Open Government

Collaboration, Transparency, and Participation in Practice

Edited by Daniel Lathrop & Laurel Ruma
Foreword by Don Tapscott

O'REILLY®

Politics & Government

Open Government

"Open government is one critical part in making happen what some may think impossible: a government that actually works."

–Lawrence Lessig, director of the Edmond J. Safra Foundation Center for Ethics and professor of law at Harvard Law School

The Web can make real-time data accessible to anyone, but how can government leverage this openness to improve operations and communication, as well as increase citizen participation? In this book, leading visionaries and practitioners both inside and outside of government share their ideas on how to achieve and direct this emerging world of online collaboration, transparency, and participation.

Contributors and topics include:

- **Beth Simone Noveck**, U.S. Deputy Chief Technology Officer for open government, "The Single Point of Failure"
- **Jerry Brito**, senior research fellow at the Mercatus Center at George Mason University, "All Your Data Are Belong to Us: Liberating Government Data"
- **Aaron Swartz**, cofounder of reddit.com, OpenLibrary.org, and BoldProgressives.org, "When Is Transparency Useful?"
- **Ellen S. Miller**, executive director of the Sunlight Foundation, "Disrupting Washington's Golden Rule"
- **Carl Malamud**, founder of Public.Resource.Org, "By the People"
- **Douglas Schuler**, president of the Public Sphere Project, "Online Deliberation and Civic Intelligence"
- **Howard Dierking**, program manager on Microsoft's MSDN and TechNet Web platform team, "Engineering Good Government"
- **Matthew Burton**, Web entrepreneur and former intelligence analyst at the Defense Intelligence Agency, "A Peace Corps for Programmers"
- **Gary D. Bass** and **Sean Moulton**, OMB Watch, "Bringing the Web 2.0 Revolution to Government"
- **Tim O'Reilly**, founder and CEO of O'Reilly Media, "Government As a Platform"

US $44.99 CAN $56.99
ISBN: 978-0-596-80435-0

Safari Books Online
Free online edition for 45 days with purchase of this book. Details on last page.
O'REILLY® oreilly.com

Open Government

Edited by Daniel Lathrop and Laurel Ruma

O'REILLY®
Beijing • Cambridge • Farnham • Köln • Sebastopol • Taipei • Tokyo

Open Government
Edited by Daniel Lathrop and Laurel Ruma

Copyright © 2010 O'Reilly Media, Inc.. All rights reserved.
Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*http://my.safaribooksonline.com*). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

Editors: Laurel Ruma and Julie Steele
Production Editor: Sarah Schneider
Copyeditor: Audrey Doyle
Proofreader: Kiel Van Horn

Indexer: Ellen Troutman Zaig
Cover Designer: Mark Paglietti
Interior Designer: David Futato
Illustrators: Nellie McKesson and Rob Romano

Printing History:
February 2010: First Edition.

O'Reilly and the O'Reilly logo are registered trademarks of O'Reilly Media, Inc. *Open Government*, the image of the Capitol building, and related trade dress are trademarks of O'Reilly Media, Inc.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and O'Reilly Media, Inc. was aware of a trademark claim, the designations have been printed in caps or initial caps.

Wiki Government: How Technology Can Make Government Better, Democracy Stronger, and Citizens More Powerful. Copyright 2009, Beth Simone Noveck. Excerpt reprinted with permission from the Brookings Institution.

While every precaution has been taken in the preparation of this book, the publisher and authors assume no responsibility for errors or omissions, or for damages resulting from the use of the information contained herein.

ISBN: 978-0-596-80435-0

[V]

1265403977

CONTENTS

	FOREWORD *by Don Tapscott*	xv
	PREFACE	xix
1	A PEACE CORPS FOR PROGRAMMERS *by Matthew Burton*	1
	Tipping Point: The Extinction of Pencils	4
	Competition Is Critical to Any Ecosystem	4
	Creating a Developer Corps	6
	Conclusion	8
	About the Author	9
2	GOVERNMENT AS A PLATFORM *by Tim O'Reilly*	11
	Government As a Platform	12
	Lesson 1: Open Standards Spark Innovation and Growth	15
	Lesson 2: Build a Simple System and Let It Evolve	18
	Lesson 3: Design for Participation	22
	A Robustness Principle for Government	25
	Lesson 4: Learn from Your "Hackers"	29
	Lesson 5: Data Mining Allows You to Harness Implicit Participation	32
	Lesson 6: Lower the Barriers to Experimentation	34
	Lesson 7: Lead by Example	36
	Practical Steps for Government Agencies	38
	About the Author	39
3	BY THE PEOPLE *by Carl Malamud*	41
	About the Author	47
4	THE SINGLE POINT OF FAILURE *by Beth Simone Noveck*	49
	The Closed Model of Decision Making	50
	New Technologies and Civic Life	53
	Participatory Democratic Theory in the Age of Networks	58
	About the Author	69
5	ENGINEERING GOOD GOVERNMENT *by Howard Dierking*	71

The Articles of Confederation and the Stovepipe Antipattern 72
Continued Maintenance: The Blob and Confederacy 77
Conclusion 80
About the Author 81

6 ENABLING INNOVATION FOR CIVIC ENGAGEMENT 83
by David G. Robinson, Harlan Yu, and Edward W. Felten

Citizen Initiatives Lead the Way 83
Providing for Reuse and Innovation 84
Data Authenticity Down the Line 87
Why Bother with Bulk? 88
Conclusion 89
About the Authors 89

7 ONLINE DELIBERATION AND CIVIC INTELLIGENCE 91
by Douglas Schuler

Definitions and Assertions 91
Democracy, Deliberation, and the Internet 95
Findings and Issues 100
Conclusion 103
About the Author 104

8 OPEN GOVERNMENT AND OPEN SOCIETY 105
by Archon Fung and David Weil

Transparency's Moment? 105
The Dark Side of Open Government 106
The Missing Diagnosis 108
Targeted Transparency 109
A Matter of Politics 111
Conclusion 111
About the Authors 113

9 "YOU CAN BE THE EYES AND EARS": BARACK OBAMA AND THE WISDOM OF CROWDS 115
by Micah L. Sifry

Change.gov Shows How to Change the Gov 116
"You Can Be the Eyes and Ears" 116
Recovery.gov Site Still Under Construction 117
Online Town Hall or "Participation Theater"? 118
Open Data and Open Government 119
Co-creation, Co-optation, or Collision? 121
About the Author 122

10 TWO-WAY STREET: GOVERNMENT WITH THE PEOPLE 123
by Mark Drapeau

Pockets of Excellence: The Goverati 124
Conclusion 129
About the Author 129

11 CITIZENS' VIEW OF OPEN GOVERNMENT 131
by Brian Reich

The First "We President" 132
The Internet Has Made Us Lazy 133
Toward a Findable Government 135
Advanced Citizenship 136
Conclusion 138
About the Author 138

12 AFTER THE COLLAPSE: OPEN GOVERNMENT AND THE FUTURE OF CIVIL SERVICE 139
by David Eaves

The Coasean Collapse 140
The Long Tail of Public Policy 141
Patch Culture 143
The End of Objectivity 144
Two Preconditions to Government As Platform: Capacity for Self-Organization and Collaboration 146
Extend the Network 148
The Next Civil Service Culture: The Gift Economy 149
Conclusion 150
About the Author 151

13 DEMOCRACY, UNDER EVERYTHING 153
by Sarah Schacht

Many Voices, Many Messages, One Government 153
My Idea 155
Revealing Obscured Government Data 157
Improving Communication without Being Crushed by Email 160
How to Improve Civic Engagement 162
Conclusion 165
About the Author 165

14 EMERGENT DEMOCRACY 167
by Charles Armstrong

Democracy As a Scaling Mechanism 167
Limiting Factors and the Internet 170
Building an Emergent Democracy 171
The Road to Emergent Democracy 175
About the Author 176

15 CASE STUDY: TWEET CONGRESS 177
by Wynn Netherland and Chris McCroskey

Tweet Congress: Build an App, Start a Movement 177
Starting the Movement: We Are All Lobbyists Now 179
So, Who Gets It? 180
Impact 180
Conclusion 182
About the Authors 182

16 ENTREPRENEURIAL INSURGENCY: REPUBLICANS CONNECT WITH THE AMERICAN PEOPLE 183
by Nick Schaper

Entrepreneurial Insurgency and Congress 183
Congress Tweets, Too 184
I YouTube, You YouTube 186
Social Media and the Fight for Transparency 188
Conclusion 190
About the Author 191

17 DISRUPTING WASHINGTON'S GOLDEN RULE 193
by Ellen S. Miller

The Bad Old Days: When Insiders Ruled 195
This Is the Mashable Now 196
What Comes Next 198
About the Author 200

18 CASE STUDY: GOVTRACK.US 201
by Joshua Tauberer

Opening Legislative Data 202
Screen Scraping Congress 205
Engaging the GovTrack Community 210
Conclusion 211
About the Author 211

19 CASE STUDY: FOLLOWTHEMONEY.ORG 213
by Edwin Bender

Accessing Political Donor Data Fraught with Problems 213
The National Institute on Money in State Politics' Role in the Fight for Greater Transparency 214
Bolstering the Spirit of Public Disclosure Laws 216
State-Level Transparency Faces Serious Challenges 217
In an Ideal World: Recommendations for Open Data 219
Conclusion 220
About the Author 221

20 CASE STUDY: MAPLIGHT.ORG 223
by Daniel Newman

Why We Founded MAPLight.org 224
MAPLight.org's Unique Contribution 225
Nuts and Bolts: Using MAPLight.org 227
Barriers to Transparency 231
Conclusion 232
About the Author 232

21 GOING 2.0: WHY OPENSECRETS.ORG OPTED FOR FULL FRONTAL DATA SHARING 233
by Sheila Krumholz

The Decision to Let Go of the Data 233
It's Not Easy Being Open 234

Creating a New Model for Transparency 236
The Future Is Now 237
Conclusion 238
About the Author 239

22 ALL YOUR DATA ARE BELONG TO US: LIBERATING GOVERNMENT DATA 241
by Jerry Brito

Liberating Government Data: Carl Malamud Versus the Man 241
Disclosing Government Data: Paper Versus the Internet 243
Accessing Government Data: Open Distribution Versus Jealous Control 244
Demanding Government Data: Public Money Versus Private Research 245
RECAP: Freeing PACER Documents for Public Use 247
Conclusion 248
About the Author 248

23 CASE STUDY: MANY EYES 249
by Fernanda Viégas and Martin Wattenberg

Policy 249
From Policy to Politicians 252
Visual Literacy 254
Conclusion 256
About the Authors 256

24 MY DATA CAN'T TELL YOU THAT 257
by Bill Allison

The How and Why of Data Collection 258
Federal Data: Approximations Galore 259
Good Data Doesn't Mean Good Results 262
Conclusion 264
About the Author 265

25 WHEN IS TRANSPARENCY USEFUL? 267
by Aaron Swartz

Sharing Documents with the Public 268
Generating Databases for the Public 268
Interpreting Databases for the Public 269
An Alternative 271
About the Author 272

26 TRANSPARENCY INSIDE OUT 273
by Tim Koelkebeck

Complexity Creates Opacity 275
Transparency, Meet Institutional Inertia 276
Kaleidoscope IT: One-Off Apps Obscure Information 278
A Market Focused on Proposals, Not Products 280
Framing the Window 284
Conclusion 286
About the Author 287

27 BRINGING THE WEB 2.0 REVOLUTION TO GOVERNMENT 289
by Gary D. Bass and Sean Moulton

Government Transparency: Three Hurdles 289
Putting It All Together: Disclosure of Federal Spending 297
Conclusion 303
About the Authors 304

28 TOADS ON THE ROAD TO OPEN GOVERNMENT DATA 305
by Bill Schrier

What Is Government? 305
Data Collection 306
Exposing the Soul of Government 307
Conclusion 313
About the Author 314

29 OPEN GOVERNMENT: THE PRIVACY IMPERATIVE 315
by Jeff Jonas and Jim Harper

Privacy-Enhancing Practices 316
Conclusion 324
About the Authors 325

30 FREEDOM OF INFORMATION ACTS: PROMISES AND REALITIES 327
by Brant Houston

The Act and Amendments 328
Conclusion 335
About the Author 336

31 GOV→MEDIA→PEOPLE 337
by Dan Gillmor

Crowdsourcing in Action 340
Conclusion 344
About the Author 344

32 OPEN SOURCE SOFTWARE FOR OPEN GOVERNMENT AGENCIES 345
by Carlo Daffara and Jesus M. Gonzalez-Barahona

Advantages of FLOSS for Government and Public Agencies 346
Best Practices: Management 350
Best Practices: Technical 354
Best Practices: Social 357
Make It Easy to Experiment and Learn 359
Conclusion 360
References 360
About the Authors 361

33 WHY OPEN DIGITAL STANDARDS MATTER IN GOVERNMENT 363
by Marco Fioretti

Badly Used Technology Hinders Progress 364

The Digital Age Explained 364
Standards and the Problems with Digital Technology 366
The Huge Positive Potential of Digital Technologies 369
Free and Open Standards and Software: The Digital Basis of Open Government 370
Conclusion 372
About the Author 373

34 CASE STUDY: UTAH.GOV 375
by David Fletcher

A Historical Perspective 375
What Today's Landscape Looks Like 376
Champions Discovered in All Branches of State Government 378
The Dramatic Shift to Web 2.0 Principles and Tools 381
Making Data More Accessible 383
Conclusion 387
About the Author 388

A MEMO FROM PRESIDENT OBAMA ON TRANSPARENCY AND OPEN GOVERNMENT 389

INDEX 391

CHAPTER TWENTY-TWO

All Your Data Are Belong to Us: Liberating Government Data

Jerry Brito

When government refuses to make itself transparent and open and fails to make public information meaningfully available, hackers will liberate the data. It has happened many times over, and it will doubtlessly happen again. Each time government data is freed, citizens gain useful access to valuable information that rightly belongs to them. But perhaps more importantly, government is forced to deal with the new reality of a networked world in which the people demand free online access to public information.

Data is liberated by hacking government. Because of how the popular press has used it, the word *hack* is often misunderstood to mean only illicit access to computer networks. In fact, to techies that is only one possible meaning. According to Wikipedia, the term usually means "a clever or quick fix to a computer program problem" or "a modification of a program or device to give the user access to features that were otherwise unavailable to them."

Liberating Government Data: Carl Malamud Versus the Man

Carl Malamud is probably the original government hacker. A technologist, activist, and entrepreneur, Malamud is well known for having forced the U.S. Securities and Exchange Commission (SEC) to make much of its public data available to the public over the Internet in 1995. At that time, the SEC did not provide free access to the corporate filings it collected. Instead, the SEC's database, the Electronic Data Gathering Analysis and Retrieval system

(EDGAR), was operated under contract with information wholesaler Mead Data, which provided data feeds to data retailers who in turn sold access to the public. According to John Markoff, author of the *New York Times* article "Plan Opens More Data to Public", "Under this system, a retail information provider, like Mead Data's own Nexis service, charge[d] about $15 for each S.E.C. document, plus a connection charge of $39 an hour and a printing charge of about $1 a page." One can imagine customers were largely limited to firms on Wall Street.

After first trying and failing to convince the SEC that it should make its database available on the Internet, Malamud began to purchase the SEC's wholesale data and made it available on his own website free of charge to anyone. The service included corporate annual reports, 10-K filings, proxy statements, and other data valuable to investors, journalists, and others. In December of that year, Malamud expanded his free offerings by adding large portions of the U.S. Patent and Trademark Office's (PTO) patent and trademark database, including full text of all patents, and text and images from the trademark database.

Malamud, however, always believed that it was government's responsibility to provide its data for free to the public, especially since the then-recently enacted Paperwork Reduction Act mandated that agencies make public information available electronically. On August 11, 1995, Malamud announced on his website that it would discontinue its free access to government data on October 1. As Malamud later recounted:

> Our goal, however, wasn't to be in the database business. Our goal was to have the SEC serve their own data on the Internet. After we built up our user base, I decided it was time to force the issue. That's when the fireworks began. When users visited our EDGAR system in August 1995, they got an interesting message:
>
> This Service Will Terminate in 60 Days Click Here For More Information
>
> Click here they did! One of the lessons I've learned from building Internet services is that when people get something for free, they want their money's worth.

The message informed users that under no circumstances would the unofficial service be continued, and suggested that it was "time for the stakeholders in this data to step up to the plate and forge a solution." It also asked users to write to Congress and the administration, which they did.

The SEC at first resisted. Eventually, however, it relented and the agency took over Malamud's service as the core of a new online EDGAR system. The public uproar apparently caught SEC commissioners off guard and they took on the responsibility of making data available before Malamud's October deadline. According to Malamud, "The commissioners of the SEC had clearly not been aware of the issue, but there is nothing like pieces in the *Wall Street Journal* and 15,000 messages to the Chairman to raise the profile of an issue" (*http://www.mundi.net/cartography/EDGAR*).

Malamud had similar plans for his patent and trademark database. In 1998, he wrote to Vice President Al Gore and Commerce Secretary William Daley (who oversaw the PTO),

announcing that unless the PTO began offering its databases on the Internet, he would create a free and robust online alternative. In the years since Malamud first put a patent database online in 1994, the PTO had not been as accommodating as the SEC. The agency was self-financed by user fees, a large portion of which came from requests for paper copies of patent and trademark information. As the commissioner for patents told the *New York Times*, "If he can [put the patent and trademark database online] we'd be out all $20 million we now receive in fees…. Why would anyone want paper?"

Malamud's strategy to overcome government's resistance was a familiar one. "I'm going to buy the trademark data and will build the user base as big as I can in a year," Malamud said at the time. "At the end of the year, I'll pull the rug out from the users and give them Al Gore's E-mail address" (*http://www.nytimes.com/1998/05/04/business/us-is-urged-to-offer-more-data-on-line.html*). The gambit worked, and less than two months later the Clinton administration announced that it would put the full patent database online. In each instance, by forcefully but legally releasing online data that the government had either not disclosed on the Internet or not made easily accessible, Malamud was able to effect a change in policy that led to a more open and transparent government (see Chapter 14).

Disclosing Government Data: Paper Versus the Internet

The United States is one of the most open and transparent countries in the world. Citizens generally have the right to inspect the records, minutes, balance sheets, and votes of almost all public bodies.

Laws encouraging government transparency and accountability have been a feature of the U.S. system of government since the founding of the Republic. The Constitution, for example, requires that each house of Congress "keep a Journal of its Proceedings, and from time to time publish the same, excepting such Parts as may in their Judgment require Secrecy." Today, the Congressional Record satisfies this requirement.

Unfortunately, many of the statutory requirements for disclosure do not take Internet technology into account. For example, the 1978 Ethics in Government Act requires the disclosure of financial information—including the source, type, and amount of income—by many federal employees, elected officials, and candidates for office, including the president and vice president, and members of Congress. The act further requires that all filings be available to the public, subject to certain limited exceptions. One might imagine, then, that every representative's or senator's information would be just a web search away, but one would be wrong.

Members of the House of Representatives must file their disclosures with the clerk of the House of Representatives, while senators must do the same with the secretary of the Senate. Each of these offices maintains a searchable electronic database of the filings. However, until very recently, to access these databases citizens had to go to Washington, D.C., and visit those Capitol Hill offices during business hours. There were no other means of searching the

databases, something that presented a major barrier to the widespread dissemination of nominally publicly available information. Even today, the disclosure forms offered are scanned images that are not easily searched or parsed.

The result is that to make the information available online, third parties such as transparency website LegiStorm must acquire paper copies of the forms and manually scan and parse them. In contrast, the clerk of the House and the secretary of the Senate could likely make their existing databases available online at little extra cost. LegiStorm also offers something the official sites still won't: financial disclosure forms for congressional staffers, not just members.

So, why would government fail to take advantage of the benefits of online disclosure? Not only would using standard Internet technologies make it easier for citizens to find and access government information, but it would probably also present efficiencies and cost savings to government itself. In most cases, the obstacle is likely bureaucratic inertia. In other cases, however, government will have little incentive, and often a disincentive, to make public information easily accessible.

Sometimes, as we saw with the PTO, government agencies make data freely available, but collect user fees for easy access to that data. This can create an incentive to protect the revenue stream at the expense of wider public access to government information. In other instances, government reticence to make data easily accessible can have political motivations.

Accessing Government Data: Open Distribution Versus Jealous Control

Much like Josh Tauberer's GovTrack.us, the *Washington Post*'s Congress Votes database allows users to easily search and sort through a database of congressional bills and votes. When the *Post* was building its site, the House offered its roll call votes in XML, a standard machine-readable format, while the Senate did not. This forced the *Post*, like GovTrack, to rely on cumbersome "screen scraping" of Senate web pages to make their roll call votes usable (see Chapter 18).

In 2007, however, co-creator Derek Willis was poking around the Senate website when he discovered a directory of XML files of vote data for past sessions. This demonstrated that the Senate had the ability to make its votes available online in a structured format. Willis was elated at the thought that perhaps there was easy access to Senate vote data after all. He wrote to the Senate webmaster asking whether structured voting data was available for the current session and, if so, whether this data would be made public. The telling response read, in part, as follows:

> A few representative votes (only a few from the early congresses) were published out to the active site during some testing periods. I really need to remove them from the site.
>
> We are not authorized to publish the XML structured vote information. The Committee on Rules and Administration has authorized us to publish vote tally information in HTML format [not a

> structured format]. Senators prefer to be the ones to publish their own voting records. As you know, looking at a series of vote results by Senator or by subject does not tell the whole story. Senators have a right to present and comment on their votes to their constituents in the manner they prefer. This issue was reviewed again recently and the policy did not change.

Senators doubtlessly would "prefer to be the ones to publish their own voting records." But jealous control over information by government is anathema to democracy. Looking at a series of votes by a senator does in fact tell the "whole story" of that senator's voting record, and despite what the webmaster said, senators do not have a "right" to present their votes to the public "in the manner they prefer." Of course, this only motivated hackers such as Willis and Tauberer further.

When third parties make government data available, it demonstrates that it is possible to do so cheaply and efficiently. In some cases, officials can be unaware of what is technically possible, or they may believe that state-of-the-art technology is prohibitively expensive. Freeing information can also generate an awareness and demand for the newly accessible data among citizens. This can lead to embarrassing questions for government: why isn't it making the data available itself? Why are citizens forced to hack the data in order to access it? Also, hacking government data can demonstrate to cautious officials that when information is made accessible and useful, the world does not end.

In fact, since GovTrack.us and the *Washington Post* Congress Votes database brought attention to the issue, the Senate Rules Committee finally relented and has recently begun to make roll call votes available in XML. Two years after Derek Willis was rebuffed by the Senate webmaster, a group of seemingly embarrassed senators wrote to Committee Chairman Chuck Schumer demanding a repeal of the prohibition on XML.

"This policy has created a situation where outside groups are forced to create databases that are more likely to contain errors and omissions," they wrote. "The suggestion that the Senate would intentionally hamstring the distribution of roll call votes so Senators could put a better spin on them is concerning. The public is capable of interpreting our votes on its own."

Demanding Government Data: Public Money Versus Private Research

The release of Congressional Research Service (CRS) reports is another example of hacking that is slowly leading to change. CRS is a think tank for Congress that is funded by U.S. taxpayers to the tune of $100 million per year. It produces objective in-depth briefings and high-quality research papers on topical public policy issues. The studies it produces are widely well regarded, and members of Congress and their staffs rely on them as they legislate.

By law, however, CRS can make its reports available only to members of Congress. There are several oft-cited rationales for this policy. First is the idea that by releasing reports directly to the public, CRS would come between lawmakers and their constituents. There is also the

concern that if CRS reports were widely disseminated, they would come to be written with a public audience in mind, rather than focusing on congressional needs. Finally, some fear a burden on CRS and congressional staff members who might have to respond to questions and comments generated by publicly available reports. Not surprisingly, several third parties who are not persuaded by these rationales have taken it upon themselves to collect and make the reports available on the Web.

Members of Congress are free to release copies of the reports to citizens if they wish, and they often do so at the request of a constituent. Unfortunately, constituents must first know of a report's existence before they can request it. There is no public listing of all the available titles, so the fear is that embarrassing reports are never released. Once a report is released, however, one is free to copy and disseminate it because works of the U.S. government are not protected by copyright.

For several years, many organizations, including the Federation of American Scientists and the National Council for Science and the Environment, have published on their websites hundreds of CRS reports related to their research areas that they have collected over the years. In 2005, the Center for Democracy and Technology (CDT), a nonprofit dedicated to Internet public policy issues, brought the different CRS collections under one roof at OpenCRS.com. There, users can search the combined collections of CDT and partner groups, which total more than 8,000 CRS reports.

More importantly, the site invites users to upload to the online library reports that they have acquired. It also provides a list of CRS reports that are missing from the library and instructions on how citizens can request them from their representatives. CDT acquires this list from a sympathetic but anonymous member of Congress who provides it on a regular basis. Further compromising the secrecy surrounding CRS reports, in February 2009 whistleblower site WikiLeaks.org released 6,780 CRS reports, which it said "represents the total output of the Congressional Research Service electronically available to Congressional offices" (*http://wikileaks.org/wiki/Change_you_can_download:_a_billion_in_secret_Congressional_reports*).

For more than a decade, lawmakers sympathetic to open government have perennially introduced bills or resolutions to make CRS reports public. While these efforts have so far failed, the vast number of CRS reports now available to the public on third-party sites undercut the rationales for a policy of selective release. Citizens have access to a wide array of CRS reports, yet the quality of those reports has not suffered, CRS's institutional character has not been diminished, and constituent relationships with representatives remain intact. As unofficial collections of CRS reports continue to grow, not only will citizens benefit from access to this information, but the selective release policy will become increasingly untenable and ripe for change.

RECAP: Freeing PACER Documents for Public Use

As long as some public documents are out of reach online, third parties will be motivated to free them. Carl Malamud's most recent intellectual descendants may be a team of hacker scholars dedicated to liberating the millions of pages of public records now locked behind a paywall on the federal court's online database. Stephen Schultze, Tim Lee, and Harlan Yu of the Center for Information Technology Policy at Princeton developed a web browser plug-in—RECAP—that distributes the hacking of the court database among many users.

Each day, federal courts around the country generate thousands of pages of court filings, transcripts, judgments, and opinions—all of public interest. Free access to these documents is guaranteed to all citizens, so long as you visit the courthouse in question during business hours. The Public Access to Court Electronic Records (PACER) system was created in 1988 as a dial-up service that charged by the minute and afforded attorneys convenient remote access to court dockets and filings. In 1998, the system was migrated to the Web, and the per minute charge was replaced by a charge per page downloaded.

Today, PACER charges 8 cents per page accessed, which doesn't sound like much. However, the charge is completely out of proportion with how much it costs the court system to make the data available. As a result, the court system has found its IT budget with a substantial surplus. Some estimate that in 2006, the cost to maintain PACER was about $27 million, yet the system brought in $62 million in revenue.

This is not a concern for lawyers, who pass their costs on to clients, or to commercial data retailers, such as Westlaw and LexisNexis, that purchase their data in bulk and benefit from the lack of convenient free access to court records. However, scholars, journalists, and average citizens are left without free online access to court records.

Schultze, Lee, and Yu's scheme to free the documents on PACER is an ingenious one. They have built a Firefox plug-in called RECAP that attorneys, librarians, and other regular users of PACER can install on their computers. When a user downloads a document from PACER, the plug-in sends a copy to RECAP's server, where it is made publicly available. If enough PACER users install RECAP, it will only be a matter of time before the entire database is liberated.

This is a brilliant system, but it raises the obvious question: altruistic motives aside, why would attorneys or other PACER users install the RECAP plug-in? The answer is just as brilliant. When a user clicks to download a document from PACER, the RECAP plug-in first searches its own server to see whether that document has already been made publicly available. If it has, then it is served to the user, who avoids PACER's download charges. Additionally, the RECAP plug-in adds some nice touches, including reformatting the document filename and adding useful metadata.

Although it is a shot across the bow of PACER's revenue stream, RECAP is entirely legal. The court documents that RECAP shares, like CRS reports, are not subject to copyright and can be freely distributed by citizens. By creating a public repository of these documents, the RECAP

team is forcing the issue and making the court system implement legal or technical countermeasures, or seek a new business model that includes public access to its documents.

Conclusion

This tough-love approach is a necessary counterpart to a public education and advocacy strategy for realizing a more open government. While we must convince citizens and government officials of the merits of transparency, talented hackers can often simply show them. When this happens the burden of proof is shifted to government. And, in turn, it's difficult to make the case that more, better, and easier access to public data isn't a good thing.

About the Author

Jerry Brito is a senior research fellow at the Mercatus Center at George Mason University, and is director of its Technology Policy Program. He also serves as adjunct professor of law at George Mason University. His research focuses on technology and telecommunications policy, government transparency and accountability, and the regulatory process. Jerry is the creator of OpenRegs.com, an alternative interface to the federal government's regulatory docketing system, and the cocreator of the accountability website Stimulus Watch.