Policy Analysis

Grading the Government's Data Publication Practices

by Jim Harper

Executive Summary

Barack Obama promised transparency and open government when he campaigned for president in 2008, and he took office aiming to deliver it. Today, the federal government is not transparent, and government transparency has not improved materially since the beginning of President Obama's administration. This is not due to lack of interest or effort, though. Along with meeting political forces greater than his promises, the Obama transparency tailspin was a product of failure to apprehend what transparency is and how it is produced.

A variety of good data publication practices can help produce government transparency: authoritative sourcing, availability, machine-discoverability, and machine-readability. The Cato Institute has modeled what data the government should publish in the areas of legislative process and budgeting, spending, and appropriating. The administration and the Congress both receive fairly low marks under systematic examination of their data publication practices.

Between the Obama administration and House Republicans, the former, starting from a low transparency baseline, made extravagant promises and put significant effort into the project of government transparency. It has not been a success. House Republicans, who manage a far smaller segment of the government, started from a higher transparency baseline, made modest promises, and have taken limited steps to execute on those promises. President Obama lags behind House Republicans, but both have a long way to go.

Jim Harper is director of information policy studies at the Cato Institute and the webmaster of WashingtonWatch. com.

CATO INSTITUTE

Introduction

As a campaigner in 2008, President Obama promised voters hope, change, and transparency.[1] Within minutes of his taking office on January 20, 2009, in fact, the Whitehouse.gov website declared: "President Obama has committed to making his administration the most open and transparent in history."[2] His first presidential memorandum, issued the next day, was entitled "Transparency and Open Government." It declared:

> My Administration is committed to creating an unprecedented level of openness in Government. We will work together to ensure the public trust and establish a system of transparency, public participation, and collaboration. Openness will strengthen our democracy and promote efficiency and effectiveness in Government.[3]

The road to government transparency is long. Nearly four years later, few would argue that American democracy has materially strengthened, or that the government is any more effective and efficient, due to forward strides in transparency and openness. Indeed, the administration has come under fire recently—as every administration does, it seems—for significant transparency failings.

Freedom of Information Act (FOIA) policy is an example. In its early days, the Obama administration committed to improving the government's FOIA practices. In March 2009 Attorney General Eric Holder issued a widely lauded memorandum ordering improvements in FOIA compliance.[4] But this September, Bloomberg news reported on its test of the Obama Administration's commitment to transparency under FOIA. Bloomberg found that 19 of 20 cabinet-level agencies disobeyed the public disclosure law when it asked for information about the cost of agency leaders' travel. Just 8 of 57 federal agencies met Bloomberg's request for documents within the 20-day disclosure window required by the act.[5]

President Obama's campaign promise to post laws to the White House website for five days of public comment before he signed them went virtually ignored by the White House in the first year of his administration. Only recently has he reached two-thirds compliance with the "Sunlight Before Signing" promise, and this is because of the multitude of bills Congress passes to rename post offices and such. More important bills are often given less than the promised five days' sunlight.[6]

There was no lack of effort or creativity around data transparency at the outset of the Obama Administration. In May 2009 White House officials announced on the new *Open Government Initiative* blog that they would elicit the public's input into the formulation of its transparency policies. In a meta-transparency flourish, the public was invited to join in with the brainstorming, discussion, and drafting of the government's policies.[7]

The conspicuously transparent, participatory, and collaborative process contributed something, evidently, to an "Open Government Directive," issued in December 2009 by Office of Management and Budget head Peter Orszag.[8] Its clear focus was to give the public access to data. The directive ordered agencies to publish within 45 days at least three previously unavailable "high-value data sets" online in an open format and to register them with the federal government's data portal, Data.gov. Each agency was to create an "Open Government Webpage" as a gateway to agency activities related to the Open Government Directive.

Many, many of President Obama's transparency promises went by the wayside. His guarantee that health care legislation would be negotiated "around a big table" and televised on C-SPAN was quite nearly the opposite of what occurred.[9] People are free to observe whether it is political immaturity, idealism, or dishonesty that prompted transparency promises of this kind. Whatever the

case, history may show that the "high-value data set" challenge was where the Obama Administration's data transparency effort began its tailspin.

Celebrated though it is, transparency is not a well-defined concept, and the administration's most concerted effort to deliver it missed the mark. The reason is that the definition of "high-value data set" it adopted was hopelessly vague:

> High-value information is information that can be used to increase agency accountability and responsiveness; improve public knowledge of the agency and its operations; further the core mission of the agency; create economic opportunity; or respond to need and demand as identified through public consultation.

Essentially anything agencies wanted to publish they could publish claiming "high value" for it.

Agencies "adopted a passive-aggressive attitude" toward the Data.gov effort, according to political scientist Alon Peled.[10] They technically complied with the requirements of the Open Government Memorandum, but did not select data that the public valued.

The Open Government Directive allowed agencies to exploit a subtle "shift in vocabulary" in the area of open government. They diverted the project away from the core government transparency that the public found so attractive about President Obama's campaign claims. "The term 'open government data' might refer to data that makes the government as a whole more open (that is, more publicly accountable)," write Harlan Yu and David Robinson, "or instead might refer to politically neutral public sector disclosures that are easy to reuse, even if they have nothing to do with public accountability."[11]

The Agriculture department published data about the race, ethnicity, and gender of farm operators, for example, rather than about the funds it spent to collect that kind of information. An informal Cato Institute study examining agencies' "high-value" data feeds found, "almost uniformly, the agencies came up with interesting data—but 'interesting' is in the eye of the beholder. And interesting data collected by an agency doesn't necessarily give the insight into government we were looking for."[12]

Genuinely high-value data for purposes of government transparency would provide insight in three areas not found in many of the early Data.gov feeds. True high-value data would be about government entities' management, deliberations, or results.[13]

"Open data can be a powerful force for public accountability," write Yu and Robinson, "It can make existing information easier to analyze, process, and combine than ever before, allowing a new level of public scrutiny."[14] This is undoubtedly true, and Americans have experienced vastly increased access to information in so many walks of life—shopping, news-gathering, and investments, to name just three. Data-starved public oversight of government appears sorely lacking in comparison.

In September a new transparency-related international initiative took center stage for the administration, the Open Government Partnership (OGP).[15] This "multilateral initiative" was created "to promote transparency, fight corruption, strengthen accountability, and empower citizens."[16] Participating countries pledged "to undertake meaningful new steps as part of a concrete action plan, developed and implemented in close consultation with their citizens." The OGP website touts a panoply of meetings, plans, and social media outreach efforts, and a recent graphic displayed on the home page said in bold letters, "From Commitment to Action." Its authors probably have no sense of the irony in that declaration. Significant actions, after all, announce themselves.

Nothing about the OGP is harmful, and it may produce genuine gains for openness in participating countries. However, it has not produced, and does not hold out, the fundamental change—data-oriented change—that

Celebrated though it is, transparency is not a well-defined concept.

was at the heart of President Obama's campaign promises.

The Obama administration is not the only actor on the federal stage, of course. House Republicans made transparency promises of their own in the course of their campaign to retake control of the House of Representatives, which they did in 2011.

"The lack of transparency in Congress has been a problem for generations, under majorities Republican and Democrat alike," said aspiring House speaker John Boehner (R-OH) in late 2009. "But with the advent of the Internet, it's time for this to change."[17]

Since 1995, the Library of Congress's THOMAS website has published information, sometimes in the form of useful data, about Congress and its activities. Upon taking control of the House for the first time in 40 years, the Republican leadership of the 104th Congress directed the Library of Congress to make federal legislative information freely available to the public. The offerings on the site now include bills, resolutions, activity in Congress, the *Congressional Record*, schedules, calendars, committee information, the president's nominations, and treaties.[18]

In an attempt to improve the availability of key information, at the beginning of the 112th Congress the House instituted a rule—not always complied with—that bills should be posted online for three calendar days before receiving a vote on the House floor.[19] The House followed up by creating a site at data.house.gov where such bills are posted. In February 2012 the House Committee on Administration held a day-long conference on legislative data,[20] evidence of continuing interest and of plans to move forward. And in September, the Library of Congress debuted beta.congress.gov, which is slated to be the repository for legislative data that ultimately replaces the THOMAS website.[21]

Between the Obama administration and House Republicans, the former, starting from a low transparency baseline, made extravagant promises and put significant ef-

fort into the project of government transparency. It has not been a success. House Republicans, who manage a far smaller segment of the government, started from a higher transparency baseline, made modest promises, and have taken limited steps to execute those promises.

The transparency problem is far from solved, of course. The information that the public would use to increase their oversight and participation is still largely inaccessible. The Republican House may be ahead, but both the administration and Congress score poorly under systematic examination of their data publication practices.

The Data that Would Make for Transparent Government

It was not disinterest that caused the Obama administration transparency effort to fade. Arguably, it was the failure of the transparency community to ask clearly for what it wants: good data about the deliberations, management, and results of government entities and agencies. So in January 2011 the Cato Institute began working with a wide variety of groups and advisers to "model" governmental processes as data and then to prescribe how this data should be published.

Data modeling is arcane stuff, but it is worth understanding here at the dawn of the Information Age. "Data" is collected abstract representations of things in the world. We use the number "3," for example, to reduce a quantity of things to an abstract, useful form—an item of data. Because clerks can use numbers to list the quantities of fruits and vegetables on hand, store managers can effectively carry out their purchasing, pricing, and selling instead of spending all of their time checking for themselves how much of everything there is. Data makes everything in life a little easier and more efficient for everyone.

Legislative and budgetary processes are not a grocery store's produce department, of course. They are complex activities involving many actors, organizations, and steps. The Cato Institute's modeling of these processes

The transparency problem is far from solved.

4

reduced everything to "entities," each having various "properties." The entities and their properties describe the things in legislative and budgetary processes and the logical relationships among them, like members of Congress, the bills they introduce, hearings on the bills, amendments, votes, and so on. The "entity" and "property" terminology corresponds with usage in the world of data management, it is used to make coding easier for people in that field, and it helps to resolve ambiguities in translating governmental processes into useful data. The modeling was restricted to formal parts of the processes, excluding, for example, the varied organizations that try to exert influence, informal communications among members of Congress, and so on.

The project also loosely defined several "markup types," guides for how documents that come out of the legislative process should be structured and published to maximize their utility. The models and markup types are discussed in a pair of *Cato@Liberty* blog posts that also issued preliminary grades on the quality of data publication about the entities.[22] The models and markup types for legislative data and budgeting/appropriations/spending data can be found in Appendixes A and B, respectively.

Next, the project examined the publication methods that allow data to reach its highest and best use. Four key data practices that support government transparency emerged. Documented in a Cato Institute Briefing Paper entitled "Publication Practices for Transparent Government,"[23] those practices are authoritative sourcing, availability, machine-discoverability, and machine-readability.

Authoritative sourcing means producing data as near to its original source and time as possible, so that the public uniformly comes to rely on the best sources of data. The second transparent data practice, availability, entails consistency and confidence in data, including permanence, completeness, and good updating practices.

The third transparent data practice, machine-discoverability, occurs when information is arranged so that a computer can discover the data and follow linkages among it. Machine-discoverability exists when data is presented consistently with a host of customs about how data is identified and referenced, the naming of documents and files, the protocols for communicating data, and the organization of data within files.

The fourth transparent data practice, machine-readability, is the heart of transparency because it allows the many meanings of data to be discovered. Machine-readable data is logically structured so that computers can automatically generate the myriad stories that the data has to tell and put it to the hundreds of uses the public would make of it in government oversight. A common and popular language for structuring and containing data is called XML, or eXtensible Markup Language, which is a relative of HTML (hypertext markup language), the language that underlies the World Wide Web.

Beginning in September 2011 the project graded how well Congress and the administration publish data about the key entities in the processes they oversee. Congress is responsible for data pertaining to the legislative process, of course. The administration has the bulk of the responsibility for budget-related data (except for the congressional budgets and appropriations). These grades are available in a pair of *Cato@Liberty* blog posts[24] and in Appendixes C and D.

With the experience of the past year, the project returned to grading in September 2012. With input from staff at GovTrack.us, the National Priorities Project, OMB Watch, and the Sunlight Foundation (their endorsement of the grades not implied by their assistance), we assessed how well data is now published. The grades presented in Figures 1 and 2 are largely consistent with the prior year—little changed between the two grading periods—but there were some changes in grades in both directions due to improvements in publication, discovery of data sources by our panel of graders, and

Four key data practices support government transparency: authoritative sourcing, availability, machine-discoverability, and machine-readability.

heightened expectations. "Incompletes" given in the first year of grading became Fs in some cases and Ds in others.

It is important to highlight that grades are a lagging indicator. Transparency is not just a product of good data publication, but also of the society's ability to digest and use information. Once data feeds are published, it takes a little while for the community of users to find them and make use of them. A new web site dedicated to congressional information, beta.congress.gov, will undoubtedly improve data transparency and the grades for data it publishes, assuming it lives up to expectations.

Government transparency is a widely agreed-upon value, sought after as a means toward various ends. Libertarians and conservatives support transparency because of their belief that it will expose waste and bloat in government. If the public understands the workings and failings of government better, the demand for government solutions will fall and democracy will produce more libertarian outcomes. American liberals and progressives support transparency because they believe it will validate and strengthen government programs. Transparency will root out corruption and produce better outcomes, winning the public's affection and support for government.

Though the goals may differ, pan-ideological agreement on transparency can remain. Libertarians should not prefer large government programs that are failing. If transparency makes government work better, that is preferable to government working poorly. If the libertarian vision prevails, on the other hand, and transparency produces demand for less government and greater private authority, that will be a result of democratic decisionmaking that liberals and progressives should respect and honor.

With that, here are the major entities in the legislative process and in budgeting, appropriating, and spending; the grades that reflect the quality of the data published about them; and a discussion of both.

> **Government transparency is a widely agreed-upon value, sought after as a means toward various ends.**

Publication Practices for Transparent Government: Rating Congress

House Membership: C-
Senate Membership: A-

It would seem simple enough to publish data about who holds office in the House of Representatives and Senate, and it is. There are problems with the way the data is published, though, which the House and Senate could easily remedy.

On the positive side—and this is not to be discounted—there is a thing called the "Biographical Directory of the United States Congress," a compendium of information about all present and former members of the U.S. Congress (as well as the Continental Congress), including delegates and resident commissioners. The "Bioguide" website at bioguide.congress.gov is a great resource for searching out historical information.

But there is little sign that Bioguide is Congress's repository of record, and it is little known by users, giving it lower authority marks than it should have. Some look to the House and Senate websites and beta.congress.gov for information about federal representatives, splitting authority among websites, rather than one established and agreed upon resource.

Bioguide scores highly on availability—we know of no problems with up-time or completeness (though it could use quicker updating when new members are elected). Bioguide is not structured for discoverability, though. Most people have not seen it, because search engines are not finding it.

Bioguide does a good thing in terms of machine readability, though. It assigns a unique ID to each of the people in its database. This is the first, basic step in making data useful for computers, and the Bioguide ID should probably be the standard for machine identification of elected officials wherever they are referred to in data. Unfortunately, the biographical content in Bioguide is not machine-readable.

Figure 1

October 2012

PUBLICATION PRACTICES FOR TRANSPARENT GOVERNMENT: RATING THE CONGRESS

How well can the Internet access data about Congress' work? The Cato Institute rated how well Congress publishes information in terms of authoritative sourcing, availability, machine-discoverability, and machine-readability.

SUBJECT	GRADE	COMMENTS
House and Senate Membership	House C− Senate A−	*The Senate has taken the lead on making data about who represents Americans in Washington machine-readable.*
Committees and Subcommittees	C−	*Organizing and centralizing committee information would create a lot of clarity with a minimum of effort.*
Meetings of House, Senate, and Committees	House B Senate B	*The House has improved its data about floor debates. The Senate is strong on commitee meetings.*
Meeting Records	D−	*There is lots of work to do before transcripts and other meeting records can be called transparent.*
Committee Reports	C+	*Committee reports can be found, but they're not machine-readable.*
Bills	B−	*Bills are the "pretty-good-news" story in legislative transparency, though there is room for improvement.*
Amendments	F	*Amendments are hard to track in any systematic way—and Congress has done little to make them trackable.*
Motions	F	*If the public is going to have insight into the decisions Congress makes, the motions on which Congress acts should be published as data.*
Decisions and Votes	B+	*Vote information is in good shape, but voice votes and unanimous consents should be published as data.*
Communications (Inter- and Intra-Branch)	F	*Transparent access to the messages sent among the House, Senate, and executive branch would complete the picture available to the public.*

Figure 2

PUBLICATION PRACTICES FOR TRANSPARENT GOVERNMENT: BUDGETING, APPROPRIATING, AND SPENDING

How well can the Internet access data about the federal government's budgeting, appropriating, and spending? The Cato Institute rated how well the government publishes information in terms of authoritative sourcing, availability, machine-discoverability, and machine-readability.

SUBJECT	GRADE	COMMENTS
Agencies	D-	*This grade is generous. There really should be a machine-readable federal government "organization chart."*
Bureaus	D-	*The sub-units of agencies have the same problem.*
Programs	D	*Program information is obscure, incomplete, and unorganized.*
Projects	F	*Some project information gets published, but the organization of it is bad.*
Budget Documents	Congress D / White House B-	*The president's budget submission and congressional budget resolutions are a mixed bag.*
Budget Authority	F	*Legal authority to spend is hidden and unstructured.*
Warrants, Apportionments, and Allocations	F	*Spending authority is divided up in an opaque way.*
Obligations	B-	*Commitments to spend taxpayer money are visible some places.*
Parties	F	*A proprietary identifier system makes it hard to know where the money is going.*
Outlays	C-	*We need real-time, granular spending data.*

As noted above, the other ways of learning about House and Senate membership are ad hoc. The Government Printing Office has a "Guide to House and Senate Members" at http://memberguide.gpo.gov/ that duplicates information found elsewhere. The House website presents a list of members along with district information, party affiliation, and so on, in HTML format (http://www.house.gov/representatives/), and beta.congress.gov does as well (http://beta.congress.gov/members/). Someone who wants a complete dataset must collect data from these sources using a computer program to scrape the data and through manual curation. The HTML presentations do not break out key information in ways useful for computers. The Senate membership page,[25] on the other hand, includes a link to an XML representation that is machine readable. That is the reason why the Senate scores so well compared to the House.

Much more information about our representatives flows to the public via representatives' individual websites. These are nonauthoritative websites that search engine spidering combines to use as a record of the Congress's membership. They are available and discoverable, again because of that prime house.gov and senate.gov real estate. But they only reveal data about the membership of Congress incidentally to communicating the press releases, photos, and announcements that representatives want to have online.

It is a narrow point, but there should be one and only one authoritative, well-published source of information about House and Senate membership from which all others flow. The variety of sources that exist combine to give Congress pretty good grades on publishing information about who represents Americans in Washington, but improving in this area is a simple matter of coordinated House and Senate efforts.

Committees and Subcommittees: C-

Like Americans' representation in Congress, lists of committees, their membership, and jurisdiction should be an easy lift. But it is not as easy as it should be to learn about the committees to which Congress delegates much of its work and the subcommittees to which the work gets further distributed.

The Senate has committee names and URLs prominently available on its main website.[26] The House does, too, at http://house.gov/committees/. But neither page offers machine-readable information about committees and committee assignments. The Senate has a nice list of committee assignments, again, though, not machine-readable. The House requires visitors to click through to each committee's web page to research what they do and who serves on them. For that, you'd go to individual committee websites, each one different from the others. There is an authoritative list of House committees with unique identifiers,[27] but it's published as a PDF, and it is not clear that it is used elsewhere for referring to committees.

Without a recognized place to go to get data about committees, this area suffers from lacking authority. To the extent there are data, availability is not a problem, but machine-discoverability suffers for having each committee publish distinctly, in formats like HTML, who their members are, who their leaders are, and what their jurisdiction is.

With the data scattered about this way, the Internet can't really see it. More prominence, including data such as subcommittees and jurisdiction, and use of a recognized set of standard identifiers would take this resource a long way.

Until committee data are centrally published using standard identifiers (for both committees and their members), machine-readability will be very low. The Internet makes sense of congressional committees as best it can, but a whole lot of organizing and centralizing—with a definitive, always-current, and machine-readable record of committees, their memberships, and their jurisdictions—would create a lot of clarity in this area with a minimum of effort.

There should be one and only one authoritative, well-published source of information about House and Senate membership.

> Can the public learn easily about what meetings are happening, where they are happening, when they are happening, and what they are about? It depends on which side of the Capitol you're on.

Meetings of House, Senate, and Committees—House: B/Senate: B

When the House, the Senate, committees, and subcommittees have their meetings, the business of the people is being done. Can the public learn easily about what meetings are happening, where they are happening, when they are happening, and what they are about? It depends on which side of the Capitol you're on.

The Senate is pretty good about publishing notices of committee meetings. From a webpage with meeting notices listed on it,[28] there is a link to an XML version of the data to automatically inform the public.

If a particular issue is under consideration in a Senate committee meeting, this is a way for the public to learn about it. This is authoritative, it is available, it is machine-discoverable, and has some machine-readable features. That means any application, website, researcher, or reporter can quickly use these data to generate more—and more useful—information about Congress.

The House does not have anything similar for committee meetings. To learn about those meetings, one has to scroll through page after page of committee announcements or calendars. Insiders subscribe to paid services. The House can catch up with the Senate in this area.

Where the House excels and the Senate lags is in notice about what will be considered on the floor. The House made great strides with the institution of docs.house.gov, which displays legislation heading for the floor. This allows any visitor, and various websites and services, to focus their attention on the nation's business for the week.

Credit is due the House for establishing this resource and using it to inform the public using authoritative, available, and machine-discoverable and -readable data. This is an area where the Senate has the catching up to do.

For different reasons, the House and Senate both garner Bs. Were they to copy the best of each other, they would both have As.

Meeting Records: D-

There is a lot of work to do before meeting records can be called transparent. The *Congressional Record* is the authoritative record of what transpires on the House and Senate floors, but nothing similar reveals the content of committee meetings. Those meeting records are produced after much delay—sometimes an incredibly long delay—by the committees themselves. These records are obscure, and they are not being published in ways that make things easy for computers to find and comprehend.

In addition, the *Congressional Record* doesn't have the machine-discoverable publication or machine-readable structure that it could and should. Giving unique, consistent IDs in the *Record* to members of Congress, to bills, and other regular subjects of this publication would go a long way to improving it. The same would improve transcripts of committee meetings.

Another form of meeting record exists: videos. These have yet to be standardized, organized, and published in a reliable and uniform way, but the HouseLive site (http://houselive.gov/) is a significant step in the right direction. It will be of greater use when it can integrate with other records of Congress. Real-time flagging of members and key subjects of debate in the video stream would be a great improvement in transparency. Setting video and video meta-data standards for use by both Houses of Congress, by committees, and by subcommittees would improve things dramatically.

House video is a bright spot in a very dark field, but both will shine brighter in time. When the surrounding information environment has improved to educate the public about goings-on in Congress in real time, the demand for and usefulness of video will increase.

Committee Reports: C+

Committee reports are important parts of the legislative process, documenting the findings and recommendations that com-

mittees report to the full House and Senate. They do see publication on the most authoritative resource for committee reports, the Library of Congress's THOMAS system. They are also published by the Government Printing Office.[29] The GPO's Federal Digital System (FDsys) is relatively new and is meant to improve systematic access to government documents, but it has not become recognized as an authoritative source for many of those documents.

Because of the sources through which they are published, committee reports are somewhat machine-discoverable, but without good semantic information embedded in them, committee reports are barely visible to the Internet.

Rather than publication in HTML and PDF, committee reports should be published fully marked up with the array of signals that reveal what bills, statutes, and agencies they deal with, as well as authorizations and appropriations, so that the Internet can discover and make use of these documents.

Bills: B-

Bills are a "pretty-good-news" story in legislative transparency. Most are promptly published. It would be better, of course, if they were all immediately published at the moment they were introduced, and if both the House and Senate published last-minute, omnibus bills before debating and voting on them.

A small gap in authority exists around bills. Some people look to the Library of Congress and the THOMAS site, and now beta.congress.gov, for bill information. Others look to the Government Printing Office. Which is the authority for bill content? This issue has not caused many problems so far. Once published, bill information remains available, which is good.

Publication of bills in HTML on the THOMAS site makes them reasonably machine-discoverable. Witness the fact that searching for a bill will often turn up the version at that source.

Where bills could improve some is in their machine-readability. Some information such as sponsorship and U.S. code references is present in the bills that are published in XML, and nearly all bills are now published in XML, which is great. Much more information should be published machine-readably in bills, though, such as references to agencies and programs, to states or localities, to authorizations and appropriations, and so on, referred to using standard identifiers.

With the work that the THOMAS system does to gather information in one place, bill data are good. This is relative to other, less-well-published data, though. There is yet room for improvement.

Amendments: F

Amendments are not the good-news story that bills are. They are "barely available," says Eric Mill of the Sunlight Foundation. "Given that amendments (especially in the Senate) can be as large and important as original legislation, this is an egregious oversight."

With a few exceptions, amendments are hard to track in any systematic way. When bills come to the House and Senate floors, amendment text is often available, but amendments are often plopped somewhere in the middle of the *Congressional Record* without any reliable, understood, machine-readable connection to the underlying legislation. It is very hard to see how amendments affect the bills they would change.

In committees, the story is quite a bit worse. Committee amendments are almost completely opaque. There is almost no publication of amendments at all—certainly not amendments that have been withdrawn or defeated. Some major revisions in process are due if committee amendments are going to see the light of day as they should.

Motions: F

When the House, the Senate, or a committee is going to take some kind of action, it does so on the basis of a motion. If the

Bills are a "pretty-good-news" story in legislative transparency.

public is going to have insight into the decisions Congress makes, it should have access to the motions on which Congress acts.

But motions are something of a black hole. Many of them can be found in the *Congressional Record*, but it takes a human who understands legislative procedure and who is willing to read the *Congressional Record* to find them. That is not modern transparency.

Motions can be articulated as data. There are distinct types of motions. Congress can publish which meeting a motion occurs in, when the motion occurs, what the proposition is, what the object of the motion is, and so on. Along with decisions, motions are key elements of the legislative process. They can and should be published as data.

Decisions and Votes: B+

When a motion is pending, a body such as the House, the Senate, or a committee will make a decision on it, only sometimes using votes. These decisions are crucial moments in the legislative process, which should be published as data. Like motions, many decisions are not yet published usefully. Decisions made without a vote in the House or Senate are published in text form as part of the *Congressional Record*, but they are not published as data, so they remain opaque to the Internet. Many, many decisions come in the form of voice votes, unanimous consents, and so on.

Voting puts members of Congress on record about where they stand. And happily, vote information is in pretty good shape. Each chamber publishes data about votes, meaning authority is well handled. Vote data are available and timely.

Both the House[30] and Senate[31] produce vote information. The latter also publishes roll call tables in XML, which is useful for computer-aided oversight. Overall, voting data are pretty well handled. But the omission of voice votes and unanimous consents drags the grade down and will drag it down further as the quality of data publication in other areas rises.

Voting puts members of Congress on record about where they stand. And happily, vote information is in pretty good shape.

Communications (Inter- and Intra-Branch): F

The Constitution requires each house of Congress to "keep a Journal of its Proceedings, and from time to time publish the same." The basic steps in the legislative process (discussed elsewhere) go into the journals of the House and Senate, along with communications among governmental bodies.

These messages, sent among the House, Senate, and Executive Branch, are essential parts of the legislative process, but they do not see publication. Putting these communications online—including unique identifiers, the sending and receiving body, any meeting that produced the communication, the text of the communication, and key subjects such as bills—would complete the picture that is available to the public.

Publication Practices for Transparent Government: Budgeting, Appropriations, and Spending

Agencies: D-

Federal agencies are the "agents" of Congress and the president. They carry out federal policy and spending decisions. Accordingly, one of the building blocks of data about spending is going to be a definitive list of the organizational units that do the spending.

Is there such a list? Yes. It's Appendix C of OMB Circular A-11, entitled: "Listing of OMB Agency/Bureau and Treasury Codes." This is a poorly organized PDF document that is found on the Office of Management and Budget website.[32]

Poorly organized PDFs are not good transparency. Believe it or not, there is still no federal government "organization chart" that is published in a way amenable to computer processing.

There are almost certainly sets of distinct identifiers for agencies that both the Treasury department and the Office of Management and Budget use. With modifications,

either of these could be published as the executive branch's definitive list of its agencies. But nobody has done that. Nobody seems yet to have thought of publishing data about the basic units of the executive branch online in a machine-discoverable and machine-readable format.

In our preliminary grading, we gave this category an "incomplete" rather than an F. That was "beyond generous," according to Becky Sweger of the National Priorities Project. We expect improvement in publication of this data, and the grades will be low until we get it.

Bureaus: D-

The sub-units of agencies are bureaus, and the situation with agencies applies to data about the offices where the work of agencies get divided up. Bureaus have identifiers. It's just that nobody publishes a list of bureaus, their parent agencies, and other key information for the Internet-connected public to use in coordinating its oversight.

Again, a prior "incomplete" in this area has converted to a D-, saved from being an F only by the fact that there is a list, however poorly organized and published, by the Office of Management and Budget.

Programs: D

It is damning with faint praise to call "programs" the brightest light on the organizational-data Christmas tree. The work of the government is parceled out for actual execution in programs. Like information about their parental units, the agencies and bureaus, data that identifies and distinguishes programs is not comprehensively published.

Some information about programs is available in usable form. The Catalog of Federal Domestic Assistance website (www.cfda.gov) has useful aggregation of some information on programs, but the canonical guide to government programs, along with the bureaus and agencies that run them, does not exist.

Programs will be a little bit heavier a lift than agencies and bureaus—the number of

programs exceeds the number of bureaus by something like an order of magnitude, much as the number of bureaus exceeds the number of agencies. And it might be that some programs have more than one agency/bureau parent. But today's powerful computers can keep track of these things—they can count pretty high. The government should figure out all the programs it has, keep that list up to date, and publish it for public consumption.

Thanks to the CFDA, data publication about the federal government's programs gets a D.

Projects: F

Projects are where the rubber hits the road. These are the organizational vehicles the government uses to enter into contracts and create other obligations that deliver on government services. Some project information gets published, but the publication is so bad that we give this area a low grade indeed.

Information about projects can be found. You can search for projects by name on USASpending.gov, and descriptions of projects appear in USASpending/FAADS downloads, ("FAADS" is the Federal Assistance Award Data System), but there is no canonical list of projects that we could find. There should be, and there should have been for a long time now.

The generosity and patience we showed in earlier grading with respect to agencies, budgets, and programs has run out. There's more than nothing here, but projects, so essential to have complete information about, gets an F.

Budget Documents—
Congress: D/White House: B-

The president's annual budget submission and the congressional budget resolutions are the planning documents that the president and Congress use to map the direction of government spending each year. These documents are published authoritatively, and they are consistently available, which is good. They are sometimes machine-

Believe it or not, there is still no federal government "organization chart" published in a way amenable to computer processing.

discoverable, but they are not terribly machine-readable.

The appendices to the president's budget are published in XML format, which vastly reduces the time it takes to work with the data in them. That's really good. But the congressional budget resolutions—when they exist—have no similar organization, and there is low correspondence between the budget resolutions that Congress puts out and the budget the president puts out. You would think that a person—or better yet, a computer—should be able to lay these documents side by side for comparison, but nobody can.

For its use of XML, the White House gets a B-. Congress gets a D.

Budget Authority: F

"Budget authority" is a term of art for what probably should be called "spending authority." It's the power to spend money, created when Congress and the president pass a law containing such authority.

Proposed budget authority is pretty darn opaque. The bills in Congress that contain budget authority are consistently published online—that's good—but they don't highlight budget authority in machine-readable ways. No computer can figure out how much budget authority is out there in pending legislation.

Existing budget authority is pretty well documented in the Treasury Department's FAST book (Federal Account Symbols and Titles). This handy resource lists Treasury accounts and the statutes and laws that provide their budget authority. The FAST book is not terrible, but the only form we've found it in is PDF. PDF is terrible. And nobody among our graders uses the FAST book.

Congress can do a lot better, by highlighting budget authority in bills in a machine-readable way. The administration can do much, much better than publishing the obscure FAST book in PDF.

Ideally, there would be a nice, neat connection from budget authority right down to every outlay of funds, and back up again from every outlay to its budget authority.

These connections, published online in useful ways, would allow public oversight to blossom. But the seeds have yet to be planted.

Warrants, Apportionments, and Allocations: F

After Congress and the president create budget authority, that authority gets divvied up to different agencies, bureaus, programs, and projects. How well documented are these processes? Not well.

An appropriation warrant is an assignment of funds by the Treasury to a treasury account to serve a particular budget authority. It's the indication that there is money in an account for an agency to obligate and then spend. "OMB has a web portal that agencies used to send apportionment requests," notes the National Priorities Project's Becky Sweger, "so the apportionment data are out there."

Where is this warrant data? We can't find it. Given Treasury's thoroughness, it probably exists, but it's just not out there for public consumption.

An apportionment is an instruction from the Office of Management and Budget to an agency about how much it may spend from a Treasury account in service of given budget authority in a given period of time.

We haven't seen any data about this, and we're not sure that there is any. There should be. And we should get to see it.

An allocation is a similar division of budget authority by an agency into programs or projects. We don't see any data on this either. And we should.

These essential elements of government spending should be published for all to see. They are not published, garnering the executive branch an F.

Obligations: B-

Obligations are the commitments to spend money into which government agencies enter. Things like contracts to buy pens, hiring of people to write with those pens, and much, much more.

USASpending.gov has quickly become the authoritative source for this information, but it is not the entire view of spending, and the data is "dirty": inconsistent and unreliable. The use of proprietary DUNS numbers—the Data Universal Numbering System of the firm Dun & Bradstreet—also weakens the availability of obligation data.

There is some good data about obligations, but it is not clean, complete, and well documented. The ideal is to have one source of obligation data that includes every agency, bureau, program, and project. With a decent amount of data out there, though, useful for experts, this category gets a B-.

Parties: F

When the government spends taxpayer dollars, to what parties is it sending the money?

Right now, reporting on parties is dominated by the DUNS number. It provides a unique identifier for each business entity and was developed by Dun & Bradstreet in the 1960s. It's very nice to have a distinct identifier for every entity doing business with the government, but it is not very nice to have the numbering system be a proprietary one.

"Parties" would grade well in terms of machine-readability, which is one of the most important measures of transparency, but because it scores so low on availability, its machine-readability is kind of moot. Until the government moves to an open identifier system for recipients of funds, it will get weak grades on publication of this essential data.

Outlays: C-

For a lot of folks, the big kahuna is knowing where the money goes: outlays. An outlay—literally, the laying out of funds—satisfies an obligation. It's the movement of money from the U.S. Treasury to the outside world.

Outlay numbers are fairly well reported after the fact and in the aggregate. All one has to do is look at the appendices to the president's budget to see how much money has been spent in the past.

But outlay data can be much, much more detailed and timely than that. Each outlay goes to a particular party. Each outlay is done on a particular project or program at the behest of a particular bureau and agency. And each outlay occurs because of a particular budget authority. Right now these details about outlays are nowhere to be found.

"Surely the act of cutting a check doesn't sever all relationship between that amount of money and its corresponding obligation/project/program," writes a frustrated Becky Sweger from the National Priorities Project. "Surely these relationships are intact somewhere and can be published."

Plenty of people inside the government who are familiar with the movement of taxpayer money will be inclined to say, "it's more complicated than that," and it is! But it's going to have to get quite a bit less complicated before these processes can be called transparent.

The time to de-complicate outlays is now. It's a feat of generosity to give this area a C-. That's simply because there is an authoritative source for aggregate past outlay data. As the grades in other areas come up, outlay data that stays the same could go down. Way down.

Conclusion

Many of the entities discussed here are low-hanging fruit if Congress and the administration want to advance transparency and their transparency grades. Authoritative, complete, and well-published lists of House and Senate membership, committees, and subcommittees are easy to produce and maintain, and much of the work has already been done.

The same is true of agencies and bureaus, at least on the executive branch side. Presidential leadership could produce an authoritative list of programs and projects within months. Establishing authoritative identi-

Outlay data can be much, much more detailed and timely.

fiers for these basic units of government is like creating a language, a simple but important language computers can use to assist Americans in their oversight of the federal government.

The more difficult tasks—amendments to legislation, for example, and discretely identified budget authorities—will take some work. But such work can produce massive strides forward in accountable, efficient, responsive, and—in the libertarian vision—smaller government.

Appendix A
Conceptual Data Model of Formal Legislative Processes in the U.S. Federal Government

Motivation

Goal is to have an
- open, authoritative set of machine-processable data covering formal actions of Congress and its members
- to *enable* (not create) a variety of uses for a variety of users:
  - for data processors (ontologies, codifications, correlations with other datasets)
  - for end users (apps, mashups, human-searchable websites, researchers, reporters)
  - for other government entities

Scope of the Specification

- a general statement of transparent data practices
- a conceptual model (descriptive and prescriptive) of desired data concerning the formal legislative process
- and not of specific publication or serialization technologies or methodologies

Transparent Data Practices

- availability
  - permanent
  - stable (always in same location)
  - complete
  - bulk accessible
  - incrementally accessible
  - open (publicly accessible and free of proprietary encumbrances)
- authority
  - authoritative (authoritative sources will emerge from consistent practices.)
  - timely/real-time
  - correctable (in response to consumers of data)
- machine-discoverability
  - internet-accessible
  - cross-referenceable
- machine-processability
  - comprehensive conceptual data model
  - semantically rich
  - well-defined, published serializations

Conceptual Data Model

Metamodel
Entities

An Entity represents an object in the world. An Entity is composed of unordered named Properties and is uniquely identified by an Identifier.

An Entity's Class defines what Properties and Identifiers compose a given Entity.

An Entity Class may be specified by other Entity Classes. Such Entity Classes are called Subclasses of the specified Entity Class. An Entity Subclass inherits the Properties and Identifiers of the Entity Class.

Properties

A Property consists of a Name and a Value. Names must be unique within an Entity. A Value must be an Entity, a Collection of Entities, or a typed literal.

A Property may be derived or computed, meaning that its value can be inferred from other Properties.

Identifiers

Identifiers uniquely identify an Entity. Identifiers are composed of the Values of one or more Properties which taken together are the minimum necessary to identify that Entity.

Identifiers should be natural where possible; if there is no natural Identifier for an Entity, a surrogate Identifier must be assigned and transmitted by an authority. Every Entity must have an Identifier.

Types

A Type describes a literal Value for a Property. Types may be simple (e.g., Integers, Strings, URIs, Currency Amounts, Dates, etc) or complex (XML documents, PDF documents, etc).

This specification does not define the textual representation for typed Values, but one should use representations that are standardized, machine-readable, and in conformance with the principles set forth in the Transparent Data Practices outlined in this document.

Collections

Collections are groups of Entities indicated together. Collections may be heterogeneous or homogeneous. Collections may have cardinality constraints.

Bag

A Bag is an unordered non-unique set of Entities. A single Entity may occur more than once within a Bag.

List

A List is an ordered non-unique set of Entities. A single Entity may occur more than once within a List. The sort order should be specified.

Set

A Set is an unordered unique set of Entities. An Entity may occur only once within a Set.

Ordered Set

An Ordered Set is an ordered unique set of Entities. The sort order should be specified.

Extending

This data model is not meant to be exhaustive. It may be extended by *augmentation* (adding additional properties to Entity Classes defined in this specification), or by *subclassing* (defining new Entity Classes inheriting from an existing Entity Class defined in this specification).

Abstract Entity Classes may not be augmented, only subclassed.

Any extensions must make use of a namespacing mechanism to prevent Property Name and Entity Class Name collisions with other extensions. No namespacing mechanism is defined by this specification—namespacing mechanisms are implementation-specific.

Metamodel Notation

The following notation is used to describe entities.

[«SuperClassName»] EntityClassName

    Description of Entity Class.

- «Identifier» (*PropertyName1, PropertyName2, . . .*) *this defines the property names that compose the Entity Class's identifier*
- *PropertyName: PropertyValueType*[*cardinality constraints*] {*collection information and other notes*}
- */DerivedPropertyName*: *PropertyValueType*

Model (Entity Classes)

Static Entities

Static Entities are those which change infrequently.

Body

An abstract Entity Class representing an official body of people.

«Body» ConstitutionalBody

An abstract Entity Class representing the House or Senate. This Entity Class is unusual in that most of its properties are derived. In principle, the value of these properties may be derived by examining all open Terms of all FederalElectiveOfficeholders.

- date: date {all other properties of this Body are assertions which are true on the date indicated by this property}
- /congress: number
- /session: number

«ConstitutionalBody» HouseOfRepresentatives

The membership of the House of Representatives on a given date

- /speaker: FederalElectiveOfficeholder
- /majorityLeader: FederalElectiveOfficeholder
- /minorityLeader: FederalElectiveOfficeholder
- /majorityWhip: FederalElectiveOfficeholder
- /majorityWhip: FederalElectiveOfficeholder
- /members: FederalElectiveOfficeholders[1..n] {Set, includes all members of the House of Representatives for a given Congress}

«ConstitutionalBody» Senate

The membership of the Senate on a given date

- /senatePresident: FederalElectiveOfficeholder {always the Vice President}
- /presidentProTempore: FederalElectiveOfficeholder
- /majorityLeader: FederalElectiveOfficeholder
- /assistantMajorityLeader: FederalElectiveOfficeholder
- /minorityLeader: FederalElectiveOfficeholder
- /AssistantMinorityLeader: FederalElectiveOfficeholder
- /members: FederalElectiveOfficeholders[1..n] {Set, includes all members of the Senate for a given Congress}

«Body» AbstractCommittee

Abstract Entity Class shared by Committees and Subcommittees

- «Identifier» code: string

- house: ConstitutionalBody_enum {house, senate, or joint}
- name: string
- jurisdiction: string {describes committee's purview}
- chairman: FederalElectiveOfficeholder[1,2] {Set, two chairmen reflects co-chairmanship}
- rankingMember: FederalElectiveOfficeholder[0,1] {leading member of the minority party, may be empty if committee has co-chairmanship}
- members: FederalElectiveOfficeholder[1..n] {Set, complete including chairman and rankingMember}

«AbstractCommittee» Committee

A Congressional Committee. Includes the Committee of the Whole.

- subcommittee: Subcommittee[0..n] {Set}

«AbstractCommittee» Subcommittee

A congressional subcommittee: *must* have only one parent Committee.

- «Identifier» code: string {full Identifier is (Committee code, SubCommittee code)}

Congress

A two-year meeting of the United States Congress composed of Sessions.

- «Identifier» number: integer
- start: date
- end: date
- sessions: Session[1..n] {OrderedSet}

Session

A meeting of a Congress. A Session *must* be part of one and only one Congress.

- «Identifier» number: integer {full identifier is (Congress number, Session number)}
- start: date
- end: date

Seat

Represents a Congressional Seat. This is an abstract class which exists solely to define Subclasses; there are no concrete Entities of this Class.

- state: usa_state

«Seat» HouseSeat

A Seat in the House of Representatives

- district: integer {0 for at-large}

«Seat» SenateSeat

A Seat in the Senate

- class: integer {senatorial class: 1, 2 or 3}

Term

Represents the time during which an official seat is held.

- start: date
- end: date
- office: House

«Term» CongressionalTerm

Represents a Congressional Term. A new Term beings when a person is sworn in.

- seat: Seat
- /congress: Congress {OrderedSet}

«Term» CongressionalOfficialTerm

Represents a Term of a Congressional office aside from Congressional Membership.

- office: congressionaloffice_enum

«Term» ExecutiveTerm

Represents a Term of an Executive office.
- office: executiveoffice_enum {president or vice president}

PartyAffiliation

Represents the time during which a person is a member of a party.
- start: date
- end: date
- party: party_enum

Person

Represents a Person. This is an abstract class which exists solely to define Subclasses; there are no concrete Entities of this Class.
- honorific: string {optional}
- firstName: string
- middleName: string {optional}
- lastName: string
- suffix: string {optional}

«Person» FederalElectiveOfficeholder

Represents a Person who holds an elective federal office. All federal elective officeholders should be identified by a single identifier system.
- terms: Term {Set}
- parties: PartyAffiliation {OrderedSet; ordered by start date}
- officialPortrait: image
- gender: gender_enum
- /currentTerm: Term
- /currentParty: {the current party affiliation of the officeholder}

«Person» Functionary

A person who is identified by title or purpose rather than by name. This is used for non-FederalElectiveOfficeholders who appear frequently in congressional proceedings but whose individual identities are not important, such as a clerk or a chaplain.
- title: string

Substantive Entities

Substantive Entities are those which contain information on the deliberations of Congress.

Bill

A Bill in Congress that has not become law.
- «Identifier» (congress, type, number)
- congress: Congress
- type: bill_type
- number: integer
- text: billtext {must include machine-extractable title and bill body text information}
- sponsor: FederalElectiveOfficeholder
- isByRequest: Boolean {indicates introduction without a show of support for the bill}
- cosponsor: FederalElectiveOfficeholder[0..n] {OrderedSet, date of cosponsorship must be recoverable through the actionlog}
- actionlog: Action {List, only actions that concern this bill}
- /state {state of the bill; inferred from actionlog and bill state machine}
- /introduced: datetime {date on which bill was introduced; inferred from actionlog}
- /introducedSession: {session during which bill was introduced; inferred from actionlog or date introduced}

PublicLaw

A Bill which has passed into law.

- «Identifier» (congress, lawnumber)
- congress: Congress
- lawnumber: integer
- bill: Bill {The bill passed to create this law.}
- dateEnacted: date

Amendment

An amendment to a Bill or to another Amendment.

- «Identifier» (object {Bill}, number) {the identifying object must be a Bill; it is found by following the object property of an amendment entity through its parent entities until a Bill is found.}
- venue: Body {where the amendment was offered}
- adoptionDate: datetime {if it was adopted, the time at which the amendment was adopted}
- number: integer {a monotonically increasing number unique among all amendments for a given Bill}
- object: Bill, Amendment {the thing amended; must eventually terminate at a Bill}
- changes: amendmentchange {the changes themselves; should be machine-processable can be applied to the object by machine}
- /afterChange: billtext, amendmentchange {optional; the text of the Bill or Amendment after applying the change}
- /introduced: datetime {the time at which the amendment was introduced, inferred from the Motion that introduces it}

Meeting

A specific temporally and spatially delineated gathering of a Body. Includes House of Representative, Senate, Committee, and Subcommittee meetings.

- legislativeDay: integer {A meeting's call to order and adjournment define the boundaries of a legislative day. Legislative days are numbered sequentially and numbering is reset at the beginning of a new Congress}
- start: datetime
- end: datetime
- location {physical location}
- title: string {optional: the official title of the meeting if one exists}
- purpose: text {optional}
- billSubject: Bill[0..n] {Set; bills discussed at a meeting}
- meetingBody: Body
- meetingType: meeting_type
- participants: Person[1..n] {Set}
- statements: LegislativeStatement {OrderedList, by time}
- records: Record {Set, transcripts of the meeting}
- materials: url {Set, reference to supplemental non-transcript documents used in the meeting}

LegislativeStatement

Something said to a Body convened in a Meeting.

- «Identifier» (meeting, time)
- time: datetime
- /meeting: Meeting {inferred}
- speaker: Person

- bill: Bill {optional, bill which is mentioned or indicated by the speaker}
- text: transcript {optional}
- officialText: transcript {optional}
- video: url {optional, reference to video files}
- /records: Record {Set, Records which include this statement, inferred from the meeting's records property}

Record

A record of the entire content of a meeting released as a single document. Although not composed of LegislativeStatements, these should be derivable from the text, video, or audio. Must include at least one of the optional properties.
- source: text {who prepared the record}
- released: date
- text: Transcript {optional}
- video: video {optional}
- audio: audio {optional}

Report

A report submitted by a committee to a house of Congress
- number: integer {identifier assigned when report is filed}
- committee: AbstractCommittee {Committee or SubCommittee}
- text: string {the text of the report in a structured markup language}
- /contains: Vote, Bill, Amendment, Decision {Set; entities present in the report itself; should be inferred from the text}
- /about: Bill

Administrative Entities

Administrative Entities are those that affect the state of a Bill.

Motion

A formal proposition put before a Body which requires the consent of that Body to be approved. The thing approved depends on the nature of the Motion, but includes Amendments, passage of Bills, adjournments, etc. Motions are closely tied to Decisions.
- «Identifier» (meeting, time)
- time: datetime
- meeting: Meeting {the meeting in which the Motion was made}
- /before: Body {the Body to whom the proposition is addressed and from whom it requires a Decision; inferred from meeting's Body}
- motionType: motion_type {optional; where the proposition is of a standard type outlined in the rule it is indicated here; otherwise the proposition text itself must suffice}
- proposition: string {the natural-language text of the motion}
- object: Bill, Amendment, Meeting {optional; where the proposition is about some object it is indicated here. The object should be evident from the proposition.}
- decisions: {OrderedSet; A motion may have several Decisions because members may object to a Decision. The last and only the last Decision in this set must be the deciding one and have an isDeciding property set to true}
- /isAdopted: Boolean {inferred from decisions property}

«Motion» Referral

The assignment of a bill to a committee for consideration. This is a Motion with a motionType of "to refer" and a Bill as its object.
- terms: referral_term {whether the bill is to be considered by all committees at once or one at a time}

- referredTo: Committee {List}

Decision

The expression of assent or dissent by a Body for or against a Motion.
- «Identifier» (motion, time)
- time: datetime
- motion: Motion {the motion being decided}
- /proposition: string {inferred from Motion}
- /object {inferred from Motion}
- objectionGrounds: text {optional; if there is an objection to the outcome of this Decision, the grounds for the objection is noted here}
- objector: FederalElectiveOfficeholder {optional; present if there is an objection}
- type: decision_type {the means of measuring assent by the Body as a whole, e.g. by roll call}
- rule: decision_rule {the type of assent required by a Body's members, e.g. simple majority, lack of objection}
- result: decision_result {the final outcome of the decision}
- isDeciding: boolean {whether this Decision was the final and deciding one for the referenced Motion; if true, it must be the last Decision for a given Motion and it must have no value for the objector and objection properties}

«Decision» RollCall

A Decision resolved by voting.
- «Identifier» (Congress, Session, number) {congress and session are inferred from the motion}
- number: integer {the number assigned to this roll call}
- votes: Vote[1..n] {Set}

Vote

An individual vote
- voter: FederalElectiveOfficeHolder
- vote: vote_cast

Communication

A formal message or communication between houses of Congress or the president and Congress
- «Identifier» (Congress, House, number) {Congress and House are inferred from the Meeting indicated by the introducedAt property}
- number: int {a monotonically increasing number uniquely identifying the communication; resets at the beginning of each Congress}
- from: Body
- to: Body
- introducedAt: Meeting
- text: communication {content of the communication with machine-processable markup}
- summary: communication {summarized content of the communication as shown in the House or Senate Journal}
- /about: Bill {Set, optional; derived from text property. If the communication references one or more bills, these should be accessible through this property}

Actions/Events
Relationship to Entity Classes

Actions are an event-based, incremental view of congressional activities. Every Action

should contain enough information to fully specify either a new entity or a set of modifications to an existing entity, or both.

The entity an action modifies is called the *object* of that action.

An ordered list of actions with an identical object is called a *actionlog*. Actionlogs should be available for entities retrieved through bulk access. For example, a Bill Entity should have some way to list all actions that affected it.

Action Entity Class
Action
- «Identifier» (meeting, timestamp)
- timestamp: datetime
- meeting: Meeting
- type: action_type
- object: Entity {Set}

Action Types
Below are the defined action types and the Entity they create or modify
- CallToOrder: Meeting, Session, Congress
- Adjourn: sets end date on Meeting, Session, or Congress
- SwearIn: Term {refers to a Person indirectly}
- Establish: Committee, Subcommittee
- Introduce: Bill
- Refer: Referral, Bill
- Report: Bill
- Cosponsor: Bill
- Remove-Cosponsor: Bill
- Amend: Bill, Amendment
- Say: Statement, Transcript
- Decide: Decision {refers to Bill, PublicLaw, or Amendment indirectly}
- Present: Report, Communication
- Pass: Bill
- Veto: Bill

State Machines
In principle the set of allowed action types and entity modifications at any point in a sequence of actions is constrained by the state of those entities. Some actions advance the state of entities in such a way that other actions upon those entities are no longer possible and new actions are possible. (For example, a Meeting that has been called to order may not be called to order again.)

The rules that govern the transitions between states are called *state machines*. Because of the complexity of the formal legislative process and because the details of this process may change over time, this specification does not rigorously define a set of state machines governing entity states.

Bill States
However, the value of the Bill *state* property is governed by a state machine because the state of a Bill is important to know and difficult to discover algorithmically.

Below is a description of the defined bill state values and their types.
introduced

Last action was a successful motion to **Introduce**.

referred

Last action was a successful motion to **Refer**.

reported

Last action was a **Report** by committee.

pass.house

Last action was a **Pass** by the House for a bill originating in the House which requires both chambers to be enacted. The bill must go to the Senate.

pass.senate

Last action was a **Pass** by the Senate for a bill originating in the Senate that requires both chambers to be enacted. The bill must go to the House.

pass_back.house

Last action was a **Pass** by the House for a bill originating in the Senate which requires both chambers to be enacted, but the bill contains modifications to which the Senate must agree. Modifications are noted by successful **Amend** actions since the **Pass** action in the House.

pass_back.senate

Last action was a **Pass** by the Senate for a bill originating in the House that requires both chambers to be enacted, but the bill contains modifications to which the House must agree. Modifications are noted by successful **Amend** actions since the **Pass** action in the Senate.

passed

Last action was a **Pass** by any chamber which was sufficient for the bill to achieve final passage.

- For simple resolutions, the bill passed in the originating chamber. This is the final state for a simple resolution.
- For concurrent resolutions, the bill passed identically in both chambers. This is the final state for concurrent resolutions.
- For constitutional amendments, the bill passed identically in both chambers, but must still be ratified by the states.
- For all other bill types, the bill passed identically in both chambers and must be presented to the President to be signed or vetoed.

vetoed

The last action was that the President vetoed a passed bill. The veto may still be overridden.

veto_override.house

The last action was that the House overrode a presidential veto, but the Senate has not.

veto_override.senate

The last action was that the Senate overrode a presidential veto, but the House has not.

enacted

The bill has become a public law or constitutional amendment either by presidential signature, veto override by both houses, or state ratification. This is a final state.

Types (property-level specifications)

The exact representation of the types below will depend on the concrete data model that implements this abstract model. Use existing standards where possible and aim for unambiguous machine-readability.

markuptype

An abstract type. A markuptype is a document with inline machine-processable markup (e.g. XML) from which it is easy to extract contained or related Entities and other

semantic information.

Special considerations:

- References to U.S. Public Laws or Codes and Statues should include an explicit machine-readable reference to the prior law affected.
- Agencies and Programs should be referenced by a standard numbering scheme, such as MAX codes or contractor codes, which are unambiguous over time.
- Where a person is mentioned, this fact should be indicated inline with a machine-readable proper name even if a unique identifier of the person is not available.
- Where a location is mentioned, this fact should be indicated inline with a machine-readable name even if a unique identifier of the location is not available.
- Where a government or agency is mentioned, this fact should be indicated inline with a machine-readable name even if a unique identifier of the government is not available. At the very least, federal agencies and U.S. state governments should have unique identifiers.

billtext (markuptype)

The text of a bill. Titles, agencies, or programs affected, U.S. Code sections affected, authorizations or appropriations of funds and their amounts, locations, people, foreign and state governments and agencies, internal section numbers, the version of the text itself, the status, and the date should be easily extractable from a billtext document.

Special considerations:

- Where Authorizations and Appropriations appear in the text of a bill, the following should be noted inline: the fact that a passage's language is authorizing or appropriating, what the mandate or agency or program is that is being authorized or appropriated, and any applicable dollar amounts attached to the authorization or appropriation

amendmentchange (markuptype)

The text of an amendment. Should be an unambiguous machine-processable description of the changes that have been proposed for another markuptype (billtext or amendmentchange) to produce the desired new document.

committeereport (markuptype)

The text of a committee report. If any other entities are expressed in the markup (e.g. Decisions, Statements, Bills), these should be expressed in a way rich enough to allow one to derive the corresponding Entity.

Where applicable, a committeereport should include: statements, minority statements, statement of authorizing clause of the Constitution, any special statements required by House or Senate rules (indicating rule number), bill texts, section-by-section summaries of bills, proposed amendments, and committee voting records.

transcript (markuptype)

The record of human speech or action. One should be able to identify every Person Entity speaking and the time of each utterance so that one may derive LegislativeStatement Entities. If any other entities are expressed in the markup (e.g. Reports, Decisions, Statements), these should be expressed in a way rich enough to allow one to derive the corresponding Entity. This markup type should be appropriate for any temporally arranged, mixed-content documents, such as the *Congressional Record*.

Special considerations:

- Support documents, such as presentation files or testimony documents, should be referenced.
- Gaps in the transcript that are caused by breaks or adjournments should be indicated.
- The entrance or exit of People into the Meeting should be indicated.

communication (markuptype)

    The text of a communication to a House of Congress from the other House of Congress or from the executive. One should be able to identify mentioned people, bills, public laws, U.S. Code sections, dockets, and regulations (RIN).

constitutionalbody_enum

- House of Representatives
- Senate
- Joint (used by some committees)
- Conference (used by conference committees)
- Presidency (used by a President)
- Vice Presidency (used by the Vice President)

congressionaloffice_enum

- House of Representatives
  - speaker
  - majorityLeader
  - minorityLeader
  - majorityWhip
  - majorityWhip
- Senate
  - senatePresident {always the Vice President}
  - presidentProTempore
  - majorityLeader
  - assistantMajorityLeader
  - minorityLeader
  - assistantMinorityLeader

executiveoffice_enum

- president
- vicePresident

party_enum

- democrat
- independent
- republican

meeting_type

- Meeting
  - Hearing
  - Markup
- Debate

bill_type

- h (House of Representatives Bill)
- hr (House Simple Resolution)
- hj (House Joint Resolution)
- hc (House Concurrent Resolution)
- s (Senate Bill)
- sr (Senate Simple Resolution)
- sj (Senate Joint Resolution)
- sc (Senate Concurrent Resolution)

vote_cast

- yes
- no

- conflict of interest {Senate only, Rule XII}
- not voting
- present {for quorum calls}

motion_types
- to adjourn
- for the previous question
- to postpone to a day certain
- to amend
- to postpone indefinitely
- to establish quorum
- to proceed with debate {cloture}
- to pass
- to reconsider
- to rise
- to refer {includes "to commit" and "to recommit"}
- to concur {committee motion}
- to disagree {committee motion}
- to close debate {committee motion}
- to subpoena {hearing motion}
- to close the meeting {hearing motion}
- to establish an investigative subcommittee {hearing motion}
- to call witnesses selected by the minority {hearing motion, Rule XI 2.(j)(1)}

decision_type
- voice vote
- roll call
- unanimous consent

decision_rule
- simple majority
- 2/3rd majority
- 3/5ths majority
- lack of objection

decision_result
- Passed
- Agreed
- Confirmed
- Failed
- Defeated
- Rejected
- Not Sustained
- Passed contrary to rule

referral_term
- single
- joint
- sequential

Special Considerations

Bill Versions

A *bill version* is the official text of a bill at a given time. The passing of an amendment to a

bill creates a new version of a bill.

There are two ways to identify a bill version given the text of the bill as introduced and a full set of adopted amendments to the bill and of adopted amendments to adopted amendments of the bill.

The first way is by date and time. Given a date and time, one must be able to apply all adopted amendments to a bill text or to adopted amendment texts in order of their *adoptionDate* up to the desired date and time. The result of these transformations should be the official text of the bill.

The second way is by amendment number. A bill's version may be identified by the amendment number of an amendment that has that bill as its object. Given the amendment identified by this amendment number, one must be able to apply all adopted amendments to a bill text or to adopted amendment texts in order of their *adoptionDate* up to and including the amendment indicated. The result of these transformations should be the official text of the bill at the moment the indicated amendment was adopted.

All amendments must express their changes as against the official version of the bill at the time of amendment adoption. Amendments must be applied sequentially in order of adoption, and may not be applied concurrently to the same version of a bill.

Only adopted amendments are considered for bill versioning. This specification does not define identifiable bill versions for amendments that have not been adopted. Thus this specification defines no mechanism for identifying proposed or unrealized versions of bills.

Appendix B
Conceptual Data Model of the
U.S. Federal Government Budgetary Process

Date
    2011-12-12
Revision
    10
Status
    draft for presentation 2011-12-14

Motivation

Goal is to have an:
- open, authoritative set of machine processable data covering federal:
  - Budgeting
  - Appropriations
  - Spending
- to *enable* (not create) a variety of uses for a variety of users:
  - for data processors (ontologies, codifications, correlations with other datasets)
  - for end users (apps, mashups, human-searchable websites, researchers, reporters)
  - for other government entities

Scope of the Specification

- a general statement of transparent data practices
- a conceptual model (descriptive and prescriptive) of desired data concerning budgets, appropriations, allocations, obligations, outlays, and reporting
- not of specific publication or serialization technologies or methodologies

Note that this specification makes reference to Entities and concepts explained in Appendix A.

Depth of the Specification

This data is sought primarily to bring transparency to the federal spending lifecycle from the original budget proposed by the president through the creation of budget authority by Congress, to the final outlay of money by executive branch agencies.

The primary problem that conforming data to this specification will solve is the current inability to correlate particular obligations and outlays to specific appropriations and budget authorities.

This specification does not attempt to model financial data detailed down to the last transaction or with the rigor required by accountants; nor does it model receipts or other sources of federal revenue; nor does it model borrowing authority; nor does it attempt to model all budget- and spending-related communication among agencies, the Treasury Department, and the Office of Management and Budget.

Transparent Data Practices

- availability
  - permanent
  - stable (always in same location)
  - complete
  - bulk accessible
  - incrementally accessible
  - open (publicly accessible and free of proprietary encumbrances)
- authority
  - authoritative (Authoritative sources will emerge from consistent practices.)
  - timely/real-time
  - correctable (in response to consumers of data)
- machine discoverability
  - internet-accessible
  - cross-referenceable
- machine processability
  - comprehensive conceptual data model
  - semantically rich
  - well-defined, published serializations

Background

This section is informative, not normative.

In order to understand the data of interest, it is necessary to understand something of the budgetary process.

1. Executive:
    1. President proposes a budget.
2. Legislative:
    1. Congress creates its own budget through House and Senate resolutions. (See Appendix A for these resolutions.)
    2. Congress creates Budget Authorities in law, which direct that money should be spent for a particular purpose.
3. Executive:
    1. Treasury Warrant Issuance: The Financial Management Service (FMS) of the Treasury Department and the Office of Management and Budget (OMB) come to agreement on which Treasury Accounts fund which Budget Authorities and issue Treasury Warrants to Agencies for all Bureau-level accounts under their purview. This creates a balance in the Treasury Account which can fund Bureau spending.
    2. Apportionment: The OMB informs individual Agencies how much of their Warrant they may spend per financial quarter. Agencies then apportion this amount to their Bureau-level accounts.
    3. Allocation: Bureaus allocate resources to Programs, which serve a given Budget Authority.
    4. Obligation: Agencies and Bureaus, through Programs, enter into various kinds of legally binding agreements (obligations), which will be satisfied by outlays.
    5. Outlay: Money transfers are executed to liquidate obligations.

6. Reporting: The OMB reports expenditures by Budget Authority and makes recommendations to Congress for future Budget Authorities.

Conceptual Data Model

Metamodel
Entities

An Entity represents an object in the world. An Entity is composed of unordered named Properties and is uniquely identified by an Identifier.

An Entity's Class defines what Properties and Identifiers compose a given Entity.

An Entity Class may be specified by other Entity Classes. Such Entity Classes are called Subclasses of the specified Entity Class. An Entity Subclass inherits the Properties and Identifiers of the Entity Class.

Properties

A Property consists of a Name and a Value. Names must be unique within an Entity. A Value must be an Entity, a Collection of Entities, or a typed literal.

A Property may be derived or computed, meaning that its value can be inferred from other Properties.

Identifiers

Identifiers uniquely identify an Entity. Identifiers are composed of the Values of one or more Properties which taken together are the minimum necessary to identify that Entity.

Identifiers should be natural where possible; if there is no natural Identifier for an Entity, a surrogate Identifier must be assigned and transmitted by an authority. Every Entity must have an Identifier.

Types

A Type describes a literal Value for a Property. Types may be simple (e.g., Integers, Strings, URIs, Currency Amounts, Dates, etc) or complex (XML documents, PDF documents, etc).

This specification does not define the textual representation for typed Values, but one should use representations that are standardized, machine-readable, and in conformance with the principles set forth in the Transparent Data Practices outlined in this document.

Collections

Collections are groups of Entities indicated together. Collections may be heterogeneous or homogeneous. Collections may have cardinality constraints.

Bag

A Bag is an unordered non-unique set of Entities. A single Entity may occur more than once within a Bag.

List

A List is an ordered non-unique set of Entities. A single Entity may occur more than once within a List. The sort order should be specified.

Set

A Set is an unordered unique set of Entities. An Entity may occur only once within a Set.

Ordered Set
> An Ordered Set is an ordered unique set of Entities. The sort order should be specified.

Extending
> This data model is not meant to be exhaustive. It may be extended by *augmentation* (adding additional properties to Entity Classes defined in this specification), or by *subclassing* (defining new Entity Classes inheriting from an existing Entity Class defined in this specification).
>
> Abstract Entity Classes may not be augmented, only subclassed.
>
> Any extensions must make use of a namespacing mechanism to prevent Property Name and Entity Class Name collisions with other extensions. No namespacing mechanism is defined by this specification—namespacing mechanisms are implementation-specific.

Metamodel Notation
> The following notation is used to describe entities.

[«SuperClassName»] EntityClassName
> *Description of Entity Class.*
> - «Identifier» (*PropertyName1, PropertyName2, …*) *this defines the property names that compose the Entity Class' identifier*
> - *PropertyName*: *PropertyValueType[cardinality constraints] {collection information and other notes}*
> - */DerivedPropertyName*: *PropertyValueType*

Model (Entity Classes)

ExecutiveAgent
> Abstract Entity representing an agent responsible for executing an Authority.
> - «Identifier» (ombAgencyCode)
> - ombAgencyCode: ombagencycode {three- or five-digit form}
> - treasuryAgencyCode: treasuryagencycode
> - name: string
> - website: url

«ExecutiveAgent» Agency
> A federal agency.
> - ombAgencyCode: ombagencycode {three-digit form}

«ExecutiveAgent» Bureau
> A bureau of a federal agency.
> - parent: Agency
> - ombAgencyCode: ombagencycode {five-digit form}

Program
> A federal program overseen by a Bureau.
> - «Identifier» (parent, programNumber)
> - programNumber: int {disambiguates a program within a Bureau}
> - name: string
> - description: string
> - parent: Bureau
> - website: url {option}

Project
> A federal project overseen by a Program.

- «Identifier» (parent, projectNumber)
- projectNumber: int {disambiguates a project within a Program}
- name: string
- parent: Program

Authority

Abstract class specifying a proposed, existing, or terminated authority to obligate funds. When a new Authority replaces an older Authority, the older authorities are referenced through the previousAuthority property.

- «Identifier» (fiscalYear, number)
- fiscalYear: year {fiscal year in which authority did, does, or would first gain force}
- number: int {number unique within fiscal year; used to disambiguate Authorities. In almost all cases this should match a treasury account symbol, which is the accountSymbol property of the AppropriationWarrant Entity}
- source: string {a reference to the legal document which proposes or establishes the authority, such as a Bill, Public Law, U.S. Code or Statute}
- sourceClass: authoritysourceclass_enum {type of document in which the budget authority was created}
- /status: authoritystatus_enum {whether the Authority is proposed, in effect, or terminated}
- authorizedAmount: dollarorindefinite {Optional; where a distinction can be made between an authorized and appropriated amount, this property may contain the authorized amount. Note that a value of "indefinite" is not the same as a missing value}
- appropriatedAmount: dollarorindefinite {Optional; where a distinction can be made between an authorized and appropriated amount, this property may contain the appropriated amount. Note that a value of "indefinite" is not the same as a missing value}
- appropriationSource: Bill {the bill that appropriates or proposes to appropriate funds for this authority}
- authorizingSource: Bill {the bill that authorizes or proposes to authorize this authority}
- authorityClass: authorityclass_enum {type of authority that is or would be exercised}
- effectiveDate: date {when the authority begins}
- terminationDate: dateorindefinite {when the authority ends}
- purpose: string {description of the Authority}
- agent: ExecutiveAgent[1..n] {Set, Optional, who or what will execute the purpose of the authority using the amount}
- estimatedOutlay: OutlayEstimate[0..n] {List, ordered by fiscalYear; estimated spending in future fiscal years; this is informational only and is not part of the Authority}
- creationDate: date {the date on which the Authority record was created; this is unrelated to effectiveDate}
- changeDate: date {the date on which the Authority record changed; usually this will change after a previousAuthority is determined}
- previousAuthority: Authority[0..n] {past Authorities which this authority replaces; Note that until the OMB and FMS officially determine whether there is a continuity among Authorities no continuity is implied. Thus, this field will often be empty if no AppropriationWarrants have been issued for the authority.}

PresidentialBudget

Represents a presidential budget Appendix. This Entity is a container for budget reports.

- «Identifier» (fiscalYear)
- fiscalYear:
- publicationDate: date
- report: PresidentialBudgetReport[1..n]

PresidentialBudgetReport

Represents an individual report in a presidential budget Appendix. Reports are organized by Authority.

- «Identifier» (authority)
- authority: Authority
- text: presidentialbudgetreporttext
- /budget: PresidentialBudget

OutlayEstimate

Represents an estimate of spending in a fiscal year for a Budget Authority. The amounts specified are not actually part of the Budget Authority and are purely informational data found in budget documents.

- fiscalYear: year
- amount: int
- /authority: Authority

AppropriationWarrant

An amount of funds assigned by the Treasury into a treasury account to serve a particular Authority which was the result of appropriating legislation.

- «Identifier» (warrantId)
- warrantId: string {a unique identifier for a warrant}
- authority: Authority {the budget authority under which the AppropriationWarrant was issued}
- treasuryAccount: treasuryaccountid
- accountHolder: ExecutiveAgent {the agency which may use the funds in the treasury account indicated by the warrant. This is also indicated first two digits of the treasuryaccountid}
- accountSymbol: int {treasuryaccountid less the first two digits. This account symbol will be a component of all budget reporting identifiers.}
- amount: int

Apportionment

An instruction from the OMB to an agency about how much an ExecutiveAgent may spend from a treasury account in service of an Authority within a given period of time.

- «Identifier» (id)
- id: int {globally and temporally unique identifier for Apportionments}
- issuanceDate: date
- authority: Authority
- agent: ExecutiveAgent
- amount: int
- periodStartDate: date
- periodEndDate: date

Allocation

A division of an Apportionment by an ExecutiveAgent into Programs or Projects.

- «Identifier» (id)
- id: int {globally and temporally unique identifier for Allocations}
- issuanceDate: date
- authority: Authority

- apportionment: Apportionment
- /treasuryAccount: treasuryaccountid {derived from the apportionment}
- subAgent: Program, Project, Obligation {the subAgent that receives the portion of the Apportionment}
- allocator: ExecutiveAgent
- amount: int
- periodStartDate: date
- periodEndDate: date
- agent: ExecutiveAgent

Obligation

An abstract class representing a binding agreement or statutory requirement that will result in outlays immediately or in the future.
- «Identifier» (obligationId)
- obligationId: string {uniquely identifies the obligation}
- authority: Authority
- allocation: Allocation {optional}
- apportionment: Apportionment {optional}
- summary: string {description of the obligation}
- amount: int
- obligor: ExecutiveAgent

«Obligation» Award

An abstract class representing an agreement between the federal government and a non-federal Entity for deliverables rendered to the federal government. Refer to [USASpending.gov](http://www.usaspending.gov/learn?tab=About%20the%20Data) for more information on these fields. Awards are typically referred to as "discretionary spending" because the authority to obligate funds is an appropriation by congress.
- contractId: string {uniquely identifies the document that creates the obligation}
- obligee: Party
- creationDate: date {date the obligation was created, such as the date a contract was signed}
- startingDate: date {date the obligation took effect}
- endingDate: date {date the obligation ceased}

«Award» Procurement Contract

An Award that acquires a good or service for the exclusive use or benefit of the Federal Government.
- federalAwardId: string {agency-defined unique award identifier}
- piid: string {Procurement Instrument ID}

«Award» Grant

An Award that produces a good or service that is not for the exclusive use of the Federal Government. If the use of the good or service is partially shared by the Federal Government, it may be called a Cooperative Agreement; otherwise it is a Grant.
- federalGrantId: string {unique grant identifier}

«Obligation» Salary

An Entity representing an aggregate of salary and wage obligations to Federal Employees.
- numberOfEmployees: int

«Obligation» DirectPayment

An Entity representing some aggregate of direct payments paid to individuals. This includes entitlement obligations (Retirement, Disability, Medicare, Unemployment, etc.) DirectPayments are typically referred to as "mandatory spending" because the obligation

is required by statue.
  - numberOfPayees: int

«Obligation» Other

An Entity representing other unclassified Obligations, such as insurance and loans.

Outlay

Spending in execution of an Obligation.
  - «Identifier» (transactionId)
  - transactionId: treasurytransactionid {uniquely identifies a financial transaction in the Treasury Transaction Reporting System}
  - authority: Authority
  - amount: int {in rare cases this may be a negative number, in which case it indicates a credit to the treasury account}
  - treasuryAccountId: treasuryaccountid
  - treasurySubAccountId: int {a subaccount under a treasuryaccountid; use of subaccounts is entirely at the discretion of the ExecutiveAgent}
  - accountHolder: ExecutiveAgent
  - transactionDate: date
  - settlementDate: date
  - payee: Party or Person {the other party to the transaction; the Person Entity is defined in the Legislative [*Conceptual Data Model of Formal Legislative Processes*](http://wac.0873.edgecastcdn.net/800873/blog/wp-content/uploads/2b-Conceptual-Data-Model-of-US-Formal-Legislative-Processes.html); for privacy reasons, this field may be omitted when payees are individuals.}
  - obligation: Obligation {the obligation that this outlay services}

Party

The recipient of an Outlay or party to an Obligation. This may include federal entities, state entities, federal employers, contractors, grant recipients, or foreign countries, which should be expressed by subclasses of this Entity as necessary.
  - «Identifier» (partyId)
  - partyId: int {identifies the party to an obligation or transaction}

Types (property-level specifications)

The exact representation of the types below will depend on the concrete data model that implements this abstract model. Use existing standards where possible and aim for unambiguous machine-readability.

authoritystatus_enum
  - *proposed*: the Authority has not come into effect either because its effectiveDate has not yet been reached or because the Authority never passes into Public Law.
  - *in effect*: the Authority is legally in effect and spending may in principle occur citing this Entity as its Authority.
  - *terminated*: the Authority has elapsed either because its terminationDate has been reached or because some condition terminating it has been met.

dollarorindefinite

A specific dollar amount, or "indefinite" if no dollar amount is specified.

dateorindefinite

A specific calendar date, or "indefinite" if no date is specified.

ombagencycode

A three- or five-digit number that indicates a specific Agency or Bureau. The first three digits indicate an Agency; the last two indicate a Bureau of that Agency. These include

Legislative and Judicial branch "Agencies" and "Bureaus." There is a mapping between ombagencycode and treasuryagencycode.

treasuryagencycode

A two-digit number that indicates a specific Agency or Bureau. These include Legislative and Judicial branch "Agencies" and "Bureaus." There is a mapping between ombagencycode and treasuryagencycode.

treasuryaccountid

Uniquely identifies a Treasury account.

authoritysourceclass_enum

A one- or two-character code indicating the type of document that created an Authority.
- PL: PublicLaw
- PB: President's Budget
- B: Bill, including House and Senate Budgets

markuptype

An abstract type. A markuptype is a document with inline machine-processable markup (e.g. XML) from which it is easy to extract contained or related Entities and other semantic information.

presidentialbudgetreporttext (markuptype)

The text of a presidential budget report. Authority, agency, bureau, schedules, treasury accounts, functional classification, object classifications, and receipts should be easily extractable from a presidentialbudgetreporttext document.

September 2011

Publication Practices for Transparent Government: Rating the Congress

How well can the Internet access data about Congress' work? The Cato Institute rated how well Congress publishes information in terms of authoritative sourcing, availability, machine-discoverability, and machine-readability.

Subject	Grade	Comments
House and Senate Membership	B+	*The variety of sources that exist combine pretty well to reveal who represents Americans in Washington, D.C.*
Committees and Subcommittees	C	*Organizing and centralizing committee information would create a lot of clarity with a minimum of effort.*
Meetings of House, Senate, and Committees	Senate B+ House D+	*Can the public learn about what meetings are happening, when, and on what subject? Depends on which side of the Capitol you're on.*
Meeting Records	C-	*There is lots of work to do before transcripts and other meeting records can be called transparent.*
Committee Reports	D+	*Committee reports are barely visible to the Internet.*
Bills	A-	*Bills are the "pretty-good-news" story in legislative transparency, though there is room for improvement.*
Amendments	House/Senate C Committees I	*Amendments are hard to track in any systematic way—especially in committee.*
Motions	I	*If the public is going to have insight into the decisions Congress makes, the motions on which Congress acts should be published as data.*
Decisions	I	*The decisions Congress makes should be published as data.*
Votes	A-	*Vote information is in pretty good shape.*
Communications (Inter- and Intra-Branch)	I	*Transparent access to the messages sent among the House, Senate, and executive branch would complete the picture available to the public.*

Appendix D

PUBLICATION PRACTICES FOR TRANSPARENT GOVERNMENT: BUDGETING, APPROPRIATING, AND SPENDING

How well can the Internet access data about the federal government's budgeting, appropriating, and spending? The Cato Institute rated how well the government publishes information in terms of authoritative sourcing, availability, machine-discoverability, and machine-readability.

SUBJECT	GRADE	COMMENTS
Agencies	I	*Believe it or not, there is NO federal government "organization chart" that is published in a way amenable to computer processing!*
Bureaus	I	*The sub-units of agencies have the same problem.*
Programs	I	*The data that identifies and distinguishes government programs is also not well published.*
Projects	D-	*Some project information gets published, but the publication is bad.*
Budget Documents	Congress D White House B+	*The president's budget submission and congressional budget resolutions are a mixed bag.*
Budget Authority	Congress C- Executive Branch D	*Legal authority to spend is there, but not well articulated in the data.*
Warrants, Apportionments, and Allocations	I	*Spending authority is divided up in an opaque way.*
Obligations	C+	*Commitments to spend taxpayer money are visible some places.*
Parties	D+	*A proprietary identifier system makes it hard to know where the money is going.*
Outlays	C-	*After-the-fact aggregated data is pretty good. We need real-time, granular spending data.*

Notes

1. An illustration of the focus on the topic, Politifact.com's "Obameter" lists 35 Obama campaign promises from 2008 related to transparency, http://www.politifact.com/truth-o-meter/promises/obameter/subjects/transparency/.

2. Macon Phillips, "Change Has Come to Whitehouse.gov," *The White House Blog*, January 20, 2009 (12:01 p.m. EDT), http://www.whitehouse.gov/blog/change_has_come_to_whitehouse-gov.

3. Barack Obama, "Transparency and Open Government," Presidential Memorandum (January 21, 2012), http://www.whitehouse.gov/the-press-office/transparency-and-open-government.

4. Eric Holder, "Memorandum for Heads of Executive Departments and Agencies: The Freedom of Information Act (FOIA)," Office of the Attorney General (March 19 2009), http://www.justice.gov/ag/foia-memo-march2009.pdf.

5. Jim Snyder and Danielle Ivory, "Obama Cabinet Flunks Disclosure Test with 19 in 20 Ignoring Law," Bloomberg.com, September 28, 2012, http://www.bloomberg.com/news/2012-09-28/obama-cabinet-flunks-disclosure-test-with-19-in-20-ignoring-law.html.

6. Jim Harper, "Sunlight before Signing: Measuring a Campaign Promise," Cato@Liberty (blog), September 6, 2012, http://www.cato-at-liberty.org/sunlight-before-signing-measuring-a-campaign-promise/.

7. Jesse Lee, "Transparency and Open Government," May 21, 2009, http://www.whitehouse.gov/blog/2009/05/21/transparency-and-open-government.

8. Peter R. Orszag, "Memorandum for the Heads of Executive Departments and Agencies, Subject: Open Government Directive," M 10-06, December 8, 2009, http://whitehouse.gov/open/documents/open-government-directive.

9. "Negotiate Health Care Reform in Public Sessions Televised on C-SPAN," Politifact.com, http://www.politifact.com/truth-o-meter/promises/obameter/promise/517/health-care-reform-public-sessions-C-SPAN/.

10. Alon Peled, "When Transparency and Collaboration Collide: The USA Open Data Program," *Journal of the American Society for Information Science and Technology* 62, no. 11 (November 2011): 2085.

11. Harlan Yu and David G. Robinson, "The New Ambiguity of 'Open Government,'" *UCLA Law Review* 59, no. 6 (August 2012): 178.

12. Jim Harper, "Grading Agencies' High-Value Data Sets," *Cato@Liberty* (blog) February 5, 2010, http://www.cato-at-liberty.org/grading-agencies-high-value-data-sets/.

13. Jim Harper, "Is Government Transparency Headed for a Detour?" *Cato@Liberty* (blog), January 15, 2010, http://www.cato-at-liberty.org/is-government-transparency-headed-for-a-detour/.

14. Yu and Robinson, p. 182.

15. Open Government Partnership, http://www.opengovpartnership.org/.

16. Office of the Press Secretary, The White House, "Fact Sheet: The Open Government Partnership," September 20, 2011, http://www.whitehouse.gov/the-press-office/2011/09/20/fact-sheet-open-government-partnership.

17. "John Boehner Introduces the House GOP Congressional Transparency Initiative," http://www.youtube.com/watch?v=hDr70qRv_9k.

18. Library of Congress, "About THOMAS," http://thomas.loc.gov/home/abt_thom.html.

19. The text of the House rules package is available at http://www.rules.house.gov/News/PRArticle.aspx?NewsID=33.

20. Committee on House Administration, Legislative Data and Transparency Conference, http://cha.house.gov/about/contact-us/legislative-data-conference.

21. Andrew Weber, "Introducing Congress gov!" *Custodia Legis: Law Librarians of Congress* (blog) September 19, 2012, http://blogs.loc.gov/law/2012/09/introducing-congress-gov/.

22. See Jim Harper, "Congress on Transparency: 'Needs Improvement,'" *Cato@Liberty* (blog), September 23, 2011, http://www.cato-at-liberty.org/congress-on-transparency-needs-improvement/; and Jim Harper, "Government Spending Transparency: 'Needs Improvement' Is Understatement," *Cato@Liberty* (blog), December 14, 2011, http://www.cato-at-liberty.org/government-spending-transparency-%E2%80%98needs-improvement%E2%80%99-is-understatement/.

23. Jim Harper, "Publication Practices for Transparent Government," Cato Institute Briefing Paper no. 121, September 23, 2011, http://www.cato.org/publications/briefing-paper/publication-practices-transparent-government.

24. See Jim Harper, "Congress on Transparency:

'Needs Improvement'"; and Jim Harper, "Government Spending Transparency: 'Needs Improvement' Is Understatement."

25. U.S. Senate, "Senators of the 112th Congress," http://www.senate.gov/general/contact_information/senators_cfm.cfm.

26. U.S. Senate, "Committees," http://www.senate.gov/pagelayout/committees/d_three_sections_with_teasers/committees_home.htm.

27. U.S. House of Representatives, "Document Naming Conventions," http://cha.house.gov/sites/republicans.cha.house.gov/files/documents/committee_docs/CommitteeRepository-NamingConventions-v1-2-1.pdf.

28. U.S. Senate, "Daily Digest Committee Meet-ings/Hearings Schedule," http://www.senate.gov/pagelayout/committees/b_three_sections_with_teasers/committee_hearings.htm.

29. U.S. Government Printing Office, "Congressional Reports," http://www.gpo.gov/fdsys/browse/collection.action?collectionCode=CRPT.

30. U.S. House of Representatives, Office of the Clerk, "Legislation and Reports," http://clerk.house.gov/legislative/legvotes.aspx.

31. U.S. Senate, "Recent Votes," http://www.senate.gov/pagelayout/legislative/a_three_sections_with_teasers/votes.htm.

32. U.S. Office of Management and Budget, "Circular A-11," http://www.whitehouse.gov/omb/circulars_a11_current_year_a11_toc.

RELATED STUDIES FROM THE CATO INSTITUTE

Publication Practices forTransparent Government by Jim Harper, Cato Institute Briefing Paper no. 121 (September 23, 2011)

Reputation under Regulation: The Fair Credit Reporting Act at 40 and Lessons for the Internet Privacy Debate by Jim Harper, Cato Institute Policy Analysis no. 690 (December 8, 2011)

Leashing the Surveillance State: How to Reform Patriot Act Surveillance Authorities by Julian Sanchez, Cato Institute Policy Analysis no. 675 (May 16, 2011)

Electronic Employment Eligibility Verification: Franz Kafka's Solution to Illegal Immigration by Jim Harper, Cato Institute Policy Analysis no. 612 (March 5, 2008)

Effective Counterterrorism and the Limited Role of Predictive Data Mining by Jeff Jonas and Jim Harper, Cato Institute Policy Analysis no. 584 (December 11, 2006)

Understanding Privacy—and the Real Threats to It by Jim Harper, Cato Institute Policy Analysis no. 520 (August 4, 2004)

RECENT STUDIES IN THE
CATO INSTITUTE POLICY ANALYSIS SERIES