

Управление данными

Введение

Радченко Ирина Алексеевна
НИУ ВШЭ
iradche@gmail.com

Создание материала на основе данных



Жизненный цикл данных



На основе: <https://blogs.ntu.edu.sg/lib-datamanagement/data-lifecycle/>

Создание данных

У нас есть два пути:

- * генерировать данные,
- * воспользоваться открытыми данными.

Поиск открытых данных

- * Порталы открытых данных

Порталы открытых данных

DataPortals.org Home Add a Portal Browse Contribute About

Data Portals A Comprehensive List of Open Data Portals from Around the World

551 Data Portals listed »

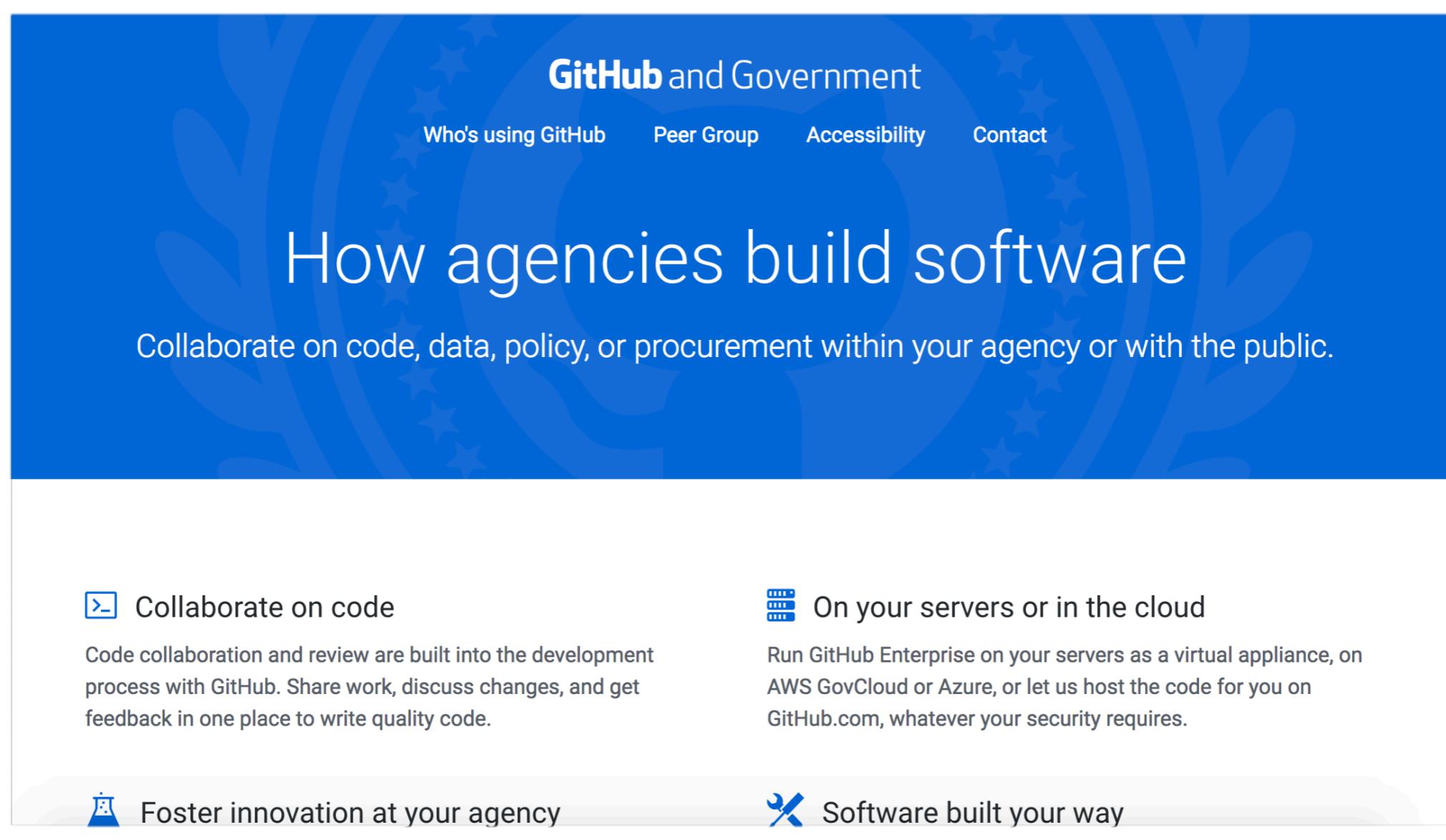
Search data ...

A world map with a legend in the top left corner showing three types of markers: yellow circles for 1-10, green circles for 11-100, and orange circles for 101+. The map shows the following counts by continent:

Continent	Count Range	Number of Portals
North America	1-10	2
North America	11-100	7
North America	101+	59
South America	1-10	10
South America	11-100	13
Europe	1-10	4
Europe	11-100	9
Europe	101+	16
Africa	1-10	2
Africa	11-100	37
Asia	1-10	2
Asia	11-100	9
Asia	101+	29
Australia	1-10	26
Total	All	551

<https://dataportals.org/>

GitHub and Government



The image shows the GitHub and Government landing page. The header features the GitHub logo and the text "GitHub and Government". Below the header are navigation links: "Who's using GitHub", "Peer Group", "Accessibility", and "Contact". The main title "How agencies build software" is displayed prominently, followed by the subtext "Collaborate on code, data, policy, or procurement within your agency or with the public." The page is divided into two columns. The left column contains the section "Collaborate on code" with a description of code collaboration and review. The right column contains the section "On your servers or in the cloud" with a description of GitHub Enterprise deployment options. At the bottom, there are two more sections: "Foster innovation at your agency" and "Software built your way".

GitHub and Government

Who's using GitHub Peer Group Accessibility Contact

How agencies build software

Collaborate on code, data, policy, or procurement within your agency or with the public.

 Collaborate on code

Code collaboration and review are built into the development process with GitHub. Share work, discuss changes, and get feedback in one place to write quality code.

 On your servers or in the cloud

Run GitHub Enterprise on your servers as a virtual appliance, on AWS GovCloud or Azure, or let us host the code for you on GitHub.com, whatever your security requires.

 Foster innovation at your agency

 Software built your way

<https://government.github.com/>

Порталы открытых данных

The screenshot shows the homepage of the data.gov.ru Open Data Portal. The page has a dark purple background featuring a map of Russia. At the top, there is a navigation bar with the Russian coat of arms, a search bar labeled "Поиск по наборам", a language switcher showing "GBR", and a search icon. Below the navigation bar, the portal's logo "data.gov.ru OPEN DATA RUSSIA" is displayed next to a colorful circular icon. The main title "OPEN DATA PORTAL RUSSIA" is centered in large white letters. To the left of the title, there is a vertical column of four blue circular icons with white symbols: a question mark, a lightbulb, a speech bubble, and a document. Below the title, the text "Datasets found: 21397" is shown, along with a button "Add dataset >". To the right of the title, there is a video thumbnail with the text "Video about portal". At the bottom of the page, there are four small boxes containing the numbers 14282, 1451, 1176, and 1053. A red vertical bar on the right side contains the text "Нашли ошибку? Расскажите нам!".

<http://data.gov.ru/>

Порталы открытых данных

The screenshot shows the homepage of the EMIS (Государственная статистика) website. At the top, there is a dark header bar with the logo of the Federal Statistical Service of Russia, the text "Е М И С С" and "ГОСУДАРСТВЕННАЯ СТАТИСТИКА", and three navigation links: "Показатели", "Ведомства", and a "Войти" button. Below the header, a large section titled "Официальные статистические показатели" displays a search bar with a magnifying glass icon and placeholder text "Поиск...", and a link to "Расширенный поиск". To the right of this section are two large numerical boxes: one showing "6135 показателей" and another showing "61 ведомство". At the bottom, there is a box for "Посетителей в день" with the number "2309" and a bar chart, and a "Популярные показатели" section with two items: "Индекс потребительских цен на товары и услуги" (with a value of 33) and "Средние потребительские цены (тарифы) на товары и услуги" (with a value of 29). A "За неделю" dropdown arrow is also visible in this section.

<https://fedstat.ru/>

Порталы открытых данных



<http://stat.gibdd.ru/>

Порталы открытых данных

Russian awesome open data list

Opendata resources in Russian / Открытые данные на русском языке

Этот список пополняется сообществом, для внесения изменений откройте этот файл [для редактирования](#) и сохраните изменения. Они будут оперативно рассмотрены модератором. Если Вы хотите редактировать файл регулярно - напишите ibegin@infoculture или <https://t.me/begin>

Table of contents / Содержание

- [Репозитории](#)
- [Проекты](#)
- [Дата журналистика](#)
- [Госпорталы открытых данных](#)
- [Большие наборы данных](#)
- [Открытый код](#)
- [Awesome Lists](#)
- [Блоги и ресурсы](#)
- [Рассылки](#)
- [Сообщества в Meetup и Timerad в Москве](#)
- [Сообщества в Facebook](#)
- [Сообщества в Vkontakte](#)
- [Каналы в Telegram](#)

<https://github.com/infoculture/awesome-opendata-rus>

Google Dataset Search

Google Dataset Search Бета-версия

Поиск наборов данных



Попробовать [boston education data](#) или [weather site:noaa.gov](#)

<https://toolbox.google.com/datasetsearch>

Kaggle Datasets

kaggle Search kaggle  Competitions Datasets Kernels Discussion Learn ... Sign In

Datasets

Documentation New Dataset

Public Sort by Hotness ▾

10,882 Datasets Sizes ▾ File types ▾ Licenses ▾ Tags ▾ Search datasets 

Rank	Dataset	Description	Tags	File Types	Size	Comments	Views	
114	Google Play Store Apps Web scraped data of 10k Play Store apps for analysing the Android market. Lavanya Gupta updated 7 days ago (Version 5)	video games, computer systems, internet, mobile web	CSV	1.9 MB	Other	15	5	17k
244	Telco Customer Churn Focused customer retention programs BlastChar updated 7 months ago (Version 1)	telecommunications, churn analysis	CSV	171.6 KB	Other	109	3	39k
12	SF Police Calls for Service and Incidents From San Francisco Open Data City of San Francisco Maintained by Kaggle updated 6 hours ago (Version 77)	socrata, crime	Other	164.1 MB	Other	4	0	3k

<https://www.kaggle.com/datasets>

Academic Torrents

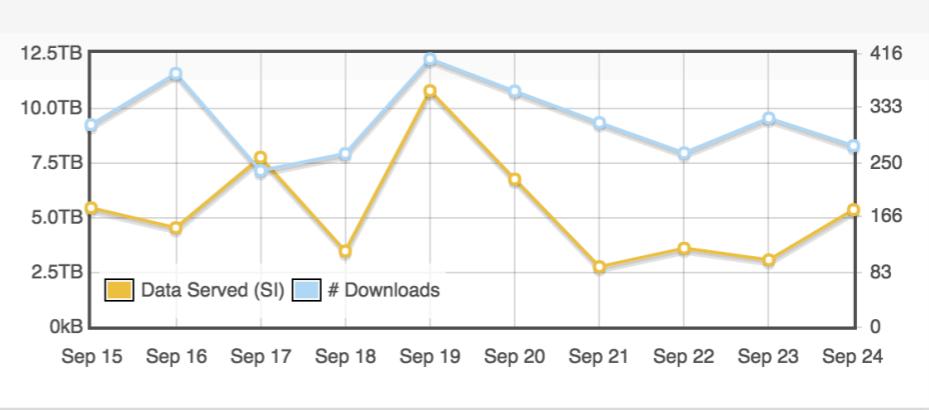
Academic Torrents Browse Upload About Donate Login

paper, author, or dataset Search

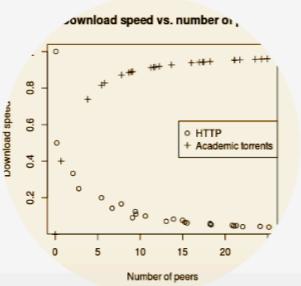
Making 27.91TB of research data available!

We've designed a distributed system for sharing enormous datasets - for researchers, by researchers. The result is a scalable, secure, and fault-tolerant repository for data, with blazing fast download speeds. Contact us at contact@academictorrents.com.

View popular! Upload a dataset!



Date	Data Served (SI)	# Downloads
Sep 15	5.0TB	250
Sep 16	4.5TB	333
Sep 17	7.5TB	416
Sep 18	3.0TB	333
Sep 19	10.5TB	250
Sep 20	6.0TB	333
Sep 21	2.5TB	250
Sep 22	3.5TB	333
Sep 23	2.5TB	416
Sep 24	4.5TB	250



Number of peers	Download speed
0	0.2
5	0.4
10	0.6
15	0.8
20	0.9



Send Feedback

<http://academictorrents.com/>