

Instructions for Completing the GOS Collection Dataset Approval for Inclusion in Data.gov

All accompanying documents associated with these instructions must be obtained through contacting your agency [FGDC Coordination Group member](#). Your FGDC Coordination Group member has access to these documents at the following controlled site: <https://my.usgs.gov/alfresco/n/browse/workspace/SpacesStore/f6e01bb4-e4b6-4f45-b0d2-30b75bce0a3c>.

Summary of Agency Participation

An overview of Agency participation in Data.gov can be seen in the “Data.gov Agency Participation Summary Spreadsheet” in the site referenced above. This summary of participation shows the number of geodata records currently shared with data.gov versus the Agency’s current downloadable datasets published in GOS. This summary provides an insight into potential candidates for inclusion to data.gov. The first tab show the Agencies with the most datasets potentially available for data.gov, the second tab show the results for all Agencies with contacts.

High Level Work Flow Overview:

I) Data sets not in GOS:

1) Publish metadata records to GOS. Datasets for inclusion in Data.gov must meet the Data.gov criteria (see last section on Criteria). Note these datasets must be downloadable datasets with direct download links.

Publishing your metadata can be done through link your Agency Clearinghouse nodes, a Web Accessible Folder (WAF), or using the GOS online form. Details can be found at <http://www.geodata.gov/gos/metadata/CreatePublishMetadata.pdf>

2) Once data sets are in GOS request new Agency Workbook through Geodata@usgs.gov and follow steps below.

II) Data Sets already in GOS:

1) Review downloadable datasets registered by your Agency in GOS. See Agency Workbooks for each FGDC member in the site referenced above. Datasets for inclusion in Data.gov must meet the Data.gov criteria (see last section on Criteria). Note these datasets must be downloadable datasets with direct download links. In conjunction with the Agency workbook you can use the GOS interface to look at your Agency publications. Instructions to assist with using the GOS interface to support your review are found in “Reviewing and Managing Records in GOS for Data.gov – FAQ” in the site referenced above.

2) Select proposed datasets - If data sets meet Data.gov criteria and has working links, flag data (Y) under column for record in the ‘Submit to POC review’ column.

3) POC Approval - Send completed Agency Workbook to your Agency POC (see “Agency Data.gov POC List” in the site referenced above) for Review. **Please send final Workbooks to your POC no later than May 1, 2010 to ensure their inclusion in the May anniversary release of Data.gov**

4) Submit approved spreadsheets to GOS - After records have been approved for Data.gov through all appropriate reviews, your spreadsheets can be returned to Geodata@usgs.gov for turning on data.gov flags. Also your Agency collection manger can turn on the flags through their GOS Admin Account.

Agency Workbook Instructions:

Within the “Agency Worksheets” space at the site referenced above, you will see a separate workbook containing lists of datasets by Agency currently published to the GOS catalog. Agencies are asked to review their respective workbooks. Note that these workbooks contain complete listings of metadata records currently in GOS those without a Y in the Data.gov column are not reflected in Data.gov. Ideally, most of these datasets will be nominated for inclusion in Data.gov – please see the Review Criteria section below.

Within each Agency Workbook, datasets have been sorted by ‘downloadable’ datasets and by ‘owner’ (the collection publisher). To view all datasets go to the ‘Type’ column and select ‘All’. This will enable you to see other potential candidate records that are not currently classified as downloadable. The workbook contains a listing of all the records published by the agency that is being asked to identify which datasets are approved for display in the Data.gov catalog. The information contained in each of the columns within this worksheet is described in the table below:

Column Label	Column Description
Docu UID (Tab No.)	Unique Identifier assigned to each metadata record
Name	Metadata Title
Agency Name	Name of the agency that owns the catalog publishing node
Owner	Agency Collection Publisher POC
EMAIL	Collection POC Email address
Affiliation	Agency GOS/OMB Code Number
Organization	Agency Name
Type	Data Type Code (Example: Code #2 = Downloadable)
ONLINK	On-link URL in the Agency metadata record
Data.gov	Existing Flag for inclusion in Data.gov (“Y” = include)
Metadata Link	Link to metadata record in GOS
Submit to POC	Flag record as candidate to include into data.gov
POC Approval	Flag if record has been approved through DMS process

Criteria for determining if a dataset in Geodata.gov should be referenced in Data.gov:

1. Begin by determining whether the agency is the owner or “authoritative data source¹” for the dataset associated with the metadata entry. You may only nominate your own data. Please use a code of “No” if the entry describes a dataset or other asset that is not originated in your agency (e.g., those owned by other Agencies or by nongovernmental organizations or by state, federal, or tribal governments).
2. Check each metadata entry to ensure that it is a) a dataset or b) data extraction or mining tool, as defined by Data.gov².
3. If the metadata entry is for a data extraction tool, again please ensure that the URL provided gets one directly to that site rather than to a contextual ‘program’ page.
4. If it is a dataset, please check the format. Only the following formats are currently being included in Data.gov: RSS, XML, CSV/TXT, Keyhole Markup Language (KML/KMZ), Microsoft Excel and ESRI Shapefile or, for live map services, OGC WMS/WFS.
 - o *Data in formats other than these will not be included.* In general, PDF & HTML documents are not published in the catalog. However, in some cases (such as Geodata maps) we will allow you to publish these data types as an exception to our rule. It is your call to determine if these documents are indeed datasets.
5. Please check the URL listed for each metadata record. The desired URL is the one that takes the user directly to a downloadable file containing geospatial data in one of the acceptable file formats.
 - o If the URL listed is to a program page (usually the link is in html format) that provides access to a number of data sets, please provide the multiple direct download URLs, and indicate whether the metadata record is identical for each of the datasets to which you are providing access
6. Datasets that require registration or a service fee for access are not acceptable for Data.gov.
7. Please look at each dataset to ensure that you feel it meets your obligations, as outlined within the certification document.

¹ A recognized or official data production source with a designated mission statement or source/product to publish reliable and accurate data for subsequent use by customers. An authoritative data source may be the functional combination of multiple, separate data sources.

² **Data Mining and Extraction Tools** offer applications that allow users to either produce maps, tables, or charts of the subset of data that are specific to the user's interests or to build their own dataset extracted from a data source. Note that many of these sites also offer downloadable data. Data mining tools that provide access to data that are not available for the raw data catalog are of particular interest.

- o Note that when determining whether the dataset is consistent your Agency's Information Quality Guidelines, you should apply a standard that is appropriate for such a high profile Presidential Initiative.
- o Examples of characteristics that make the dataset questionable for Data.gov include datasets that are:
 - outdated (e.g., boundaries have now changed),
 - not documented adequately enough for the Data.gov user to determine fitness for use,
 - not of known, documented quality.

When reviewing their GOS collection datasets, agencies are also encouraged to consider the guidelines in "Geospatial Data for Inclusion in Data.gov - Metadata Quality Considerations" in the site referenced above to identify other potential improvements in how geospatial metadata are currently presented in the GOS catalog.

For technical assistance with this request please contact Jerry Johnston (johnston.jerry@epa.gov) at 202-564-3175 or your agency's [Senior Agency Official for Geospatial Information \(SAOGI\)](#) or [FGDC Coordination Group member](#).