

# Bulk Data Task Force Update

## Government Publishing Office

Lisa LaPlant, Program Manager

July 2019

# Govinfo System



Content  
Management  
System

Securely controls digital content throughout its lifecycle to ensure authenticity and integrity.

Preservation  
Repository

Follows archival system standards to ensure long-term preservation and access of digital content.

Public Access  
Website

Combines extensive metadata with modern search technology.

# Govinfo Developer Hub

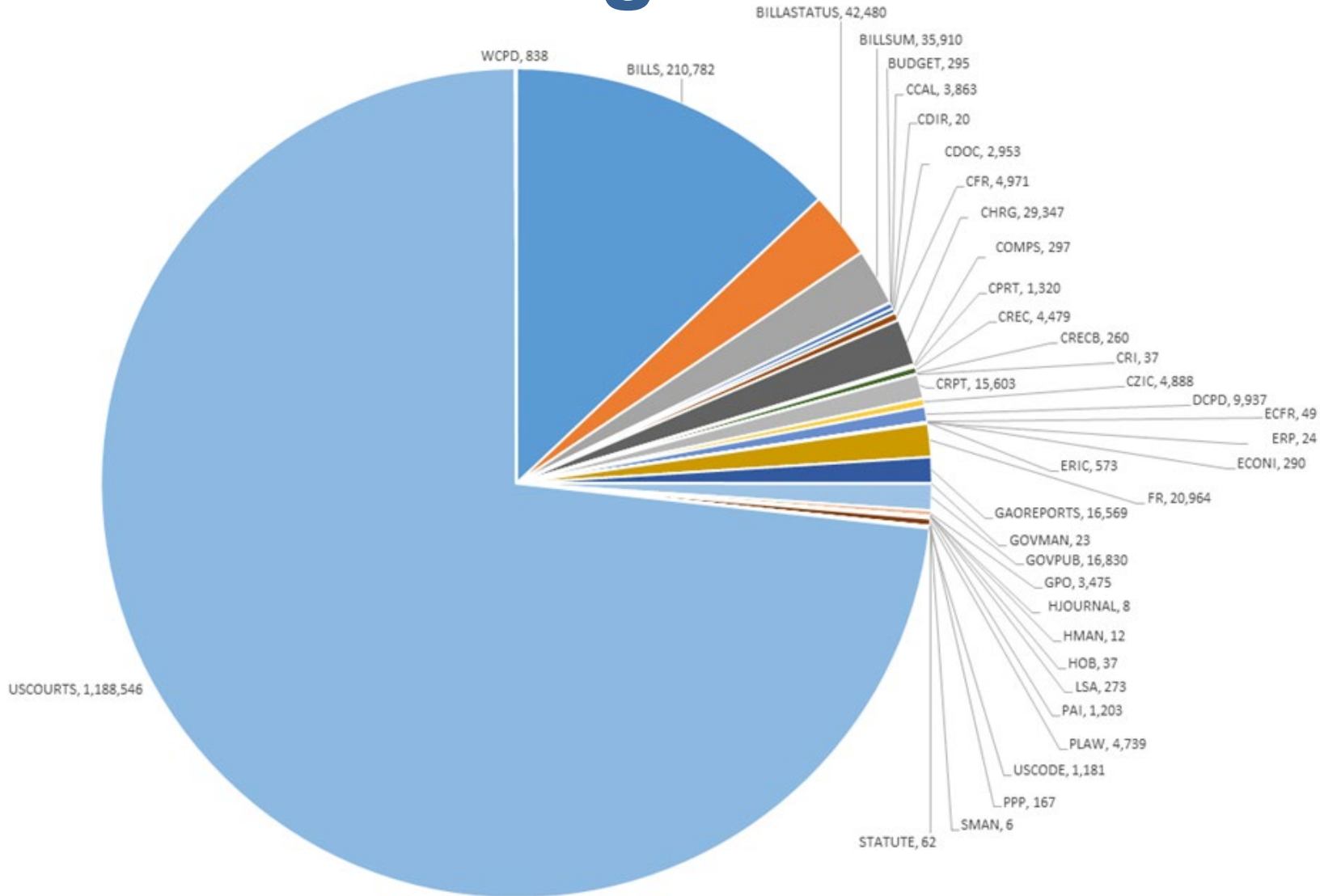
- **Developer Hub**  
[www.govinfo.gov/developers](http://www.govinfo.gov/developers)
- **Bulk Data Repository with JSON and XML Endpoints**  
[www.govinfo.gov/bulkdata](http://www.govinfo.gov/bulkdata)
- **RSS Feeds for Content and Metadata**  
[www.govinfo.gov/feeds](http://www.govinfo.gov/feeds)
- **Sitemaps**  
[www.govinfo.gov/sitemaps](http://www.govinfo.gov/sitemaps)
- **API**  
[api.govinfo.gov](http://api.govinfo.gov)
- **Link Service**  
[www.govinfo.gov/link-docs](http://www.govinfo.gov/link-docs)
- **GitHub Repositories and Issue Tracking**  
[www.github.com/usgpo](http://www.github.com/usgpo)

# FY18 Metrics and Fun Facts

- 3.5B retrievals from the public website since January 2009.
- 35M average monthly retrievals from the public website in FY18.
- 1.55M Access Content Packages on the public website.
- Over 6.4M PDF files and 12.5M image files in the preservation repository.
- 239K annual package ingest rate excluding replacements.
  - 4,605 average weekly rate
  - 1,626 median weekly rate
- More than 50 Production code deployments completed since 2011.
- Over 800 requests for changes to the system completed since 2011.
- 14,836 issues of the digitized Federal Register available.
- 51 volumes of the digitized Statutes at Large available.
- 144 volumes of the digitized Bound Congressional Record available.
- 131 Federal, district, appellate, and bankruptcy courts depositing opinions.

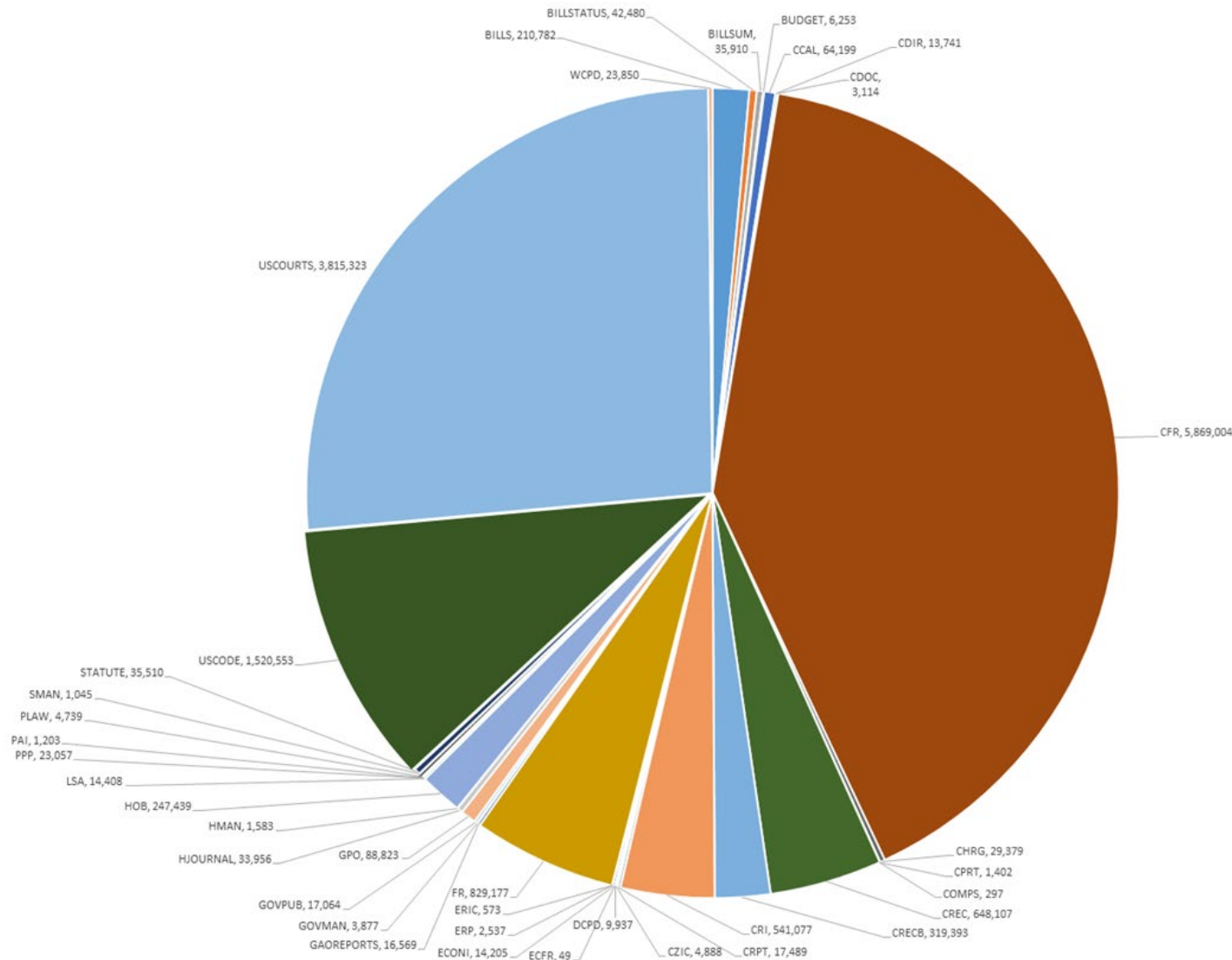
# Public Site Packages

June 2019



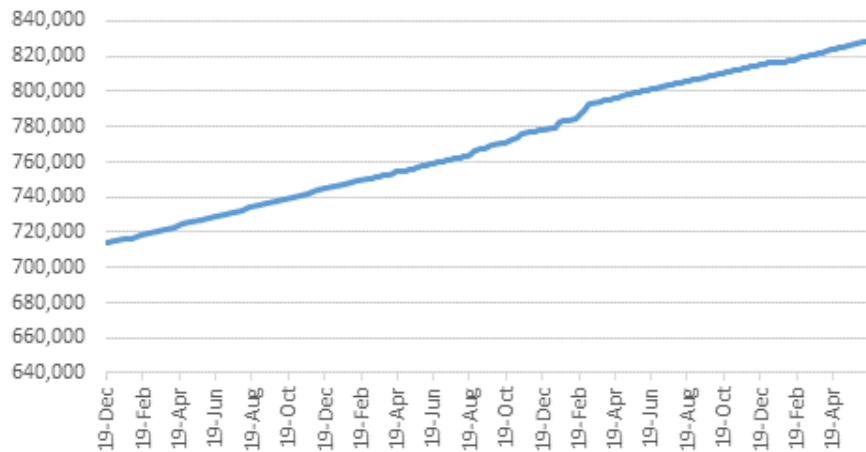
# Public Site Granules

June 2019

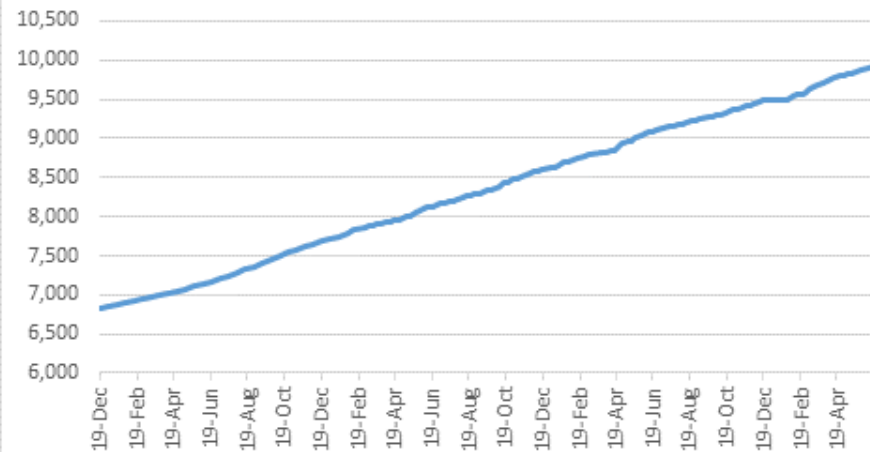


# Sample Monthly Growth Rate

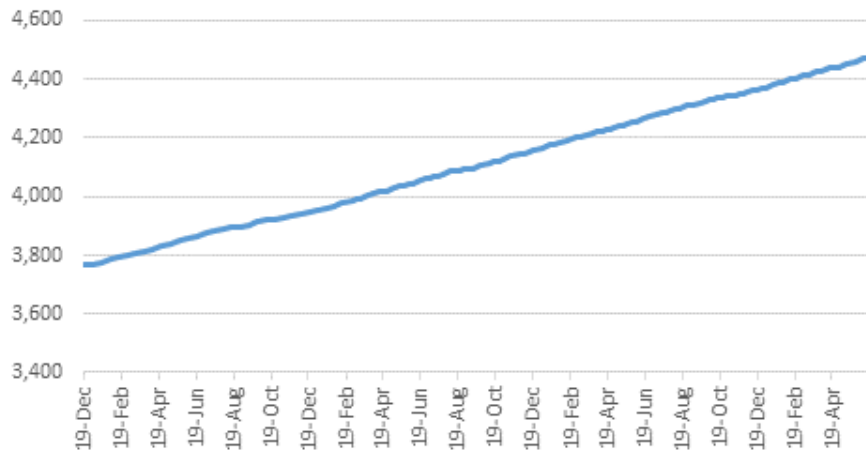
FR Granules - 828,579 (+569)



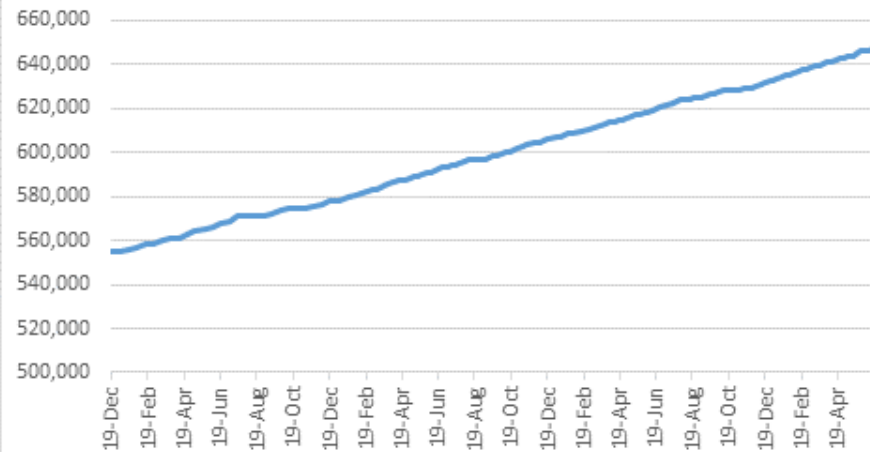
DCPD Packages - 9,919 (+11)



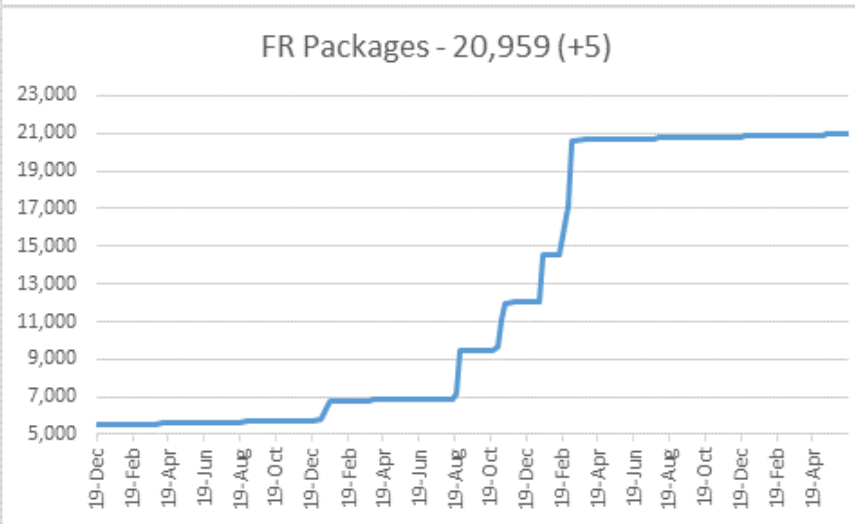
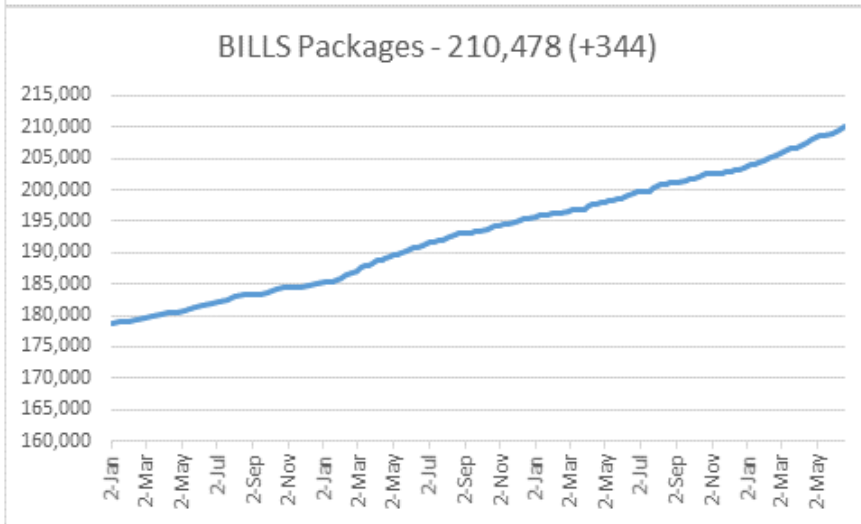
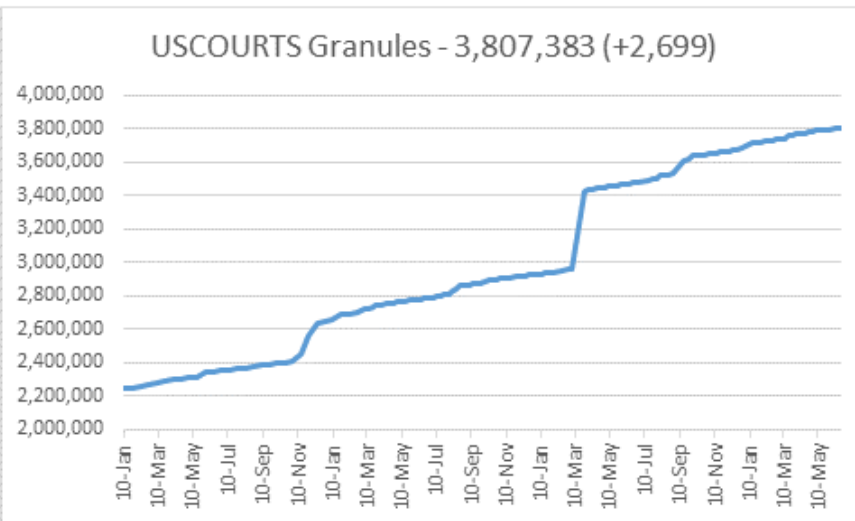
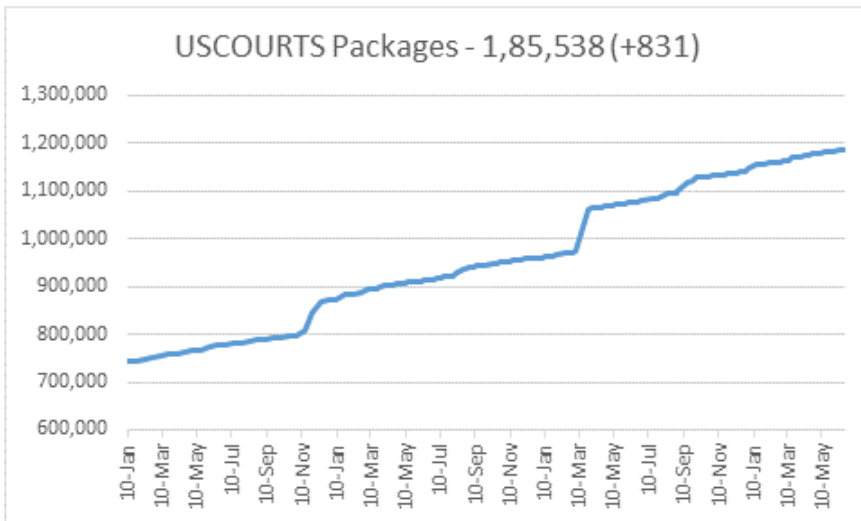
CREC Packages - 4,474 (+5)



CREC Granules - 647,451 (+789)



# Sample Monthly Growth Rate





# Govinfo Update Highlights

- FDsys website was retired and replaced by govinfo.
- GPO is the first organization in the United States and second organization in the world to receive ISO 16363 Trustworthy Digital Repository certification.
- See also [www.govinfo.gov/features/release-notes](https://www.govinfo.gov/features/release-notes).
- Newly added content includes Federal Register and CFR Indices, over 1700 National Recovery Administration publications in partnership with the University of Florida, the Public Papers of the Presidents back to Hoover in 1929, and responsive linked HOB and CRI with MD5 hashes in API response JSON (GitHub API [#10](#)/[#11](#)/[#41](#)).

# CRI and MD5 JSON Examples

- <https://www.govinfo.gov/content/pkg/CRI-2019/html/CRI-2019-GOVERNMENT-PUBLISHING-OFFICE.htm>
- [https://api.govinfo.gov/packages/CRI-2019/granules/?api\\_key=DEMO\\_KEY&pageSize=100&offset=0](https://api.govinfo.gov/packages/CRI-2019/granules/?api_key=DEMO_KEY&pageSize=100&offset=0)

## GOVERNMENT PUBLISHING OFFICE

### *Bills and resolutions*

Committee on Printing (Joint): election of members (see [S. Res. 86](#)), [S1594](#) [28FE] (see [H. Res. 226](#)), [H2717](#) [13MR]

Federal Register: modernize (see [S. 1078](#)), [S2328](#) [9AP] (see [H.R. 1654](#)), [H2614](#) [8MR]

Legislative branch of the Government: making appropriations (see [H.R. 2779](#)), [H3922](#) [16MY]

Public documents: establish and maintain a public website allowing access to electronic copies of all congressionally mandated reports (see [S. 195](#)), [S315](#) [19JA] (see [H.R. 736](#)), [H1171](#) [23JA]

### *Remarks in House*

Committee on Printing (Joint): election of members (H. Res. 226), [H3169](#) [9AP]

Federal Register: modernize (H.R. 1654), [H2671-H2673](#) [12MR]

Legislative branch of the Government: making appropriations (H.R. 2779), [H4425](#) [11JN]

### *Remarks in Senate*

Committee on Printing (Joint): election of members (S. Res. 86), [S1606](#) [28FE]

### *Reports filed*

Access to Congressionally Mandated Reports Act: Committee on Homeland Security and

```
{
  "count": 9721,
  "offset": 0,
  "pageSize": 100,
  "nextPage": "https://api.govinfo.gov/packages/CRI-2019-granules/?api_key=DEMO_KEY&pageSize=100&offset=100",
  "results": [
    {
      "title": "SMALL BUSINESS ACT",
      "granuleId": "CRI-2019-SMALL-BUSINESS-ACT/summary",
      "md5": "2B08B92756DF1F99B796085F18217460",
      "granuleLink": "https://www.govinfo.gov/content/pkg/CRI-2019-SMALL-BUSINESS-ACT/summary"
    },
    {
      "title": "SMALL BUSINESS ADVOCACY IMPROVEMENTS ACT",
      "granuleId": "CRI-2019-SMALL-BUSINESS-ADVOCACY-IMPROVEMENTS-ACT/summary",
      "md5": "11AA5B6C50AE694289F63295C02DA224",
      "granuleLink": "https://www.govinfo.gov/content/pkg/CRI-2019-SMALL-BUSINESS-ADVOCACY-IMPROVEMENTS-ACT/summary"
    },
    {
      "title": "SMALL BUSINESS CYBER TRAINING ACT",
      "granuleId": "CRI-2019-SMALL-BUSINESS-CYBER-TRAINING-ACT/summary",
      "md5": "9ACB7EEC816B460EE2CF2F369D4C40F5",
      "granuleLink": "https://www.govinfo.gov/content/pkg/CRI-2019-SMALL-BUSINESS-CYBER-TRAINING-ACT/summary"
    },
    {
      "title": "SMALL BUSINESS INVESTMENT IMPROVEMENT ACT",
      "granuleId": "CRI-2019-SMALL-BUSINESS-INVESTMENT-IMPROVEMENT-ACT/summary",
      "md5": "9ACB7EEC816B460EE2CF2F369D4C40F5",
      "granuleLink": "https://www.govinfo.gov/content/pkg/CRI-2019-SMALL-BUSINESS-INVESTMENT-IMPROVEMENT-ACT/summary"
    }
  ]
}
```

# GitHub Bill Status and MARC XML

- Bill Status Repository
  - <https://github.com/usgpo/bill-status>
  - Issues and Requests
    - 119 Closed
    - 9 Open
  - Highlights
    - Monitored and closed operational issues.
    - Added API for Bill Status in Bulk Data Repository.
    - Added <textVersions> to Bill Status XML - [#45](#).
- Cataloging Records Repository
  - <https://github.com/usgpo/cataloging-records>.
  - Added MARC XML for new or changed cataloging records.

# Bill Status API Examples

- [https://api.govinfo.gov/collections?api\\_key=DEMO\\_KEY](https://api.govinfo.gov/collections?api_key=DEMO_KEY)

```
{
  "collections": [
    {
      "collectionCode": "BILLS",
      "collectionName": "Congressional Bills",
      "packageCount": 211122,
      "granuleCount": null
    },
    {
      "collectionCode": "BILLSTATUS",
      "collectionName": "Congressional Bill Status",
      "packageCount": 20907,
      "granuleCount": null
    },
    {
      "collectionCode": "BUDGET",
      "collectionName": "United States Budget",
      "packageCount": 6256,
      "granuleCount": 5961
    },
    {
      "collectionCode": "CCAL",
      "collectionName": "Congressional Calendars",
      "packageCount": 64384,
      "granuleCount": 60510
    }
  ]
}
```

- [https://api.govinfo.gov/collections/BILLSTATUS/2019-07-01T20%3A18%3A10Z?offset=0&pageSize=100&congress=116&api\\_key=DEMO\\_KEY](https://api.govinfo.gov/collections/BILLSTATUS/2019-07-01T20%3A18%3A10Z?offset=0&pageSize=100&congress=116&api_key=DEMO_KEY)

```
{
  "count": 905,
  "message": null,
  "nextPage": "https://api.govinfo.gov/collections/BILLSTATUS/2019-07-01T20:18:10Z?offset=100&pageSize=100&congress=116",
  "previousPage": null,
  "packages": [
    {
      "packageId": "BILLSTATUS-116hr3164",
      "lastModified": "2019-07-03T14:07:56Z",
      "packageLink": "https://api.govinfo.gov/packages/BILLSTATUS-116hr3164/summary",
      "docClass": "hr",
      "title": null,
      "congress": "116"
    },
    {
      "packageId": "BILLSTATUS-116hr3055",
      "lastModified": "2019-07-03T14:07:51Z",
      "packageLink": "https://api.govinfo.gov/packages/BILLSTATUS-116hr3055/summary"
    }
  ]
}
```

- [https://api.govinfo.gov/packages/BILLSTATUS-116s1843/summary?offset=0&pageSize=100&congress=116&api\\_key=DEMO\\_KEY](https://api.govinfo.gov/packages/BILLSTATUS-116s1843/summary?offset=0&pageSize=100&congress=116&api_key=DEMO_KEY)

```
{
  "collectionCode": "BILLSTATUS",
  "collectionName": "Congressional Bill Status",
  "category": "Bills and Statutes",
  "dateIssued": "2019-07-03",
  "packageId": "BILLSTATUS-116s1843",
  "download": "https://www.govinfo.gov/bulkdata/BILLSTATUS/116/s/BILLSTATUS-116s1843.xml",
  "branch": "legislative",
  "governmentAuthor1": "Library of Congress",
  "governmentAuthor2": "Congressional Research Service",
  "billType": "s",
  "congress": "116",
  "billNumber": "1843",
  "lastModified": "2019-07-03T14:04:46Z"
}
```

- <https://www.govinfo.gov/bulkdata/BILLSTATUS/116/s/BILLSTATUS-116s1843.xml>

```
<textVersions>
  <item>
    <type>Introduced in Senate</type>
    <date>2019-06-13T04:00:00Z</date>
  </formats>
  <item>
    <url>
      https://www.govinfo.gov/content/pkg/BILLS-116s1843is/xml/BILLS-116s1843is.xml
    </url>
  </item>
</textVersions>
```

# Try It and Provide Feedback via GitHub

The screenshot shows a web browser at the URL <https://api.govinfo.gov/docs/>. The navigation menu includes 'govinfo', 'Browse', 'About', 'Developers' (highlighted), 'Features', 'Help', and 'Feedback'. An 'AUTHORIZE' button with a lock icon is visible in the top right. The main content is organized into two sections: 'Collections' and 'Packages', each with a dropdown arrow. The 'Collections' section lists three endpoints: 1) GET /collections: Request list of collections. Response includes collectionCode, collectionName, package and granule counts. 2) GET /collections/{collection}/{startDate}: Retrieve new or updated packages for a collection given a start date and time. 3) GET /collections/{collection}/{startDate}/{endDate}: Retrieve new or updated packages for a collection within a date range. The 'Packages' section lists five endpoints: 1) GET /packages/{packageId}/summary: Return json summary for specified package. 2) GET /packages/{packageId}/{contentType}: Return content or metadata for specified package. 3) GET /packages/{packageId}/granules: Get a list of granules associated with a package. 4) GET /packages/{packageId}/granules/{granuleId}/summary: Return json summary for specified granule. 5) GET /packages/{packageId}/granules/{granuleId}/{contentType}: Return content or metadata for specified granule. Each endpoint is shown in a light blue box with a lock icon on the right.

← → ↻ <https://api.govinfo.gov/docs/> ☆ 👤

**govinfo** Browse About **Developers** Features Help Feedback

**AUTHORIZE** 🔒

## Collections

 Identify new and updated documents ▾

- GET** /collections Request list of collections. Response includes collectionCode, collectionName, package and granule counts 🔒
- GET** /collections/{collection}/{startDate} Retrieve new or updated packages for a collection given a start date and time 🔒
- GET** /collections/{collection}/{startDate}/{endDate} Retrieve new or updated packages for a collection within a date range 🔒

## Packages

 Return content and metadata for individual packages ▾

- GET** /packages/{packageId}/summary Return json summary for specified package 🔒
- GET** /packages/{packageId}/{contentType} Return content or metadata for specified package 🔒
- GET** /packages/{packageId}/granules Get a list of granules associated with a package 🔒
- GET** /packages/{packageId}/granules/{granuleId}/summary Return json summary for specified granule 🔒
- GET** /packages/{packageId}/granules/{granuleId}/{contentType} Return content or metadata for specified granule 🔒

# Documents in USLM

- Completed BDTF project to convert a subset of enrolled bills, public laws, and the Statutes at Large from GPO locator-coded text format into Beta USLM XML.
- Scope of project was enrolled Bills and Public Laws (113th Congress, 2013 forward) and Statutes at Large (108th Congress, 2003 forward).
- Beta USLM XML is available on the govinfo website and in the Bulk Data Repository.
- More information and the draft 2.0.4 USLM XML schema is available at <https://github.com/usgpo/uslm/tree/proposed>.

# Statute Compilations in USLM

Goal is to provide a uniform set of laws in USLM to enable downstream processes and increase efficiencies.

## Phase 1 – 297 COMPS Available

- [www.govinfo.gov/app/collection/comps](http://www.govinfo.gov/app/collection/comps)
- In coordination with HOLC/SOLC/Clerk/Secretary, make select Statute Compilations in PDF available on govinfo as a Beta.
- Ingest legacy COMP DTD XML & locator files into govinfo CMS.

## Phase 2 – Procurement Pending

- Convert legacy COMP DTD XML or locator files into USLM XML.
- Provide access to COMPS in USLM XML as Bulk Data.

# Composition System Replacement

- XML-based composition system to replace legacy Microcomp system and locator-coded text format.
- Currently composing the 2018 main edition of the United States Code through the new XML-based system.
- First volumes are going to press very soon.
  - Volume 1: Titles 1-4, part of 5
  - Volume 2: Titles 5, 6, part of 7
  - Volume 3: Title 7



# Contacts and Sites

[llaplant@gpo.gov](mailto:llaplant@gpo.gov)

[@lisalaplant](https://twitter.com/lisalaplant)

[github.com/usgpo](https://github.com/usgpo)

[usgpo.github.io/innovation](https://usgpo.github.io/innovation)

[github.com/usgpo/uslm](https://github.com/usgpo/uslm)

[govinfo.gov](https://govinfo.gov)

[api.govinfo.gov](https://api.govinfo.gov)