Metadata, Search and Discovery

November 2, 2018

John Davidson
GeoPlatform Technical Team, Image Matters LLC

GEOPLATFORM.gov

FGDC.GOV
FEDERAL GEOGRAPHIC DATA COMMITTEE

Search: Opportunities and Challenges

- Expand the NSDI with curated authoritative data and services aggregated across levels of government
- Find the authoritative, relevant data and services is still too difficult
  - Google Search is not good enough
  - Agency level catalogs and portals are disconnected and inconsistent
- #1 challenge: dearth of curated, authoritative, machine-readable metadata
- Key elements for search:
  - Controlled vocabularies to ground keywords, themes, and places to unique, unambiguous concepts
  - Consistent, unambiguous identification of the responsible agency
  - Descriptive titles and complete descriptions that explain why the dataset was produced e.g., its intended use
  - Unambiguous links to related resources (e.g., dataset->service, layer ->dataset)

For example, from the 2018 HIFLD Data Catalog

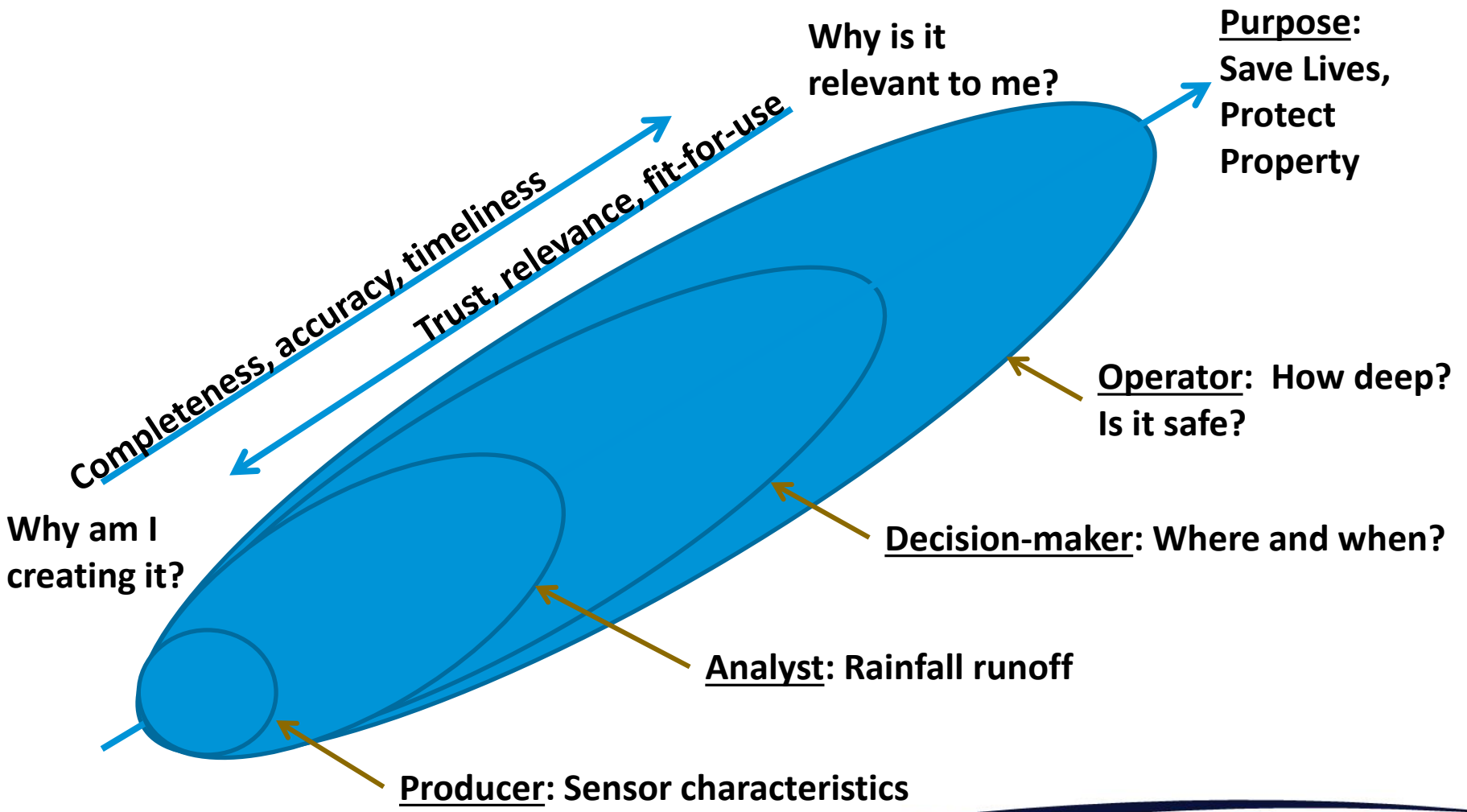
- A sampling of 112 HIFLD services (1/3 of published services)
  - 8% Unsuccessful URL Response Rate (9 Objects)
  - 92% Successful URL Response Rate (102 Objects) (import and validate with SSC)
    - 98% No Title in the service metadata (100 Objects)
    - 84% No Description in the service metadata (86 Objects)
    - 85% No Publisher in the service metadata (87 Objects)
    - 92% No Keywords in the service metadata (94 Objects)
    - 87% Metadata quality rating: 1 - Not good service metadata (89 Objects)
    - 10% Metadata quality rating: 2 - Serviceable service metadata (11 Objects)
    - 1% Metadata quality rating: 3 - Good service metadata (1 Objects)

This is a hurdle we're trying to overcome everywhere (not just HIFLD)!

Many Search Options ("NGDA Levee")

- Google Dataset Search
  - https://toolbox.google.com/datasetsearch/search?query=NGDA%20Levee&docid=rL4roxnCdsjrUpsZAAAAAA%3D%3D
- Esri ArcGIS Online
  - https://hub.arcgis.com/search?q=NGDA%20Levee
  - http://open-fedmaps.opendata.arcgis.com/datasets?q=NGDA+Levees&sort_by=relevance
- Data.gov
  - https://catalog.data.gov/dataset?q=NGDA+Levee&sort=score+desc%2C+name+asc
- GeoPlatform.gov
  - https://www.geoplatform.gov/geoplatform-search/#/?q=NGDA%20Levee

Common purpose, distinct vocabularies



MODI: A most excellent vocabulary (an Ontology) for Disaster data

- FEMA ESFLG Model and Data Inventory (MODI)
- An Inventory of Datasets/ Models used by Emergency Management
- https://gis.fema.gov/Model-and-Data-Inventory/index.html

Lifeline

MODI Vocabularies

Core Capability

Resource Type (Dataset/Model Type)

User (Agency)

Owner (Agency)

Recovery Support Functions (RSF)

Emergency Phase

Hazard

Emergency Support Functions (ESF)

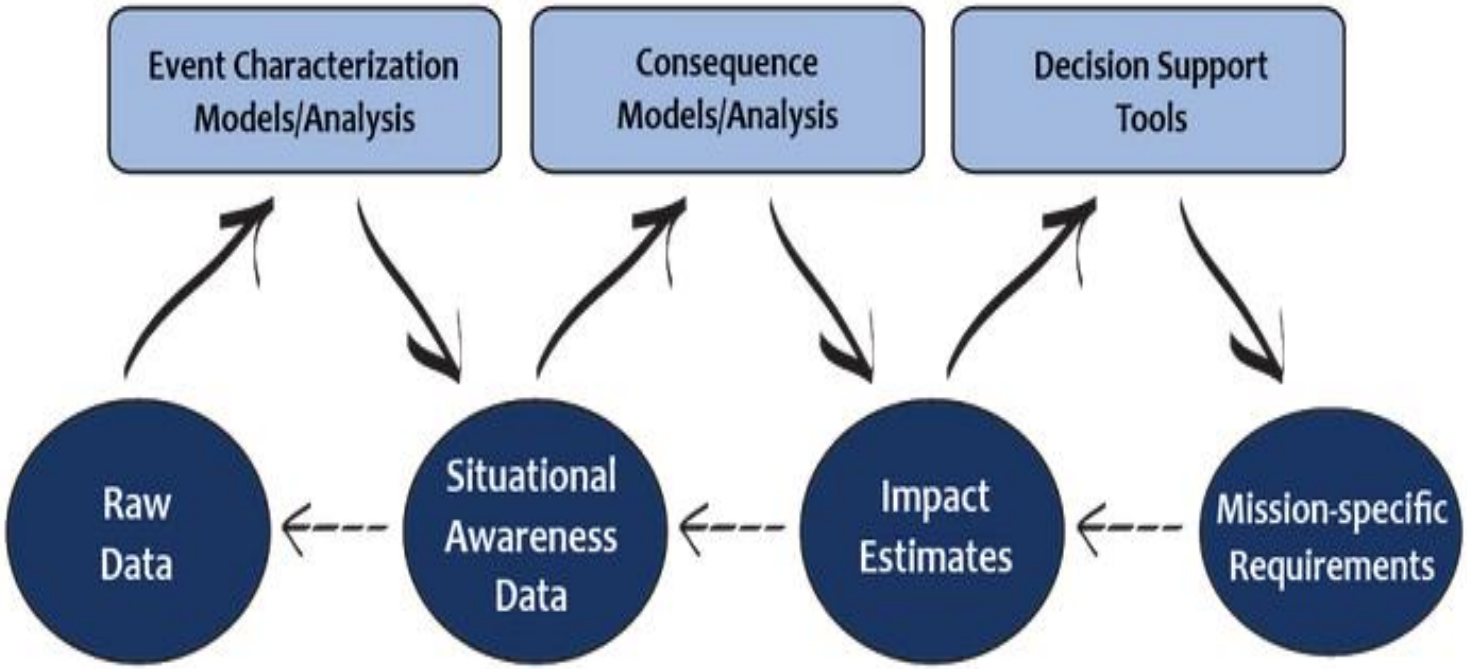
Keyword

Output File Type

Access Mechanism

ESFLG Model and Data Inventory

https://gis.fema.gov/Model-and-Data-Inventory/



Flow of information through iterative rounds of data and modeling. Data sources are shown in dark blue; models and data processing tools are in light blue.

A slice of MODI

https://gis.fema.gov/Model-and-Data-Inventory/index.html#search

Constraints: Phase - Short-term Recovery; User - USDA; RSF - Health & Social Services

HAZARD	OWNER (AGENCY)	ESF	RSF	KEYWORD
None Selected	None Selected	None Selected	Health And Social...	None Selected
CORE CAPABILITIES	**USER (AGENCY)**	**EMERGENCY PHASE**	**ACCESS MECHANISM**	**OUTPUT FILE TYPE**
None Selected	USDA	Short-term Recove...	None Selected	None Selected

Show all/default filters

FILTERED BY: PHASE: *Short-term Recovery* ✘ USER: *USDA* ✘ RSF: *Health and Social Services* ✘ Remove All

2 DATASETS/MODELS FOUND

Shaded Name = *This dataset/model has multiple functions and is tagged by more than one resource type.*

Datasets/models are sorted, in descending order, by the number of federal agency-level users using the dataset/model.

Click on the column headers or hover over a dataset/model to view additional details.

RAW DATA	EVENT CHARACTERIZATION	SITUATIONAL AWARENESS	CONSEQUENCE MODELS	IMPACT ESTIMATES	DECISION SUPPORT TOOLS	MISSION-SPECIFIC REQUIREMENTS
	NARAC Modeling System	NARAC Modeling System	NARAC Modeling System	NARAC Modeling System		
	QUIC (IMAAC)		QUIC (IMAAC)			

2 Datasets/Models

MODI to GeoPlatform RIM/KG Mappings

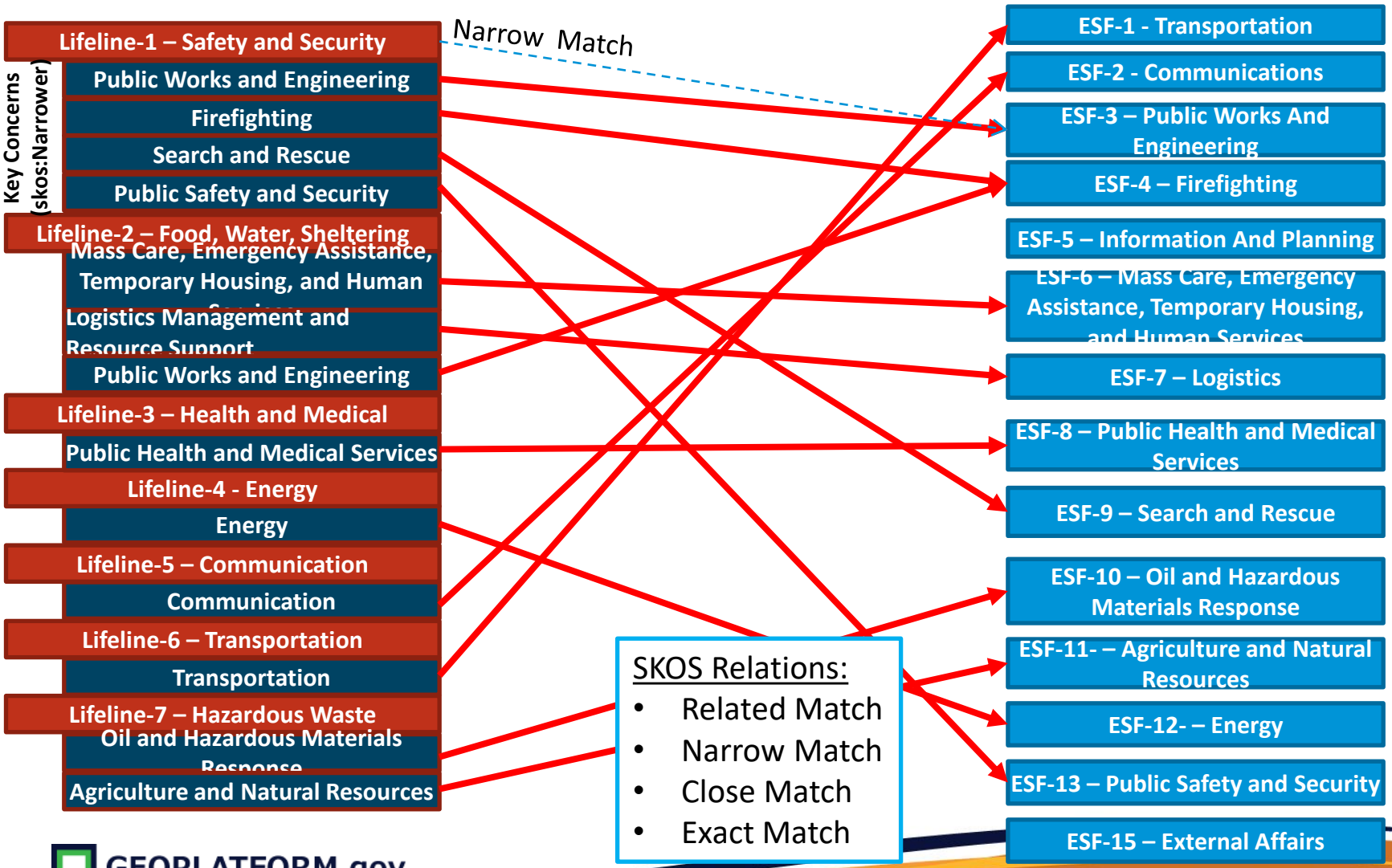
GeoPlatform RIM/KG

- Purpose
- Function
- Topic
- Subject
- *Community*
- Audience
- *Place*
- Category
- Publisher
- Keyword

FEMA MODI Vocabularies

- *Lifeline*
- Resource Type (Data/Model Type)
- Hazard
- Owner (Agency)
- ESF
- RSF
- Keyword
- Core Capability
- User (Agency)
- Emergency Phase
- *Access Mechanism*
- *Output File Type*
- *…*

FEMA Lifelines to ESF Mappings



Demonstration of GeoPlatform Search

Open Discussion

GEOPLATFORM.gov

FGDC.GOV

FEDERAL GEOGRAPHIC DATA COMMITTEE