Data.gov, GeoPlatform.gov and ISO Metadata
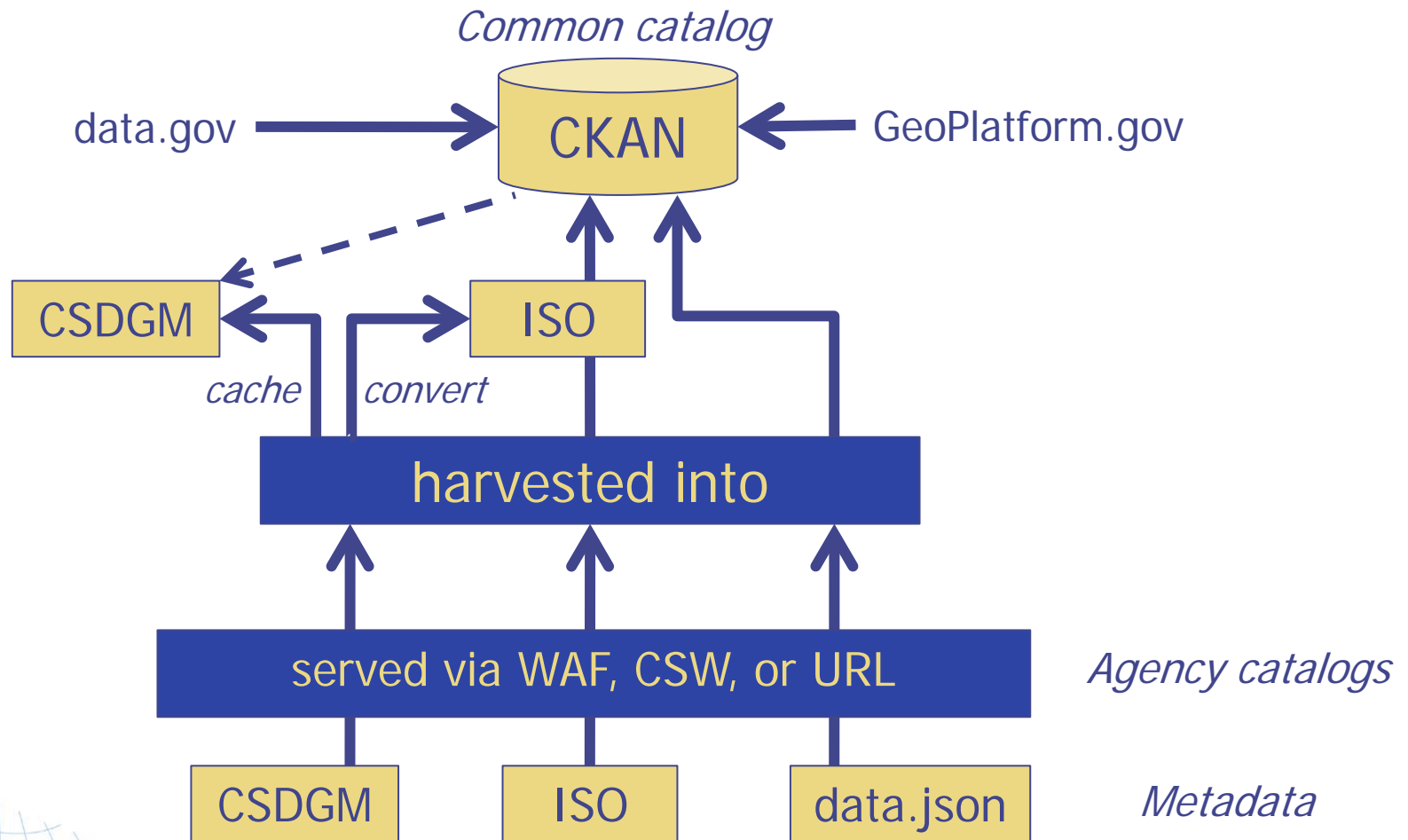
Doug Nebert

FGDC Secretariat

ISO Support

- The common CKAN catalog software used by GeoPlatform.gov and data.gov expects ISO 19115 or -2 metadata as XML

- Metadata style guidance is available under the Resources tab of geoplatform.gov

- Other forms of input metadata can be processed, such as FGDC/CSDGM and data.json
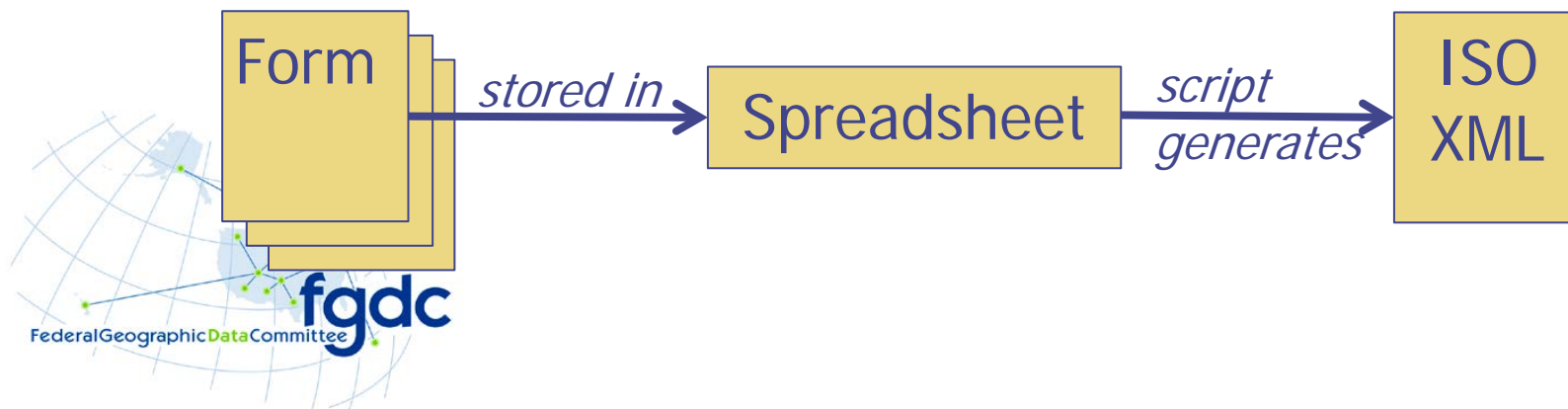
fgdc
FederalGeographicDataCommittee

Metadata flow

Open Data Policy

- OMB/OSTP published an Open Data Policy in May 2013 that requires all federal agencies to report their holdings

- Require use of a data.json format for syndication and reporting of metadata to OMB

- Styled after the DCAT catalog practice

- Metadata fields can be mapped to ISO

fgdc
Federal Geographic Data Committee

GoogleDocs approach

- DOI developed a metadata form that collects metadata fields that enable compliance with data.json and ISO 19115

- Deployed in GoogleDocs as a spreadsheet with a script

- Available across DOI, but others can use

| Form | *stored in* → | Spreadsheet | *script generates* → | ISO XML |

Metadata Collection Form for Project Open Data

This is a Google Form that saves your input in a Google Spreadsheet. Fill out the form for each piece of metadata you want to record. After you're finished, you can go to the spreadsheet associated with this form and export your information as ISO XML. The fields here are a combination of are mandatory fields in Project Open Data and the ISO 19115 Metadata Standard.

* Required

title *

Human-readable name of the asset. Should be in plain English and include sufficient detail to facilitate search and discovery. Example: NOAA Geotiff - 5 meter LiDAR Reflectivity, U.S. Caribbean - Puerto Rico (southwest), 2006, UTM 19N NAD83

description *

Human-readable description (e.g., an abstract) with sufficient detail so that a user can quickly understand if the asset is of interest.

identifier *

A unique ID for your resource. You can use any system you like here as long as each ID is unique within your Bureau or Office metadata, such as bureauName_accessURL, or bureauName_number, or DOI, or URN. Examples: bia_001 or gov.noaa.nodc.ncddc.file:http://nccddc.nodc.noaa.gov/data/ofr-98-38

theme

OPTIONAL. Select the category most suited to the data asset. These categories are ISO 19115 Topic Categories, and further explanation for what falls under each category is at
http://gcmd.nasa.gov/add/difguide/iso_topics.html

6

keyword *

Tags or Keywords help users discover your dataset. Please include terms that would be used by technical and non-technical users. Words should be quoted and have commas between entries. You can put your program name or acronym here to help people searching for your data. Example: "water","Arizona","hydrology","New Mexico","CERSC","Energy Program","Earthquakes","Hazards"

[   ]

bureauCode *

OMB code for your Bureau. Please refer to this document to find your code: http://goo.gl/ylqlFd

[   ]

contactPoint *

Contact person's name or position

[   ]

mbox *

Email associated with the POC

[   ]

modified *

Date of last update to the data resource

[mm/dd/yyyy   ]

accessLevel *

Can this dataset be made publicly available? Public means that your data asset is *or could be* made publicly available to all without restrictions. Restricted public means that your data asset is made available under certain use restrictions. Non-public means that your data asset is not available to members of the public. Example: A dataset that I can share without restriction if someone asks me for it, even though it isn't available for download online, would still be categorized as "public" here.

[▲▼]

Fed

access_Level_Comment

IF APPLICABLE. If the data asset is not already public, comment on the timeline for making it available, or the reason it cannot be made available. Example language: Non-public restricted: Data has not completed Bureau's Publication Process for open data. Data and documentation is incomplete. Non-public: Data contains information that is proprietary in nature/contains PII, or has other restrictions that exclude FOIA/etc.

accessURL

IF APPLICABLE. URL providing direct access to the described resource. Leave blank if data can't be accessed publicly.

format

IF APPLICABLE. The format of the resource as can be accessed at the resource URL. Please only enter one format type. If there are multiple available, choose the most prominent. Examples: csv, xls, pdf

webService

IF APPLICABLE. Endpoint of web service to access dataset. Leave blank if data can't be accessed publicly.

spatial

OPTIONAL. Geographic location. Location or extent of the resource. Examples: Lincoln, Nebraska or Chesapeake Bay

[ ]

westBoundLongitude

OPTIONAL. West bounding longitude in decimal degrees.

[ ]

eastBoundLongitude

OPTIONAL. East bounding longitude in decimal degrees.

[ ]

southBoundLatitude

OPTIONAL. South bounding latitude in decimal degrees.

[ ]

northBoundLatitude

OPTIONAL. North bounding latitude in decimal degrees.

[ ]

fgdc
FederalGeographicDataCommittee

dataDictionary

OPTIONAL. Enter the URL to data dictionary documentation.

dataDictionaryText

OPTIONAL. Provide in comma-separated, quoted list form, one row per element, in the order: "tablename.elementName", "elementDescription","elementDataType","unitsOfMeasure". Example: "HEADER_PGS.LITHO2","Gross lithology #2 Encoded text field.","text(20)"

dataQuality

OPTIONAL. Does this data asset meet your Bureau or Office's data quality guidelines?

Submit

Never submit passwords through Google Forms.

ISO Metadata Form

♦ Current form generates an XML file for each row in the spreadsheet on Google Drive, then can be uploaded/posted

♦ Plans are to modify the CKAN default entry form to collect similar data

- Cache as ISO XML

- Generate data.json and POD tags

- Support full query and retrieval via CSW