



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

# Форматы открытых данных и работа с открытыми данными

**Радченко Ирина Алексеевна**

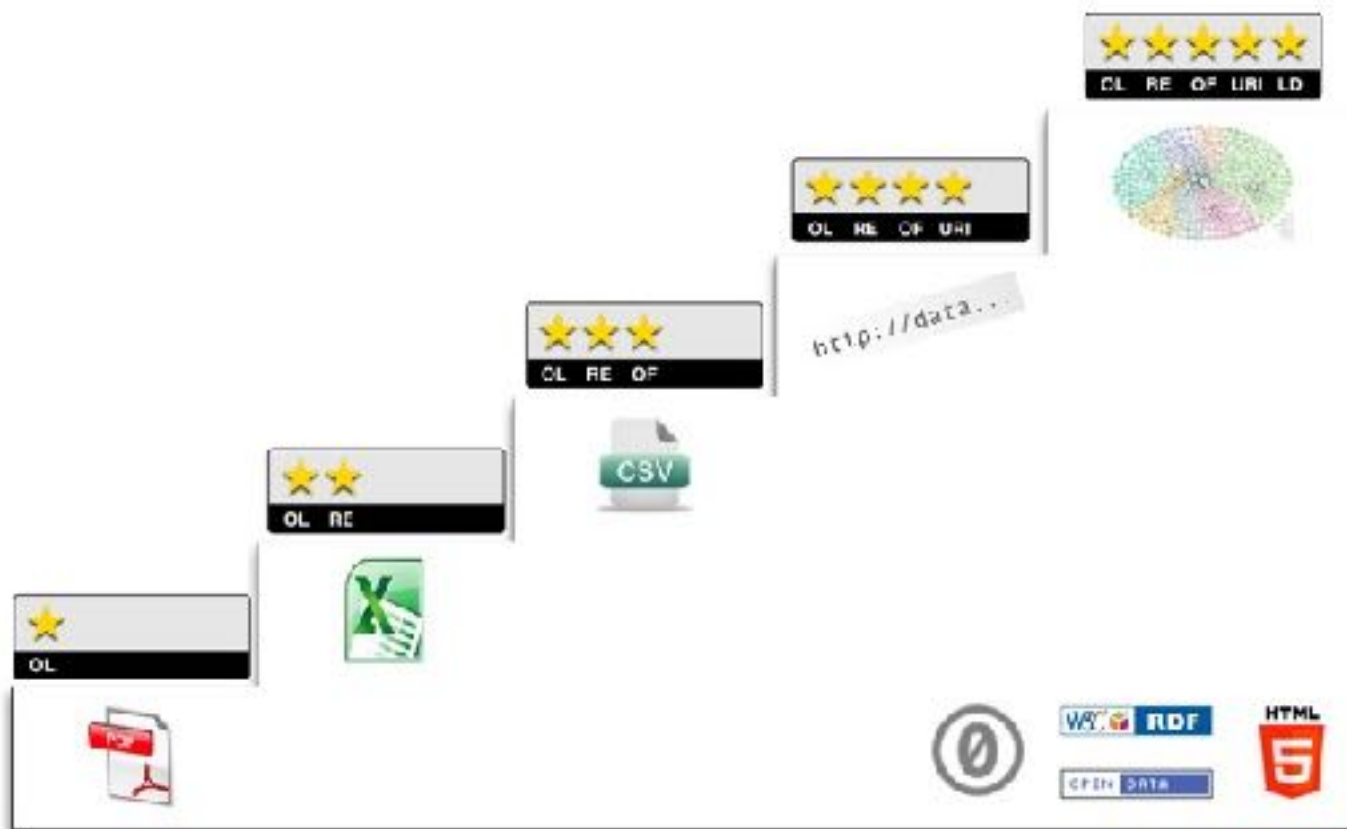
кандидат технических наук, доцент

**<https://t.me/dataplace>**

# Требование машиночитаемости данных



# Пятизвездочная модель открытых данных по Тиму Бернерсу-Ли



Источники: <http://5stardata.info/>, <http://www.w3.org/DesignIssues/LinkedData.html>

# Формат PDF (Portable Document Format)

Представление в электронном виде полиграфической продукции



Достоинства	Недостаток
Представление документов в электронном виде	Мало пригоден для машинной обработки
Кроссплатформенность	
Стандарт ISO 32000-1:2008	
Много программного обеспечения (в том числе, свободного) для работы с данным форматом	

# Excel-формат

## Представление данных в табличной форме



Представление данных в табличной форме

Пригодность для машинной обработки

Много программного обеспечения для работы с данным форматом

# Формат CSV

## Представление данных в табличной форме



Представление данных в табличной форме

Пригодность для машинной обработки

Открытый формат данных

Много программного обеспечения (в том числе, онлайн-сервисов) для работы с данным форматом

# Формат JSON (JavaScript Object Notation)

Текстовый формат обмена данными



Читается как компьютерами, так и людьми

Пригодность для машинной обработки

Может использоваться с любым языком программирования

Подходит для сериализации сложных структур данных

# Формат XML (eXtensible Markup Language)



Читается как компьютерами, так и людьми

Рекомендован консорциумом W3C

Пригодность для машинной обработки

Удобен для обмена данными (а также документами) в интернете



# Модель представления данных RDF



Пригодность для машинной обработки

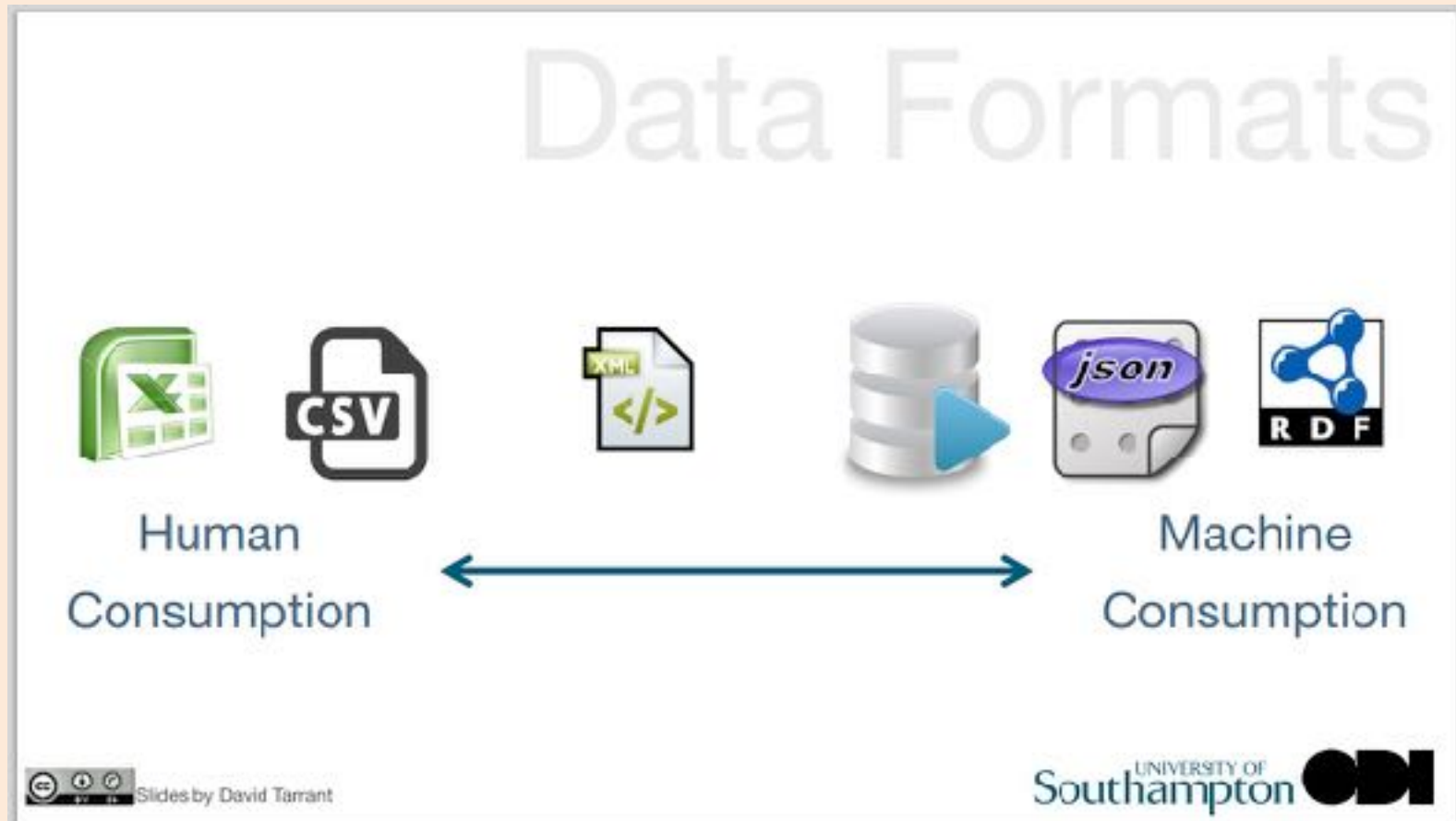
Является частью Семантического веба  
(Semantic Web)

Используемые форматы данных: RDF/XML,  
RDF/JSON, RDFa, N-Triples, Turtle, N3

Использует словари, таксономии и онтологии



# Модель представления данных RDF



Источник: [http://odinprac.theodi.org/September\\_2013/resources/Discovering\\_Data.pdf](http://odinprac.theodi.org/September_2013/resources/Discovering_Data.pdf)

# Процесс работы с открытыми данными



## Полезные ссылки

1. Список программного обеспечения для работы с PDF: [http://en.wikipedia.org/wiki/List\\_of\\_PDF\\_software](http://en.wikipedia.org/wiki/List_of_PDF_software)
2. ISO 32000-1:2008: [http://www.iso.org/iso/catalogue\\_detail?csnumber=51502](http://www.iso.org/iso/catalogue_detail?csnumber=51502)
3. Описание CSV: <http://tools.ietf.org/html/rfc4180>
4. Описание JSON: [https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Global\\_Objects/JSON?redirectlocale=en-US&redirectslug=JavaScript%2FReference%2FGlobal\\_Objects%2FJSON](https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Global_Objects/JSON?redirectlocale=en-US&redirectslug=JavaScript%2FReference%2FGlobal_Objects%2FJSON)
5. Описание RDF: <http://www.w3.org/RDF/>





# Спасибо за внимание!

<http://iRadche.ru>

<http://about.me/Irina.Radchenko>



@iRadche



<http://iRadche.livejournal.com/>



<https://www.facebook.com/iRadche>



<http://www.slideshare.net/iRadche>