

XII международная конференция
CEE-SECR / РАЗРАБОТКА ПО

28 - 29 октября, Москва



Восхождение Гражданской науки о данных

Докладчик

Ирина Радченко
Университет ИТМО

Ольга Максименкова
Высшая школа экономики



Ирина Радченко

Кандидат технических наук,
Доцент университета ИТМО

Сооснователь проекта DataDrivenJournalism.Ru

Основатель ODI Moscow (позднее ODI St.Petersburg)



@iradche

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

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

<https://github.com/iradche>

iradche@gmail.com

Потребности в дата-ученых



THE BIG DATA MARKET

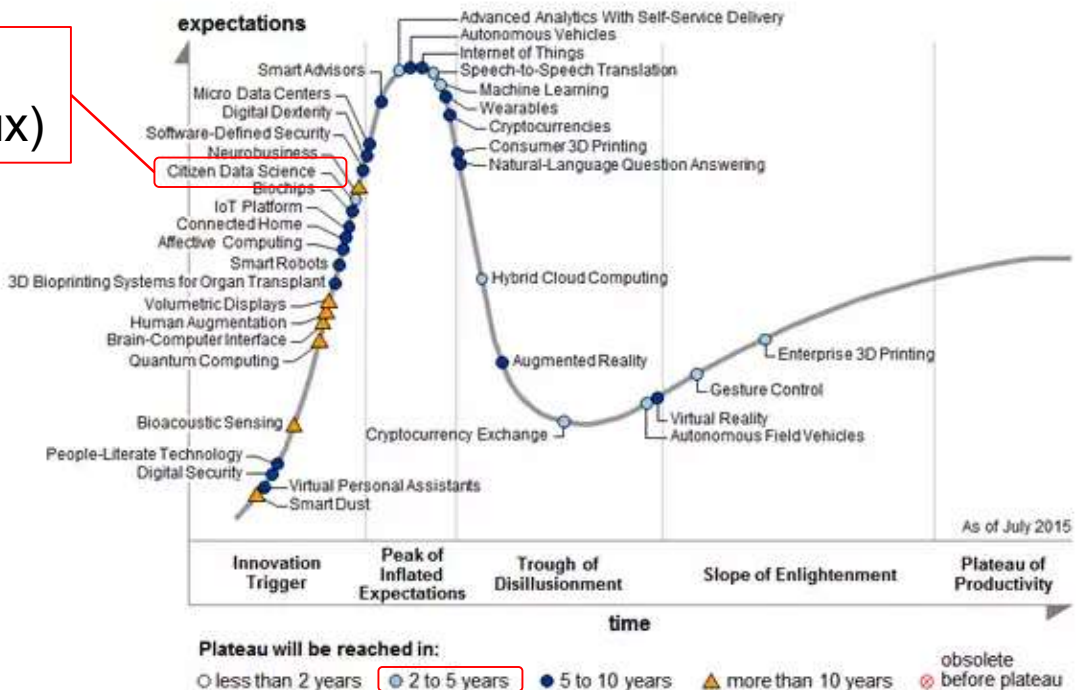
The Need for Data Scientists Is Exploding

A LARGE NUMBER OF USE CASES FOR HADOOP and Spark are in and around data science. A quick search for "Data Scientists" on LinkedIn shows that there are 16,000 employed in the US. What is shocking is that there are a whopping 5,000 open job posts for people with data-science skills. One can appreciate the demanding difference if you compare this to clinical scientists at pharmaceutical companies. There are 3,200 clinical scientists in the U.S., according to LinkedIn, but a crawl of job posts shows only 200 vacancies for that position across American companies.

A Data-Driven Analysis of Companies Using Hadoop, Spark, and Data Science
Источник: <http://www.oreilly.com/data/free/the-big-data-market.csp>

Pure Cycle for Emerging Technologies, 2015

Citizen Data Science,
Гражданская наука (о данных)



The Mirage of a Citizen Data Scientist

◀ Previous post

Next post ▶



64



Tweet

Tags: [Citizen Data Scientist](#), [Data Analyst](#), [Data Scientist](#), [Gartner](#), [Overfitting](#)

The term "citizen data scientist" has been irritating me recently. I explain why I think it both a bad term and a bad idea, and what we need instead.

By [Gregory Piatetsky](#), [KDnuggets](#).

comments

The term "citizen data scientist" has been irritating me recently.

It started appearing in 2015 following Gartner research director Alexander Linden suggestion to cultivate

"citizen data scientists" - people on the business side that may have some data skills, possibly from a math or even social science degree-and putting them to work exploring and analyzing data.

This term was picked up by publications like [TechRepublic](#) and [NGData](#) which proclaim the rise of the citizen data scientist.

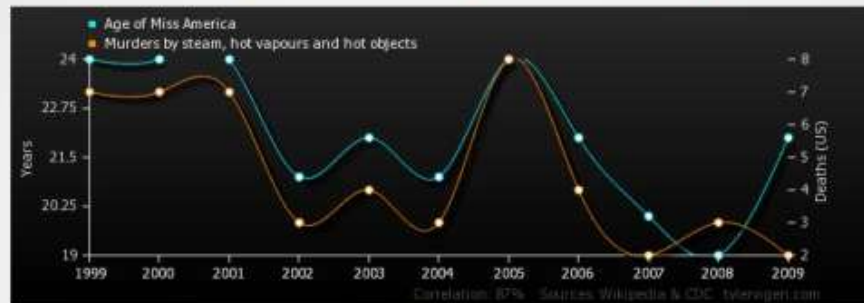
I dislike "citizen data scientist" for two reasons.

First, the word "citizen" has very misleading connotations, especially given the heated debate now in the US regarding immigration.

Having organized and attended many US conferences on data mining and

Age of Miss America correlates to Murders by Steam, hot vapours, and hot objects.

Age of Miss America correlates with Murders by steam, hot vapours and hot objects



Here are a couple examples from my experience. KDnuggets has recently tested [IBM Watson Analytics for social media analysis](#), which is one of the semi-automated tools that is supposed to be suitable for "citizen" Data Scientist.

We applied IBM Watson to KDnuggets Twitter data from January and one of the key findings was that there was a much higher engagement on Thursdays.

The data was convincing enough, however when I checked the February data, Thursday was no longer the most engaging day for tweets. When I dug deeper, I found that January results were due to one particularly popular tweet. However, someone without understanding Twitter would have reached the conclusion that Thursday is always the best day for tweeting.

As the comments from my previous post dealt with correlation, I thought I'd follow

Источник: <http://www.gartner.com/newsroom/id/3114217>

Определение



- Гражданская наука — это научная работа, производимая публично, в сотрудничестве или под руководством профессиональных ученых и научных организаций



Levels of Citizen Science

Level 4 'Extreme'	• Collaborative Science – problem definition, data collection and analysis
Level 3 'Participatory science'	• Participation in problem definition and data collection
Level 2 'Distributed Intelligence'	• Citizens as basic interpreters
Level 1 'Crowdsourcing'	• Citizens as sensors



Принципы Гражданской науки



1. Вовлечение потенциальных участников
2. Выгода для всех участников
3. Политика открытости
4. Открытая методология
5. Открытые научные данные
6. Открытый доступ
7. Популяризация науки и распространение культуры



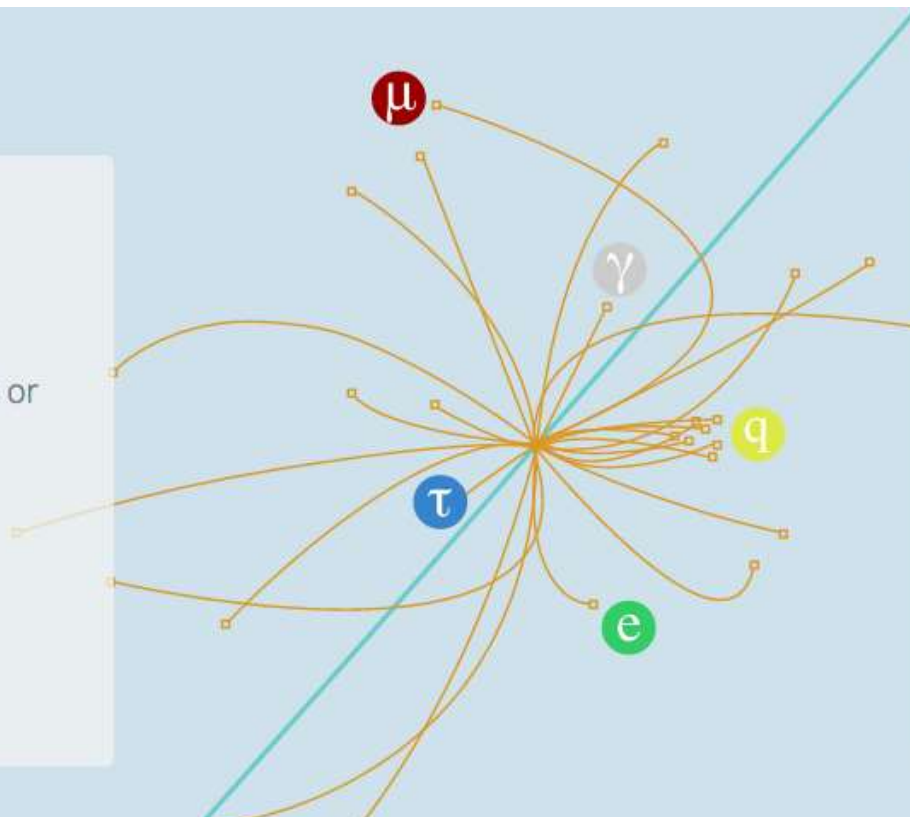
На основе принципов European Citizen Science Association
(Natural History Museum, London)

1. Вовлечение потенциальных участников

Education

Visualise events, check reconstructed data, run tools or build your own!

Start learning



Research

Get the genuine working environments, virtual machines and datasets to start your research!

Start analysing

2. Выгода для всех участников



Фотография используется с разрешения автора:
Carola Radke / Museum für Naturkunde Berlin

3. Политика открытости

- Использование открытой лицензии, открытой политики



4. Открытая методология

Innovations in scholarly communication - global survey on research tool usage [version 1; referees: awaiting peer review]

✉ Bianca Kramer, Jeroen Bosman

+ Author affiliations

+ Grant information

Abstract

Many new websites and online tools have come into existence to support scholarly communication in all phases of the research workflow. To what extent researchers are using these and more traditional tools has been largely unknown. This 2015-2016 survey aimed to fill that gap. Its results may help decision making by stakeholders supporting researchers and may also help researchers wishing to reflect on their own online workflows. In addition, information on tools usage can inform studies of changing research workflows.

The online survey employed an open, non-probability sample. A largely self-selected group of 20663 researchers, librarians, editors, publishers and other groups involved in research took the survey, which was available in seven languages. The survey was open from May 10, 2015 to February 10, 2016. It captured information on tool usage for 17 research activities, stance towards open access and open science, and expectations of the most important development in scholarly communication. Respondents' demographics included research roles, country of affiliation, research discipline and



Views: 390
Downloads: 82



Get PDF
Get XML
Cite
Export
Tracking
Email
Share

Referee Status: *AWAITING PEER REVIEW*

Comments on this article

All comments (0)

Add a Comment

Sign up for Table of Contents Alerts

Your email address

SIGN UP

Articles that may interest you

RESEARCH ARTICLE



Are scientific abstracts written in poetic verse an effective representation of the underlying research? [version 2; referees: 2 approved with



Источник: <http://f1000research.com/articles/5-692/v1>

5. Открытые научные данные

[Submit data now](#)

[How and why?](#)

Files in this package

Content in the Dryad Digital Repository is offered "as is." By downloading files, you agree to the [Dryad Terms of Service](#). To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data.  

Title	Sci-Hub download data
Downloaded	52 times
Description	These data include 28 million download request events from the server logs of Sci-Hub from 1 September 2015 through 29 February 2016. The uncompressed 2.7 gigabytes of data are separated into 6 data files, one for each month, in MapInfo TAB format.
Download	scihub_data.zip (684.5 Mb)
Details	View File Details

Title IPython Notebook for Sci-Hub raw data

Search for data

Enter keyword, DOI, etc. [Go](#)

[Advanced search](#)

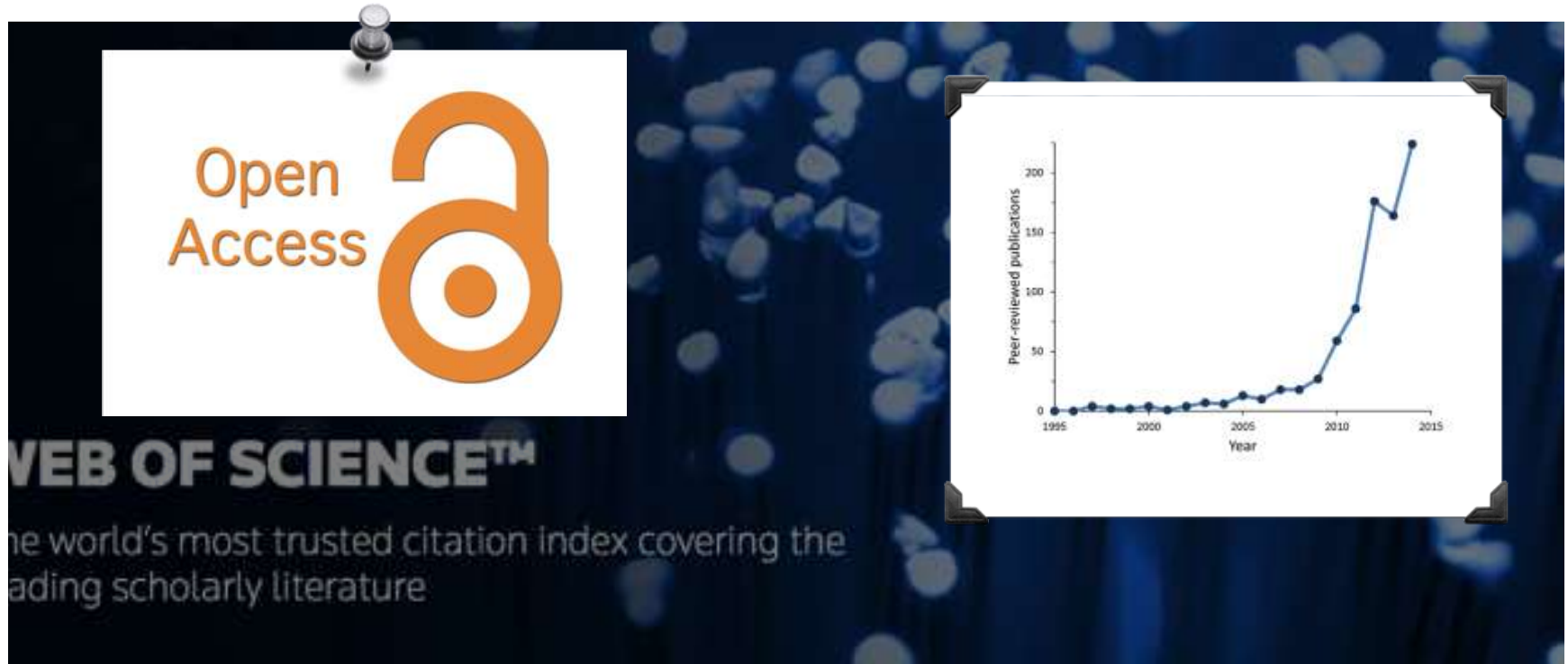
Be part of Dryad

We encourage organizations to:

- [Become a member](#)
- [Sponsor data publishing fees](#)
- [Integrate your journal\(s\), or](#)
- All of the above

Источник: <http://datadryad.org/resource/doi:10.5061/dryad.q447c>

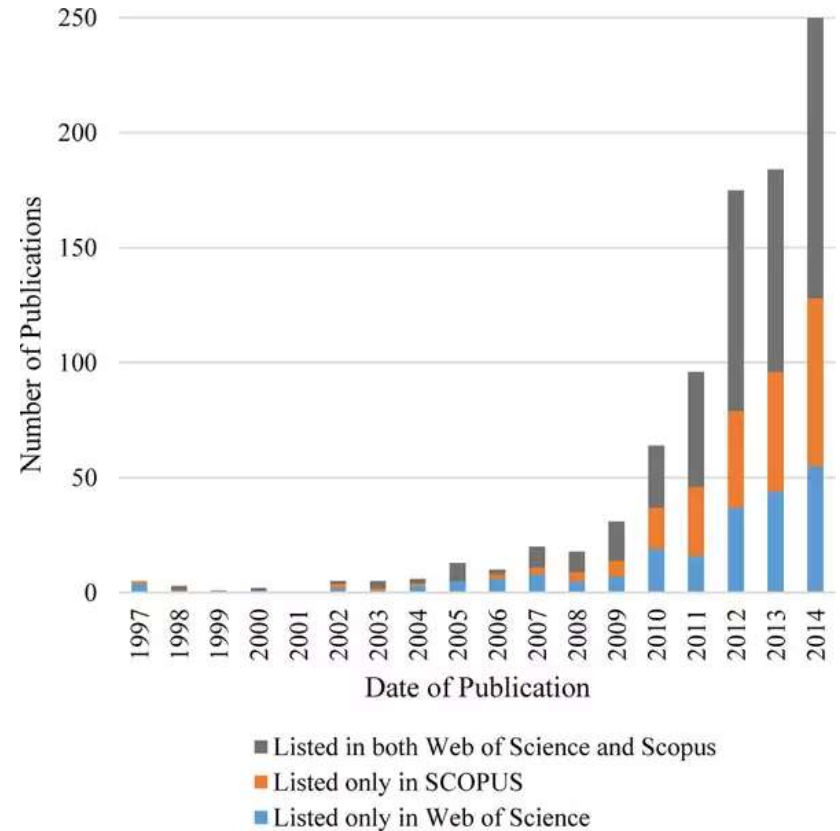
6. Открытый доступ



Источник: <http://www.esa.org/esa/expanding-the-reach-of-environmental-research-with-citizen-science/>

6. Открытый доступ

Figure showing the growth of published peer reviewed articles on citizen science, from 1997 to 2014. Figure from Follet and Strezov, PLOS ONE.



Источник: <http://dx.doi.org/10.1371/journal.pone.0143687.g003>

7. Популяризация науки и распространение культуры

The image shows a screenshot of a Wikipedia article titled "List of citizen science projects". The article is pinned to a background that includes a "WIKI PARTICIPATE" banner on the left and a cartoon illustration of a person on a motorcycle on the right. The article text discusses citizen science projects, mentioning examples like the Old Weather project, EyeWire, and TurtleSAT. It also notes that the list is dynamic and provides a link to a list of 1,100 active projects. A table of active projects is visible at the bottom of the article.

List of citizen science projects

From Wikipedia, the free encyclopedia.

Citizen science projects are activities sponsored by a wide variety of organizations so non-scientists can meaningfully contribute to scientific research. Activities vary widely from transcribing old ship logbooks to digitize the data as part of the Old Weather project to observing and counting birds at home or in the field for eBird.^{[1][2]} Participation can be as simple as playing a computer game for a project called EyeWire that may help scientists learn more about neural neurons.^[3] It can also be more in depth, such as when citizens collect water quality data over time to assess the health of local waters.^[4] An emerging branch of Citizen Science are Community Mapping projects that utilize smartphones and tablet technology. For example, TurtleSAT^[5] is a community mapping project that is mapping freshwater turtle deaths throughout Australia.

Lists of citizen science projects may change. For example, the Old Weather project website indicates that as of January 10, 2015, 51% of the logs were completed.^[6] When that project reaches 100 percent, it will move to the completed list.

There are more than 1,100 active and searchable global citizen science projects listed on [SciStarter](#).^{[7][8][9][10]}

Contents (book)

- 1 Active citizen science projects
- 2 Completed or inactive projects
- 3 See also
- 4 Notes
- 5 References
- 6 External links

Active citizen science projects [edit]

This list is incomplete; you can help by expanding it.

Project name	Discipline(s)	Sponsoring organization(s)	Area	Begin	Web site
AgeDance	Biology, Aging	Center for Research and Interdisciplinarity, University of Southern	Worksite	2012	AgeDance

Источник: https://en.wikipedia.org/wiki/List_of_citizen_science_projects

ЖУРНАЛИСТИКА ДАННЫХ

[ПРАКТИКУМ](#)[ПОЛЕЗНЫЕ МАТЕРИАЛЫ](#)[О ПРОЕКТЕ](#)[ОБ АВТОРАХ](#)[ПРИСОЕДИНЯЙТЕСЬ!](#)[ENGLISH](#)[МЫ НА GITHUBЕ](#)

Экспедиции данных

Что такое Дата-экспедиция?

Экспедиция данных, или дата-экспедиция, — это неформальное общедоступное образовательное онлайн-мероприятие, строящееся на принципах проектного обучения и взаимопомощи. Под "проектным" имеется в виду, что итогом участия в идеале должно стать создание индивидуального или коллективного проекта. Даже если идеала достичь по тем или иным причинам не удастся, каждая экспедиция сама по себе строится как процесс работы над определенным проектом. Говоря о "взаимопомощи", мы исходим из того, что у каждого из участников экспедиции, помимо желания получить результат, есть те или иные навыки и опыт. Будучи у всех людей разными, они могут дополнять друг друга. Чем интенсивнее обмен знаниями между участниками в ходе работы над проектом, тем у всех больше шансов научиться чему-нибудь новому и полезному. С другой стороны, если участник просто наблюдает и следует инструкциям, но не хочет вступать



Data Journalism



АВТОРЫ

[Alina Pinchuk](#)[Alia Rybina](#)[Anastasia Vafeeva](#)

Источник: <http://www.datadrivenjournalism.ru/practicum/data-expeditions/>

Экспедиции данных



 Download .zip

 Download .tar.gz

 View on GitHub

Что такое Дата-экспедиция?

Экспедиция данных, или **дата-экспедиция**, - это неформальное общедоступное образовательное онлайн-мероприятие, строящееся на принципах проектного обучения и взаимопомощи. Под "проектным" имеется в виду, что итогом участия в идеале должно стать создание индивидуального или коллективного проекта. Даже если идеала достичь по тем или иным причинам не удастся, каждая экспедиция сама по себе строится как процесс работы над определенным проектом. Говоря о "взаимопомощи", мы исходим из того, что у каждого из участников экспедиции, помимо желания получить результат, есть те или иные навыки и опыт. Будучи у всех людей разными, они могут дополнять друг друга. Чем интенсивнее обмен знаниями между участниками в ходе работы над проектом, тем у всех больше шансов научиться чему-нибудь новому и полезному. С другой стороны, если участник просто наблюдает и следует инструкциям, но не хочет вступать во взаимодействие с остальными, против этого тоже никто возражать не будет. У каждого свой стиль обучения и работы. Термин «экспедиция данных» заимствован из практики OKF/School of Data (<http://schoolofdata.org/>). Принципы взаимного

Источник: <https://dataexpedition.github.io/DataExpedition/>

(<http://schoolofdata.org/>) и P2P University (<https://p2pu.org/en/>).



Целевые группы дата-экспедиций

- ◎ Студенты
- ◎ Журналисты
- ◎ Чиновники
- ◎ Все заинтересованные в обучении работе с данными



Этапы работы с данными



Источник: <http://www.slideshare.net/iradche/data-expeditions-methodology>

Про открытые данные и светлое будущее

[НОВОСТИ](#)[ПОЛЕЗНЫЕ МАТЕРИАЛЫ](#) ▾[ССЫЛКИ](#) ▾[ПРАКТИЧЕСКИЕ ЗАНЯТИЯ](#) ▾[О САЙТЕ](#)[ОБ АВТОРЕ](#) ▾[ENGLISH](#)

Мои публикации

На русском языке

- [Использование открытых данных в научных исследованиях](#)
- [Школа открытых данных: эффект](#)
- [Связанные данные в университетах](#)
- [О проекте ODI nodes](#)
- [Открытое правительство запускает образовательный курс по работе с открытыми данными](#)

На английском языке

- [Transformation of Communication Processes: Data Journalism](#)
- [On some Russian Educational Projects in Open Data and Data Journalism \(full text\)](#)
- [The View on Open Data and Data Journalism: Cases, Educational Resources and Current Trends \(full text\)](#)
- [Data Expeditions and Data Journalism project as OER in Russian](#)
- [Overview of Russian Open Education Projects](#)
- [Experimenting with Data Expeditions](#)



Categories

[Aaron Swartz](#)[Big Data](#)[Business Model](#)[Cloud Computing](#)[Data](#)[Data Science](#)[Data-Driven Journalism](#)[Data-expedition](#)[Definition](#)[Digital Storytelling](#)[Featured](#)

Спасибо за внимание!



<http://iRadche.ru>

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



 **@iRadche**

 <https://github.com/iRadche>

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

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