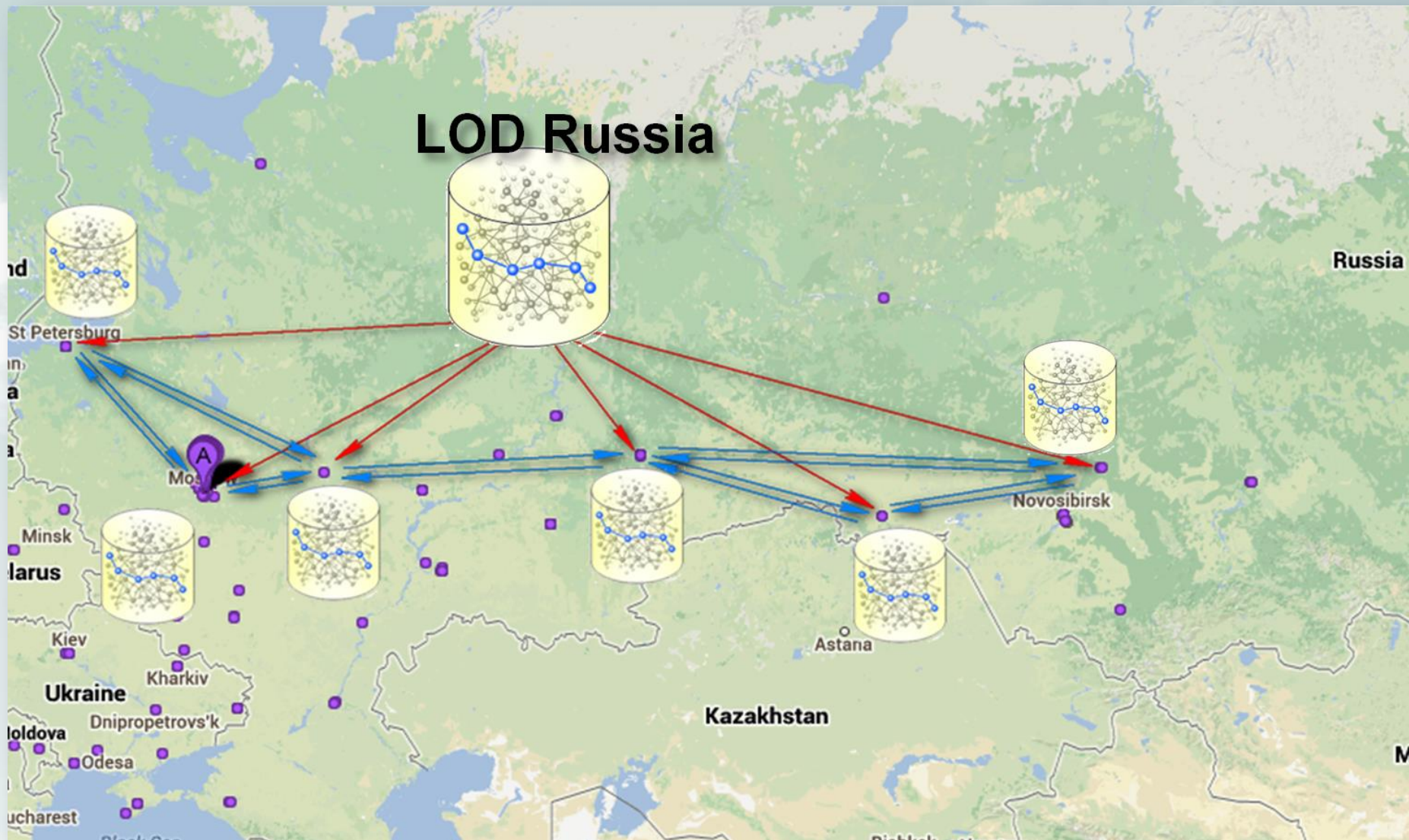
После публикации Связанных Данных любая система, способная обрабатывать семантические запросы (при помощи языка запросов SPARQL) сможет извлекать и повторно использовать релевантную информацию.



[8] <http://linkeddatabook.com/editions/1.0/>

[9] <https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris>

# «5 звезд». Инфраструктура публикации научно-образовательных данных России






# Webometrics сегодня, – это:

**1. Visibility (50%)** известный ранее, как "impact", - число уникальных входящих внешних ссылок ([Majestic SEO](#) и [ahrefs](#)) и число **доменов**, генерирующих эти ссылки.

**2. Activity (50%)** состоит из:

- **Presence (1/3)** известный ранее, как "size". Количество страниц, фиксируемое Google.
- **Openness (1/3)**, известный ранее, как "rich" (pdf, doc, docx, ppt). Количество статей, опубликованных на сайтах в соответствии с поисковиком [Google Scholar](#).
- **Excellence (1/3)**, известный ранее, как "scholar". Статьи, опубликованные в высокорейтинговых международных журналах в соответствии с Scholar и [Scimago group](#).

**Бороться надо за:**

**Visibility + Presence = 50% + 16.6% = 66.6%** 



## Шаг 1 – создание ядра инфраструктуры:



- **Выгрузить (контролируемо) из транзакционной базы «Карты науки России» данные в LOD формате – LOD Russia**
- Включить в “Примерный перечень критериев общероссийской системы оценки деятельности высших учебных заведений” критерии **Visibility** и **Presence** Webometrics. Разработать стандарты для **обязательных разделов сайтов** российских университетов.

# Шаг 2-й (1) – что нужно сделать:



**СМИРНОВ ДМИТРИЙ АЛЕКСАНДРОВИЧ**

кафедра теоретической и прикладной механики

Нижегородский государственный технический университет

Методы локализации производства оборудования и технологий в системе стратегического управления электросетевой компанией // Вестник Научно-исследовательского центра корпоративного права, управления и венчурного инвестирования Сыктывкарского государственного университета. 2012. № 1

**ШУВАЛОВА ДАРЬЯ ГЕОРГИЕВНА**

НИУ ВШЭ. Институт проблем ценообразования и регулирования естественных монополий

Сайт НИУ ВШЭ

О Вышке → Преподаватели и сотрудники → Шувалова Дарья Георгиевна



**Шувалова Дарья Георгиевна**

Институт проблем ценообразования и регулирования естественных монополий  
Научный сотрудник

(Сотрудничает с НИУ ВШЭ с 2010 года.)

**Образование, учёные степени**

- Кандидат наук: 2009
- Московский энергетический институт, 2004, специальность «Экономика и управление в энергетике»
- Бакалавриат: Московский энергетический институт, 2002, специальность «Экономика и управление в энергетике»

**Публикации**

Шувалова Д. Г., Бобылева Т. Н. Оценка эффективности внедрения альтернатив на примере корпусов НИУ МЭИ // В кн.: Энергосбережение - теория и практика: труды семинара молодых учёных и специалистов. М.: Издательский дом МЭИ, 2013. С. 380-383.

Шувалова Д. Г., Тарапова Т. П. Экономические аспекты анализа эффективности энергосбережения и повышения энергоэффективности // В кн.: Энергосбережение. Всероссийской школы-семинара молодых учёных и специалистов. М.: Издательский дом МЭИ, 2013. С. 380-383.

Шувалова Д. Г., Бадаев О. Экономический эффект внедрения элементов интеллектуальной цифровой подстанции // В кн.: Энергосбережение - теория и практика: труды семинара молодых учёных и специалистов. М.: Издательский дом МЭИ, 2013. С. 380-383.

Волкова И. О., Смирнов Д. А., Шувалова Д. Г. Методы локализации производства оборудования и технологий в системе стратегического управления электросетевой компанией // Корпоративное управление. Вестник Научно-исследовательского центра корпоративного права, управления и венчурного инвестирования Сыктывкарского государственного университета. 2012. № 1

Контактная информация  
E-mail: dshuvanova@hse.ru

**Новая** страница на Сайте НГТУ

**1** Личная карточка ученого предоставит детальную информацию о нем, а также ссылки на другие объекты в базе знаний

**Proscience**

Иванов Иван Иванович, к.ф.-м.н. Сравнить Анализ связей Критерии

2006-... Институт радиотехники и электроники им. В.А. Котельникова РАН (Москва)  
Научные интересы: Физика, Химия, Гидродинамика, ...

2000-2008 Московский физико-технический институт (Москва)  
Научные интересы: Химия, Биология, Радиобиология, ...

Выбор направления фильтрует содержимое остальных блоков карточки ученого

Публикации  
1. СИНХРОНИЗМ ДИНАМИКИ И ВНУТРЕННЕЕ ДВИЖЕНИЕ В МАССОВОРАЗВЛЕКАТЕЛЬНЫХ МАНТАИТАХ ПО ДАННЫМ ЭПР. Смирнов В.А., Демидов В.В., Готовцев Д.Г., Нополова Н.Е., Иванов И.И., Бах Р. Журнал экспериментальной и теоретической физики. 2004. Т. 126. № 1. С. 229-238.

2 NMR AND SPIN RELAXATION IN LGA1-XMNXO3: EVIDENCE FOR THERMALLY ACTIVATED INTERNAL DYNAMICS  
Noginova N., Arthur E., Weaver T., Ivanov I.I., Atsarkin V.A., Gotovtsev D.G. Physical Review B: Condensed Matter and Materials Physics. 2004. Т. 69. № 2. С. 244061-2440610.

Все публикации...

Гранты  
08-05-12060-офн «Разработка научных основ дистанционного мониторинга изменений природных комплексов» (РФФИ)

Все гранты...

Патенты  
Патент РФ № 2110005 (27/04/1998) «Классификация объектов»

Все патенты...

1. В страницу публикаций ученого загружаются Веб-адреса его **соавторов** из других университетов
2. С Веб-страниц формируются прямые перекрестные ссылки на страницы соавторов





18.07.2013

Dear Sirs,

During the STI summit in Suzdal we had a good chance to acquaint ourselves with the Russian delegation's vision of how the Open Data concept can be further developed in Russia, and how it can influence the ICT development in this country. We can affirm that a substantial progress has been achieved in this field by Russian structures involved in the process: during less than one year 1000+ datasets were published by 50 Ministries; a work on improving the quality of published datasets was performed (the "5 Star" project of RIA Novosti); a Russian DBMS ensuring the storage of data according to Web standards was developed.

We also considered a potential Russian project dealing with the creation of an infrastructure for publishing scientific and educational data in conformity with the "5 Star" standards and basing on data published on the Web that are within the competence of the Ministry of Education and Science of the Russian Federation.

The said project, in case of its implementation, can produce a material effect upon the development of scientific and educational structures not only in Russia, but also in countries the scientific research community of which we have the honor to represent.

The publication of data in the LOD (Linked Open Data) form obtained as a result of implementing the "Map of Russian Science" project not only can be used for the creation of a Russian infrastructure for publishing scientific data, but can also contribute to the better positioning of the Russian education research system in the worldwide rating systems.

We, being official representatives of an association that develops innovative standards and methods for implementing business processes on the World Wide Web, as well as full professors and research scientists, are willing and prepared to assist in making our universities active participants of this initiative.

We believe that the implementation of this project will give a new impetus to the development of innovative economies in our countries.

Sincerely yours,

**Prof. John Domingue** - The Open University United Kingdom,  
President STI International

**Prof. Sören Auer** - University of Bonn and Fraunhofer IAIS,  
Co-Chair STI International, Germany

**Prof. Dieter Fensel** - Director STI Innsbruck, University of Innsbruck,  
Austria

**Prof. Georg Gottlob** - University of Oxford, UK

**Dr. Tim Furge**, University of Oxford, UK

**Prof. Key-Sun Choi** - KAIST, Korea

**Prof. Ho-Jin Choi** - KAIST, Korea

**Dr. Anna Fensel** - University of Innsbruck, Austria

**Dr. Ioan Toma** - University of Innsbruck & STI-R, Austria

**Dr. Dmitry Mouromtsev** - NRU ITMO, Russia


**Ivan Ermilov** - University of Leipzig, Germany

Инфраструктуры публикации  
научных данных.

Открытый доступ.  
Что это дает.



# ...на примере диссертаций




РОССИЙСКАЯ ФЕДЕРАЦИЯ

**БЕСПЛАТНАЯ НАУЧНАЯ БИБЛИОТЕКА АВТОРЕФЕРАТОВ ДИССЕРТАЦИЙ**


- ГЛАВНАЯ
- КАНДИДАТСКИЕ АВТОРЕФЕРАТЫ
- ДОКТОРСКИЕ АВТОРЕФЕРАТЫ
- НАУЧНЫЕ СТАТЬИ
- МЕТОДИЧЕСКИЕ ПОСОБИЯ
- ОФОРМЛЕНИЕ КАНДИДАТСКОЙ
- ДОКТОРАНТУРА
- НАУЧНЫЕ ЖУРНАЛЫ
- АВТОРЕФЕРАТЫ УКРАИНЫ
- АВТОРЕФЕРАТЫ БЕЛОРУССИИ
- ОФИЦИАЛЬНЫЕ ДОКУМЕНТЫ
- СЛОВАРЬ ТЕРМИНОВ
- КОНТАКТЫ


Добро пожаловать!

[Яндекс.Директ](#) [Все объявления](#)

 [Бесплатные диссертации РГБ](#) Скачать диссертацию бесплатно. Библиотека диссертаций и авторефератов.  
dslib.net


[Закажи диссертацию у Dissertation](#) Профессиональное написание диссертации! Индивидуальный подход. Гарантии.  
dissertation.msk.ru

 [Ищешь реферат? Могу написать](#) Уникальные рефераты на заказ. Быстро, недорого, гарантия. Узнай цену сейчас  
[Адрес и телефон](#) zaochnik.ru

 [Авторефераты по всем темам](#) >> [Авторефераты по сельскому хозяйству](#)

**Повышение эффективности производства свинины на основе метода гибридизации**

*Автореферат докторской диссертации по сельскому хозяйству*

 [СКАЧАТЬ ОРИГИНАЛ ДОКУМЕНТА](#) Страницы: 1 2 3

Суслина елена николаевна  
**ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ПРОИЗВОДСТВА СВИНИНЫ НА ОСНОВЕ МЕТОДА ГИБРИДИЗАЦИИ**  
06.02.07 - Разведение, селекция и генетика  
сельскохозяйственных животных  
**АВТОРЕФЕРАТ**  
диссертации на соискание ученой степени  
доктора сельскохозяйственных наук  
п. Лесные Поляны Московской области  
2011 г.

Работа выполнена во Всероссийском научно-исследовательском институте племенного дела.

Реклама от Google

[СВИНЕЙ](#)  
[КОРМЛЕНИЕ КРС](#)  
[ОТКОРМ КРС](#)

[Яндекс.Директ](#) [Все объявления](#)  
[Диссертации на заказ](#)  
Диссертации: докторские, кандидатские, магистерские. Любые специальности.  
[Адрес и телефон](#) диплома@yandex.ru

[Дни премиум-седанов VOLVO!](#)  
Роскошные предложения на Volvo с 8 по 10 февраля

# Как это можно сделать...

## На примере диссертаций

Суслина Елена Николаевна

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ  
ПРОИЗВОДСТВА СВИНИНЫ НА  
ОСНОВЕ МЕТОДА ГИБРИДИЗАЦИИ

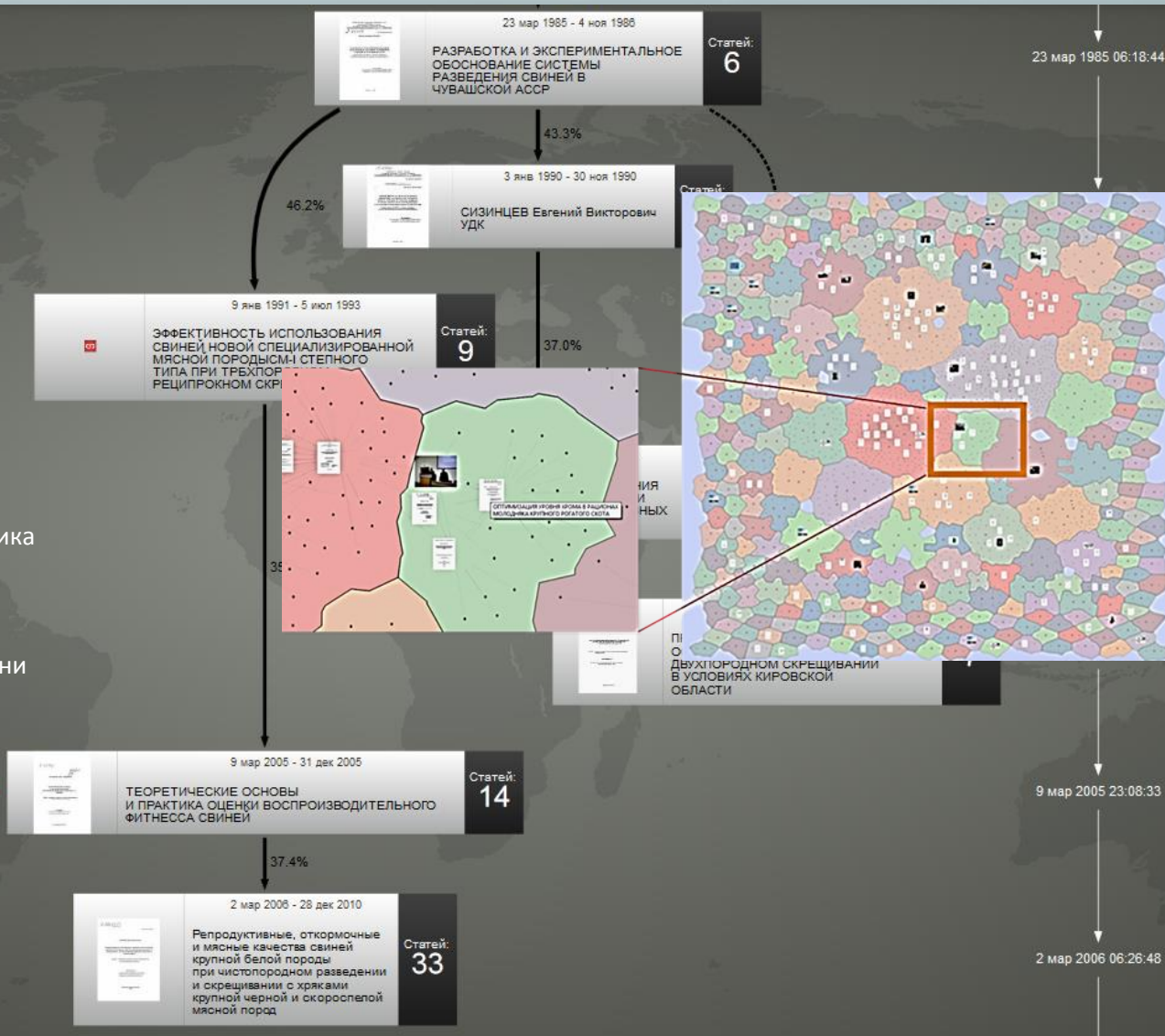
06.02.07 - Разведение, селекция и генетика  
сельскохозяйственных животных

АВТОРЕФЕРАТ

диссертации на соискание ученой степени  
доктора сельскохозяйственных наук  
п. Лесные Поляны Московской области

На эту же тему

релевантность 16.0% (β-beta)



# «Рост науки»

1665 г.  
«Journal des  
sçavans» -  
первый  
научный  
журнал

Д.Прайс: «We  
can say that 80  
to 90 percent of  
all the scientists  
that have ever  
lived are alive  
now»

с 1986 по 2002  
количество  
научных  
журналов  
увеличилось  
на 58%

Ученый около 60% времени тратит  
на поиск научной информации (Коэн, 2006)

# Развитие Open Access

«Serials crisis» с 1984 по 2005 рост стоимости доступа к USA Periodical в 6,4 раза превысил рост CPI

В 2009 году OA охватывает 20,6% ISI articles и 20,4% Scopus+ISI articles

JISC: стоимость доступа подписка £2,337, стоимость OA - £1,524

«Под «открытым доступом» мы подразумеваем открытые для всех публикации в Интернете, которые можно читать, загружать, копировать, распространять, распечатывать...»



# Open Access

Green OA  
(~12%)

Gold OA  
(~8,5%)

