

Model Catalog Working Group

Version 2, May 15, 2020.

Convened by Community for Data Integration facilitators

Contact: Leslie Hsu, cdi@usgs.gov

This is a document that summarizes initial discussions to get diverse perspectives on expectations for a USGS Model Catalog. This is just one effort informing requirements for a pilot model catalog, a project in the Science Analytics and Synthesis Program for FY20. Background and motivation for this effort is on [this wiki page](#) and copied at the end of this document.

Process	2
Perceived Gaps	2
Highlights from Discussions	2
What models to include in a pilot:	2
With regard to USGS-authored-only or USGS-used models	3
What information to include about the models:	3
How to organize and filter the models? How would one search for models?	4
What cases could this catalog support?	4
What does model mean to you?	5
Feedback on the draft catalog at the ScienceBase page	5
Other comments	5
Participants and Further Suggested Participants	6
Responses to Questions	7
What are your expectations/criteria when you hear “USGS model catalog”?	7
What are the major, established models in USGS that come to your mind first?	8
Where are existing locations to find the various data/code pieces for USGS models?	10
Brainstorm and crowdsource a list of “_____ models” (what model types do you think of? No overthinking. Just how would you complete the phrase “_____ models”.)	11
Sharing an INITIAL list of models, what else belongs on the list? What other initial reactions do you have? (se...	13
Bonus question: How do you define “model”?	14
“Anything else you would like to share. If you’d like to share your name, program, etc. you can do so here. Other topics of interest – how about models that are heavily used by USGS but do not ori...”	16

Other Notes	19
References and Resources	20

Process

1. Initial participants were chosen because of their previous involvement with the CDI (see section on Participants for more information)
2. Additional participants were added from Leslie's network or suggestions from invitees
3. [Background information](#) was provided in the invitation.
4. We scheduled two group discussions with 3-5 participants for 60 minutes, where that was not possible, meetings with 1-2 participants were scheduled for 30-60 minutes. In one case the form was filled out without a meeting.
5. All meetings were conducted on MSTeams and were a hybrid of conversation and filling in a [Microsoft Form](#).
6. [Form responses](#) were collected at the end of this document and highlights from the conversations extracted and grouped for the body of this document.
7. The focus of the conversations evolved as the meetings were held, for example, later conversations focused more on use cases and model labels and filtering, and other issues that arose from initial discussions.

Perceived Gaps

Social science, decision making, geophysics besides NEIC, more Energy and Minerals, and more.

Highlights from Discussions

What models to include in a pilot:

1. In narrowing down what type of model to start with for a pilot, and for the purposes of integrated modeling (EarthMAP), **participants emphasized process-based models (also referred to as physics-based models, deterministic models, numerical models) as a priority.**
2. It was noted that there is a much larger amount of statistical models and it would be hard to comprehensively deal with them.
3. "I would focus on models that are used consistently to update USGS data."
4. **"I would start with: 'what are we trying to model', instead of just a listing of a bunch of things that might not go anywhere."**
5. One participant asked "Are we saying that the USGS stands behind and will offer help for any of the models in this catalog?" One implication being (as I read it) that some model authors may not really want to be contacted.
6. A user would want to see up front the **relevant, maintained models with a contact person, not abandoned or out of date models.**

7. It was noted that one didn't think that half of the scripts on their website were worth putting into a catalog for the purposes of meeting an EarthMAP integrated modeling use case. "There are 100s to 1000s of supporting functions to large models. But there are diminishing returns to populating everything under that."
8. Whatever is decided to be the target group of models, just be sure to make that distinction up front and acknowledge what is not prioritized in the first stage (none of the working group participants were thinking of conceptual models or data models).

With regard to USGS-authored-only or USGS-used models

1. Most participants thought that it was important for the use case of EarthMAP integrated modeling to **include models authored outside of the USGS. USGS-used seemed like a good scope.** Other categories brought up were models that were created by USGS scientists, with USGS funding, or models that underlie USGS products.
2. Some USGS models include components that are authored outside of the USGS.
3. **For the use case of EarthMAP integrated modeling, one participant was adamant that it would be a mistake to use USGS-authored models only, because the discipline has its own community-supported, well-used, best-in-class models. To him, using USGS-authored only models would leave out the best model and best science for the job.**
4. One participant's job was to showcase and connect USGS modeling expertise, so it was not important to her to include non-USGS models.
5. Many participants noted that their peers used models that were authored outside of the USGS and those would be important if the use case was discovery of the best model to use.
6. "Yes, models from outside the USGS but are used within it are VERY common."
7. There was the other view too "We often use models that do not originate from the USGS but I don't think this should go in the USGS model catalog." "If we look at the USGS Publications Warehouse as a parallel, it does not include non-USGS publications."
8. For the case of a scientist in the Office of Foreign Disaster Assistance that assists with advising appropriate landslide models, having non-USGS models would be useful, because USGS models don't cover all of the possible situations. Some non-USGS models are commercial and come with high price, but might be the best model for the job. One could envision a catalog where non-USGS models were listed at lower levels of detail, but for completeness, and that could be toggled off the view.

What information to include about the models:

1. **Depending on the participant and their desires for using the catalog, the important fields/metadata were very different.** Important fields for some would be irrelevant for others.
2. Most participants assumed a catalog optimized for human-readability and navigation, with one participant thinking about machine-readability.
3. One participant wanted to see how much "support" the model had by the community, meaning, how much is it used, is it actively developed?
4. What are the inputs and outputs? Knowing in a detailed way what inputs and outputs are needed would help coordinators point partners to the correct model to use. For example, for Landslide models, one would want to know very specific information: what scale of digital elevation model is required, soil properties, soil strength, if the data is raster or vector.

5. For certain well-used Water models, having all of the possible inputs and outputs attached or linked would complicate things.
6. Number of citations (indicating the amount of use)
7. The date that the model was last updated.
8. What language is the model written in?
9. Is there a graphical user interface or is it command line?

How to organize and filter the models? How would one search for models?

1. Process-based / statistical / conceptual models
2. By discipline
3. What climate/environment the model is applicable to. (arid, etc.)
4. By mission area (important for some, not relevant for others)
5. By scripting language
6. By author
7. By contact person
8. By “application” or some notion of what the model is applicable to
9. The Water Mission Area is familiar with the terminology “core software” versus “application models” and that is a way they differentiate models. The difference is reusable core software and applications of that software to specific cases.
10. The Water Mission Area is used to categorizing things in the legacy structure: Water Quality and Chemistry; Groundwater; Statistics & Graphics; Surface Water
11. For landslides common categories would be: Susceptibility model, run-out model, initiation model, inundation model, Warning model. One would want to filter by hydrologic (soil moisture, pore pressure, rainfall threshold); landslide type (deepseated, shallow, dry/wet), climate.

What cases could this catalog support?

1. Pointing partners to USGS model expertise, (for example for contaminant response, “**connecting science to events**”). This is different from another use case of pointing partners to the best model to use, whether USGS-authored or not.
2. **Helping to understand what models had been run with what parameters in the past, to avoid re-doing model runs.** (Part of a much larger issue where model results are subset for a publication-supporting data release and the rest are discarded. Waste of resources/time when that data might be used for other studies.)
3. **Helping with the Water Mission Area Model Archive requirement** – a requirement unique to the water mission area. How exactly to do this was not fleshed out but some way the information in this catalog could support that would be a big benefit.
4. **Model intercomparison and evaluation.**
5. **For selecting models to integrate for EarthMAP.**
6. **For showcasing to the public (Congress or otherwise) the work of the USGS.**
7. **For discovery of relevant models for a single researcher’s work.**
 - a. “trying to find a model outside of your domain is difficult and this model catalog would help facilitate collaborations for model coupling projects.”

What does model mean to you?

1. A large number of respondents prioritized process-based models.
2. "A conceptual (and often quantified) representation of a process."
3. "I think clarifying what you mean by "model" will be crucial...there are MANY statistical models that hypothetically could be included, so avoiding these will be very helpful."
4. "There are conceptual, process and statistical models. For this catalog it seems the focus might be on process models."
5. "In this context, I believe "model" should be a process-based, mechanistic simulation model that has been supported by USGS."
6. "Something that represents our best understanding through equations or reproducible computations. Not just qualitatively talking about how things work."
7. There was considerable questioning of the differences between models, software, and tools, and what we were interested in for this catalog.
8. More answers at the end of this document.

Feedback on the draft catalog at the ScienceBase page

1. A description of the acronyms would be very helpful.
2. Also note in Land Resources (and likely elsewhere) there are a number of land-use related models not on this list, models that have a long history of use within USGS. FORE-SCE, CBLCM, and LUCAS are the ones that come to mind. This may be because there is a lot of internal use without a lot of publication about them.
3. **"The description is very comprehensive and may not be a reasonable aspiration** (to have everything!) I would rather start top down with the main goals (is it populating EarthMAP or making things discoverable). Phrases like "unambiguously link" and "everything associated with them," sets an aspiration that does not seem reasonable.
4. **"I would love to see a more systematic survey of USGS so that the content isn't biased** by who is and isn't asked. Collecting the information at a Science Center level would seem to me like the easiest way to do so."

Other comments

1. Question: **"How do you fit in machine learning models which don't have a specific name other than the broad tool** (e.g. Recurrent Neural Network – used in water resources for temperature prediction)"
 - a. Response: I am not sure if what I think of as "methods" like Recurrent Neural Network, fit into this catalog as we are envisioning it. That would be more like a catalog of "tools or methods."
2. **Keeping the catalog up to date will be a big issue** – even contact people change often. Maybe there can be a way to indicate the last time the information was vetted and have an "expiration date" or disclaimer.
3. A vision of what could be, but as an editorial comment this is not the dominant perspective of who we talked to: "I would add another use case: in a machine reasoning-supported vision of EarthMAP, **computers would visit the model catalog**, to see what models exist for a particular phenomenon (for which they cannot directly find adequate observational data), to rank models for the same phenomenon based on suitability to the problem at hand (i.e., scale and context), extract the algorithms, pair them with data, run the models, and link to other needed models."

4. About the old Water Mission Area software page:
 - a. Keeping the old list of Water Mission Area models up to date was a time-consuming effort. People were assigned to try to keep things up to date, if they knew someone had retired, etc.
 - b. The audience of the old page was, first, internally, Water Science Center staff – that is the target audience internally. Externally, there are lot of different audiences – scientists/engineers/students/grad students; water resource managers at Federal level / state level, state departments using the models. The “general public” was less of an audience.

Participants and Further Suggested Participants

Link to [updated O365 spreadsheet of participants](#).

Participants and suggested participants as of 15 May 2020.

	Status	Name	Mission Area	Descriptor	Org
1	complete	Bagstad, Ken	Ecosystems	research economist, ecosystem services	Geosciences and Environmental Change
2	complete	Barlow, Paul	Water Resources	hydrologist, previously groundwater	Water Resources MA
3	complete	Bradford, John	Ecosystems	research ecologist, ecosystems, vegetation	Southwest Biological SC
4	complete	Cerovski-Darriau, Corina	Natural Hazards	landslides	Landslide Hazards Program
5	complete	Driscoll, Jessica	Water Resources	national water modeling	Water Resources MA
6	complete	Guterriez, Ben	Natural Hazards	geologist, coastal	Woods Hole SC
7	complete	Hinck, Jo Ellen	Environmental Health	contaminant response	Columbia Environmental RC
8	complete	McKerrow, Alexa	Core Science Systems	species distribution	
9	complete	Morgan, Leah	Energy and Minerals	research geologist, geochronology	Geology, Geophysics, and Geochemistry
10	complete	Peters, Jeff	Land Resources	spatial modeling	Western Geographic SC
11	complete	Rosemartin, Alyssa	Ecosystems partner	phenology, ecosystems	National Phenology Network
12	complete	Signell, Rich	Natural Hazards	coastal/marine	Woods Hole SC
13	complete	Sohl, Terry	Land Resources	land resources	EROS
14	complete	Slaughter, Stephen	Natural Hazards	Landslides	Landslide Hazards Program
15	complete	Wald, David	Natural Hazards	earthquakes	National Earthquake Information Center
16	complete	Zwart, Jacob	Water Resources	process-guided, lake temperature	Data Science Branch Water MA
17	awaiting response	Glynn, Pierre	Water Resources		
18	suggested	Farmer, Will		statistical models in water	
19	suggested	Fienen, Mike			

20	suggested	Jacobson, Pete			
21	suggested	Kreitler, Jason		carbon modeling	
22	suggested	Letcher, Ben			
23	suggested	Liu, Jinxun		ecosystem modeling	
24	suggested	Michael, Andy	Natural Hazards	earthquakes	Earthquake Science Center
25	suggested	Reilly, Tim	Environmental Health		
26	suggested	Rowell, Gareth			
27	suggested	Schuster, Rudy		economic models	
28	suggested	Skalak, Katie		IWP Integrated Water Prediction	
29	suggested	Sleeter, Ben			
30	suggested	Steevens, Jeff	Environmental Health		
31	suggested	Wiltermuth, Mark	Ecosystems		
32	suggested others	Bamzai, Aparna		climate adaptation	
33	suggested others	Ludwig, Kris	Natural Hazards		
34	suggested others	Walker, Jess	Land Resources		
35	suggested others	Wieferich, Daniel	Core Science Systems		

Responses to Questions

What are your expectations/criteria when you hear “USGS model catalog”?

At it’s most basic...a list of USGS models and their basic capacities, preferably characterized by main model function.

Ideally, it would also include:

- Dependencies / data inputs
- Resolution
- Output (information, format, etc.)
- Perceived gaps/needs for improvement
- Source code/structure
- Availability / access
- PI / Contact info?

A website where you can find (and search for) different models, perhaps those largely those created by USGS scientists?

Model name, link to software download and/or development (e.g. Github repo), brief description of the model, applications of the model (published or not), include versions of the model, link to publications of the model, model input/output, model resolution (temporal / spatial), interdependencies

An organized list or database of models developed at USGS to support science. It would include links to the model (code, software, etc.), a description and any development documentation, and science publications using the model.

Something very similar to sciencebase for data products, for models, with description, source/citation, example outputs/uses. Some keywords to help you find them.

Search by category, application, type (numerical vs other),

It would inform the audience about models that can be implemented for other research efforts – i.e. a hydrologic model that could be customized/trained for a new region.

A resource for USGS scientists, the scientific community, and outside users of USGS science.

Want to be able to filter on: How many people contributed? How recently have their been contributions?

To be most useful, should it be “Models used by people at USGS”?

See below.

A place to find information about the very diverse models we use in our many mission areas. My inner cynic thinks “a repository that can quickly become dusty if we don’t think of a good way to keep it current, i.e., linked to ScienceBase” My EarthMAP vision thinks of “a place for computers to search for models to couple together, so humans don’t have to” (more on that in question 7).

Working with WMA models such as the National Hydrologic Model, and would like to see different configurations of models – where work has been done already. Searchable by geography, topical interest (water balance, surface water, ground water, statistical, deterministic), temporal and spatial extent. “Zappos for models” (groundwater, SouthWest, time range, see what is available.) What has been applied, published model archive – input/output.

Help USGS connect with other DOI partners and beyond to correct tools. Demonstrate USGS’s work. Have a common place to pull models and tools from. Connection between model expert and a person with questions – a direct link to the expertise. Discovery of existing relevant models.

What are the major, established models in USGS that come to your mind first?

MODFLOW / GSFLOW – groundwater/surface water flow

FORE-SCE – Land use change

SPARROW – Water Quality

PRMS – precip/runoff

LUCAS (Land Use Change Analysis System)

CBLCM (Chesapeake Bay Land Change Model)

nothing really. Perhaps some geologic maps (and/or cross sections) can be considered models?

SPARROW, MODFLOW, PRMS, StreamMetabolizer, WRTDS, GFLOW, GSFLOW

Pedestrian Evacuation Analysis Tool (PEAT), Land Use and Carbon Scenario Simulator (LUCAS), CosMos, DCLAW, LaharZ

sea-level rise, flooding, species niche models

I'm not in the water mission area, but my impression is that there are some established models used there. Also, I assume there are many established models in areas like earthquakes, landslides, etc.

I'm in ecosystems, and I'm not sure if we have any models with sufficient recognized importance that they are supported by long-term USGS programmatic funds.

Water Program's Stream temperature models and hydrological models, fore-sce,

MODFLOW, NAQWA, COAWST, CosMOS, and landslide or earthquake models that are used for hazards assessments. May also apply to the USGS Coastal Vulnerability Index (USGS-CVI) which is a hazard index that is widely used.

- COAWST – two way coupled <https://www.usgs.gov/software/coupled-ocean-atmosphere-wave-sediment-transport-coawst-modeling-system>

- COSMOS (collection of different community models) – Delft family of models, ADCIRC

- National Probabilistic Seismic Hazard Model (probabilistic seismic hazard assessment) (empirical) (Ground motion prediction equations (under the National Map), and all submodels (e.g., fault models, earth models, shaking models) Petersen et al., 2020.

- PAGER – (many submodels) looking at losses from earthquakes (empirical)

- 3D Waveform Simulations – earthquake rupture simulated and seismograms in different materials generated – many codes do the same thing (numerical / physics-based)

- ShakeMap – could be under PAGER or standalone. (empirical)

- Aftershock forecasting

- Earthquake monitoring – early warning, magnitude/location – forecast tools.

- Note this is not comprehensive yet!

I'm most familiar with land, water, and ecosystem service-based models like the National Water Model, SPARROW, MODFLOW, LUCAS, FORE-SCE, GAP models, SolVES. There are also of course many models used by USGS researchers that did not originate with USGS, ranging from coastal processes (e.g., SLOSH), land cover change (CLUE), ecosystem services (InVEST), hydrology (SWAT, VIC, many, many others), economic models (IMPLAN, CGE models, etc.). USGS obviously also contributes to gobs of R packages, models built withing GIS environments, etc.

That raises the question of how we would distinguish what is a "USGS model"? Is it something developed at its core by USGS scientists (or whose development was funded by USGS)? Developed with just one or a few USGS collaborators? That was developed at a university or another agency but that has been adopted wholesale by a group of USGS researchers? It also raises the question of the function of the USGS model catalog: if just to document USGS models, it will NOT be a place researchers go to discover and compare different hydrologic models – that would necessitate wider intercomparison than just across USGS models!

PRMS, (NHM is a configuration of PRMS), MODFLOW, GS-FLOW, PHREEQC, Monthly Water Balance Model, SPARROW

We use: water, geospatial, contaminant spread. Working with UAS data. ; It's not an official "contaminants model." Contaminant, DOI resource, habitat. Depends on the geographical area/climate. Affects on animals. Transport models from WMA, StreamStats (McCarthy) , Faith Fitzpatrick – Oil aggregates. But need to know that these models exist. Link to Economic models (Rudy Schuster's group) is important and needed. T&E (Threatened and Endangered species) – in Ecosystems (distribution modeling). NPN. Categories for filtering – hydrologic model, geospatial, plant vs. animal based. Geographic applicability* (e.g., sagebrush), or nation-wide.

In addition to the models on the Landslide Hazards Program Software page, LAHARZ and VS2D

Where are existing locations to find the various data/code pieces for USGS models?

Specific web addresses...unsure, but know that there are models available through public USGS websites (e.g., NAWQA pages have access to a number of hydrology related models
- Also GITHUB/GITLAB for public access (not sure how much USGS models are distributed here)
- ScienceBase

ScienceBase?

Github, Gitlab, ScienceBase, NWIS, WQP, USGS National Map, Yeti,

I don't of a one good location. During the Drupal transition it was mentioned that our Drupal sites have a "Tool" category but no Model category. Is a tool a model and vice versa.

I'm not sure where I would look for info or code on USGS models, some is in ScienceBase (and GitHub?), I'm sure. Maybe through the pub warehouse

We manage and provide code for our models (which I'm not claiming are "USGS models") on GitHub.

I think there have been various local installations of code – USGS GitHub, and I think SAS had some local versions

Model specific websites that are maintained by Pis/USGS publications as resources for the various models that USGS has produced or produced with collaborators.

ScienceBase entries for COAWST and COSMOS.
GitLab

<https://earthquake.usgs.gov/research/software/> (old)
<https://www.usgs.gov/natural-hazards/earthquake-hazards/software>
(but there are both codes and models) Probably only half of those are important enough for the use case of building EarthMAP.

Either searching on USGS websites or journal articles, then looking for associated code repositories or executable files to install and run a modeling package.

There's a project model catalog – mix of URLs, code.usgs.gov and GitHub mix, local files. Software releases do not have all of the research code. And code.usgs.gov links may be private (Jess will send).

<https://www.usgs.gov/centers/cerc/science/natural-resource-damage-assessment-and-restoration-nrdar> (look at categories here). To create this list, Jo reached out to different Mission Areas, Susanna Soileau (Ecosystems Comm Person), but also had to rely on institutional knowledge to search for people/websites/publications. Contacts in the Econ field.

<https://water.usgs.gov/software/>

Water Mission Area Model Archives (requirement since ~2016) are all on the "Water Node" AKA NATWEB. They are in .zips.

Brainstorm and crowdsource a list of "_____ models" (what model types do you think of? No overthinking. Just how would you complete the phrase "_____ models".)

age (or age-depth) models (Bacon: <https://chrono.qub.ac.uk/blaauw/bacon.html>), thermal models (Hefty, QTQt), diffusion models, geologic maps as models?: <https://ngmdb.usgs.gov/www-nadm/>, thermodynamic modeling (<http://melts.ofm-research.org/>)

Land cover / land use model

Urbanization Model

Climate Model

Water Quality

Ground Water

Bioclimatic Envelope

Species Distribution Model

Vegetation Succession

Fire / Disturbance / Recovery Model

Post-fire Debris-flow Model

Coastal Marsh Migration Model

Carbon and GHG fluxes (landscape)

Sediment Transport

Aerosols

pheno-climatic models

climate change projection models

data model

conceptual model

coupled socio-ecological models

All mine would be spatial: GIS, Climate, Flow, Evacuation models

process-based models, deep learning models, data intensive models, machine learning models, water quality models, 1D models, coupled models, climate models, statistical models, empirical models,

(Computational), statistical, process-based, deterministic (HPC – solving equations of motion – solving physics equations on grid systems, MPI-using)

ocean model

tectonics model

hydrodynamic

As an interdisciplinary researcher, I think of specific disciplines – hydrologic models, ecological models, land-use change models, coastal process models. Perhaps that's because integrated modeling/model coupling across disciplines is so unwieldy, it's easiest to think of a particular discipline and how I might find the pieces to couple to answer my interdisciplinary research question.

Vegetation dynamics, fire behavior, land-,use dynamics, smoke modeling, erosion models, flood models, hydrologic models, climate models, statistical models – (i.e. occupancy – Presence, metapopulation)

I think pretty broadly in terms of models that USGS is known for such as MODFLOW, NAQWA, COAWST, CosMOS, and landslide or earthquake models that are used for hazards assessments. May also apply to the USGS Coastal Vulnerability Index (USGS-CVI) which is a hazard index that is widely used.

Vegetation dynamic
species distribution
urbanization model
socioeconomic models

Sediment transport
Sea level rise

Sharing an INITIAL list of models, what else belongs on the list? What other initial reactions do you have? (se...

Other than what was discussed on the call...I'd also note that there are a number of models used within the USGS that may not have their origin within USGS itself. Just a couple of Examples...Adam Terando and the use of SLEUTH, a very well known model of urbanization that developed at UC Berkley and is now used by people across the globe. Another example is LANDIS...a forest landscape model that has a user community and is used by people within USGS as well.

Also note in Land Resources (and likely elsewhere) there are a number of land-use related models not on this list, models that have a long history of use within USGS. FORE-SCE, CBLCM, and LUCAS are the ones that come to mind.

Makes sense, is roughly what I was expecting. Makes me wonder, if all models should have a cool acronym name.

The list should include more info about each model, organized in a table. So the user can filter on different model categories or keywords, authors etc.

Awesome work!

Do all the models have to be named? How do you fit in machine learning models which don't have a specific name other than the broad tool (e.g. Recurrent Neural Network – used in water resources for temperature prediction)

Random Forest Model, etc..

I think it would be useful if the model catalog is organized by higher level topics (e.g. Hydrology, Biogeochemistry, etc..)

Would conceptual models fit in here? E.g. River continuum concept

I would think that process models are more relevant than statistical/mathematical. I would focus on models that are used consistently to update USGS data. I think the individual models (child items) should have descriptive names and/or organized by discipline. If there are linked models they can be linked using tags.

Is it possible to categorize the model list by application or discipline?

I think clarifying what you mean by “model” will be crucial...there are MANY statistical models that hypothetically could be included, so avoiding these will be very helpful.

Also, our models (documented here: <https://github.com/DrylandEcology>) didn't appear.

- “unambiguously link” gives a chuckle
- “everything associated with them” also gives a chuckle

The description is very comprehensive and may not be a reasonable aspiration (to have everything!) Would rather start top down with the main goals (is it populating EarthMAP or making things discoverable) (because there may even be things not worth discovering)

It's unsurprisingly incomplete, though a good start for being a few weeks. It also includes models to my knowledge developed outside USGS (Delft3D), which makes me wonder about how we account for/include “outside” models. I would love to see a more systematic survey of USGS so that the content isn't biased by who is and isn't asked. Collecting the information at a Science Center level would seem to me like the easiest way to do so. That would obviously require clear guidelines about what you define as a model, what's considered internal vs. external to USGS, etc.

The way the catalog is now – you still need to have some basic knowledge of what these acronyms stand for. More context in the quick view. Descriptors may add assumptions, but Mission Areas would help to categorize.

Bonus question: How do you define “model”?

a way of describing a state or process

At its simplest...a conceptual (and often quantified) representation of a process.

Any representation of reality that helps us better understand or predict future conditions. For the purposes of this catalog, I could see limiting it to numerical models, and then categories within, as you said, spatially-explicit etc.

Because of my background I think of models as simulations of something, like people evacuating an area or debris flowing down a slope.

A tool used to organize, understand, and predict the world around us.

There are conceptual, process and statistical models. For this catalog it seems the focus might be on process models.

Model can be considered as a representation of a system via physics, mathematics, empirical, that can be used to approximate a system in question and used to conduct research or apply research to a management application.

In this context, I believe “model” should be a process-based, mechanistic simulation model that has been supported by USGS. The process being simulated can vary among mission areas.

Also, possibly this catalog could include established complex workflows for statistical analysis...like SDM generation from BBS data.

Something that represents our best understanding through equations or reproducible computations. Not just qualitatively talking about how things work.

Noun or verb?

SLAB 2.0 uses model as a noun – it is a product that represents the earth’s structure, static, no temporal aspect.

Many times we mean codes that look at physical processes. (SLAB 2.0 could fall under PSHA – need to know all faults, and SLAB represents faults of subduction zones.)

A mathematical approach to quantify some phenomenon in the world for which we cannot directly gather complete data across all locations and times needed.

For water: Mass balance water. Representation of physical or chemical processes.

SPARROW is more an application of statistical methods.

There could be a lot of use for the statistical methods, but not independent of the physical processes. Statistical models could guide process-based models. (Such as process-guided statistical models.)

set of mathematical equations that describe some aspect of a hydrologic system, state, stresses, boundary condition.

“Anything else you would like to share. If you’d like to share your name, program, etc. you can do so here. Other topics of interest – how about models that are heavily used by USGS but do not ori...”

Yes, models from outside the USGS but are used within it are VERY common.

TO make this catalogue useful, make sure it’s more than a list. Characterize dependencies, linkages, needs/gaps, etc. as well as model function and output, so that it provides a means of potentially finding and filling gaps/needs among models. In short, make sure it can be used as a tool to help improve model integration across USGS.

I’m Alyssa, with National Phenology Network

I do think that thinking more porously than just within USGS would be good, with some limit of course – models that were created by USGS scientists, with USGS funding, or models that underlie USGS products.

I often need to search for models of growing degree thresholds for species activity, but I wouldn’t care the source, would just want a general model database.

I don’t know many USGS scientist who use model developed out side of USGS but I can think of one. I think it would be valuable to capture these models as well, but maybe in a different list or catalog to separate them form what’s developed at USGS or with partners. <http://www.apexrms.com/state-and-transition-simulation-models/>

Jacob Zwart, USGS WMA Data Science Branch

We often use models that do not originate from the USGS but I don’t think this should go in the USGS model catalog.

I think this would be most useful for model interoperability and model coupling – trying to find a model outside of your domain is difficult and this model catalog would help facilitate collaborations for model coupling projects.

If I was starting on a research project and knew I could customize an existing model for that work I would go to the catalog to get the latest. Which brings up some need for maintaining the linkages to the latest – circling back to the person responsible for the model.

Ben Gutierrez- USGS Coastal and Marine

Who needs this catalogue? USGS scientists, outside collaborators both scientists and managers, so that they have a resource to identify USGS products and investigators that may apply to the problems that they are trying to address.

The catalog could be a resource for researchers outside (or inside) of USGS looking to apply established analytical tools developed by USGS.

If looking for integration of model pieces, consider community-supported best known models, not just USGS models.

Not sure if/how useful to me.

There are 100s to 1000s of supporting functions to large models. But there are diminishing returns to populating everything under that.

Categories: numerical vs. empirical

How to categorize models at the very highest level? Are we forecasting? Trying to understand processes? Sensitivity analysis? How about basic science – that may be behind the curtain. Not the priorities for contributing to EarthMAP.

Start with: what are we trying to model, instead of just a listing of a bunch of things that might not go anywhere.

Non-USGS code would be significant. Work with FEMA to do estimates of losses (used for residential/state declarations), HAZUS (Hazards US) from FEMA allows loss model. If we want comprehensive earth model that includes loss estimates for EQ (as EarthMAP should), this collaboration is key. Important part of forecasting capability.

Ken Bagstad, Land Change Science program (happy to talk more Leslie to clarify any answers – this is a very interesting effort)

See my comments in question 2 re: models used by but not originating in USGS. I think this is a definitional problem that will need to be addressed early on. These are extremely common in my field.

One other point: USEPA a number of years back put together an ecosystem services model library. It may be instructive to read more about their goals, process, and how they achieved this:

<https://www.epa.gov/eco-research/ecoservice-models-library>. Having built it, it's now starting to gather dust because there's no update mechanism aside from authors voluntarily filling out a form for each of their studies that develops a new model (you can imagine how often that happens). This is a big advantage of working with an organization like USGS: if our data, code, and model repositories are properly linked this should be able to update as USGS releases new models.

Who would visit the catalog: the 3 use cases you describe here

<https://my.usgs.gov/confluence/display/cdi/Model+Catalog+Working+Group> in “Who needs this catalog, and what is it going to do for them?” seem like you'll get an incomplete view of a discipline's models if you only include USGS-originated models (maybe this is less the case for subdisciplines where USGS is a more undisputed global leader, like in earthquake and volcano hazards).

I would add another use case: in a machine reasoning-supported vision of EarthMAP, computers would visit the model catalog, to see what models exist for a particular phenomenon (for which they cannot directly find adequate observational data), to rank models for the same phenomenon based on suitability to the problem at hand (i.e., scale and context), extract the algorithms, pair them with data, run the models, and link to other needed models. This is of course requires semantics and perhaps a different view of the model catalog, but is the vision we're working toward through <http://www.integratedmodelling.org/>. It also raises the question of where you put model coupling or integration frameworks like ARIES, CSDMS, etc.

Issue: Getting models into publishable format. Access to certain time steps or variables. (need standardization of outputs). Publications may only report dozens out of hundreds of variables. Want to avoid re-running model.

Purposes that the model has been used for. If you want to choose a model, what has been used before for particular circumstances.

Lots of collaborations with other agencies / academia and contribute to those models.

Which models could be integrated with other models? Clearly communicate what integrations are happening/are important.

For the use case that we discussed here today, we'd want the other author-organizations to put their models in the catalog.

Model intercomparison and evaluation – ways to find the “best” model for your purpose.

As for supporting use of the model – this will vary among models, may not be able to support.

Importance of some sort of standards body – accessible, some sort of standard, etc.

Keeping the model catalog up to date – (has seen that CSDMS has had issues with this, contact person, etc.). It is difficult. Is there a partially automated way to check this? (maybe an expiration date in the catalog).

There is already the Water model archive requirement – can this catalog help with that requirement?

There is a single person that approves every model archive in water (Steve Markstrom – may change). No uniform solution. There's a PDF form – send folders of files. (very hard for NHM)

Will Farmer (statistical models), Katie Skalak (IWP), Mike Fienen, Jess Driscoll.

Context: Science coordinator for NRDAR. Chemical or oil spills. Protect DOI resources. Expertise on science side/models to the situation. Data resources URL shared, pulled together for NRDRE practitioners. Also inland oil spills group. **Connect science to the events. Could use catalog to inform our partners** (for example coast guard). Disaster supplemental funding – help tell the accomplishments. Show Congress what work was done, how it prepares us for the future, and future opportunities. ; Contaminant issues with cascading hazards. ; Pipeline collapse – has been contacted by State Dept about event in Ecuador. Contaminants, geomorphology, multi-disciplinary, but no catalog to pull stuff out of. There's a progression of Incident, settlement, restoration (different set of models). Challenging for research scientists to get to contribute sometimes; hydrogeophysics branch has a ton of different models (Woods Hole). How to navigate the current list of resources. Is the purpose more internal or external, need to know the target audience. “How does it help the public?” Non-USGS models are less important in this case.

Other Notes

13 April

- The catalog project should be able to identify gaps or perceived gaps and address them

- Internal vs. external models (land resources may have a lot of models that are more for internal use and have not been highly published on – should these be represented?)
- Machine learning – deep learning, recurrent neural network – how to be represented here, For example, deep learning model as applied to another model – would that be represented in the catalog?

21 April 2020

- Several participants thought primarily of process-based models – not statistical models. There may be “too many” statistical models to comprehensively catalog.
- Maybe models with USGS institutional support should be prioritized. Some models are supported completely by outside-of-USGS funds. Some USGS models rely on external models – where is the line drawn?
- In participants’ disciplines, concentrating only on models with institutional support would leave out a lot of discipline-specific models that are presented in papers, may be in development – what is that threshold?
- Is there the expectation that this is a USGS-vetted model and we believe in it and will provide support on running the model?
- How to keep this up to date? Contact people change. Or the PI could update the model.
- How to search –
 - Helpful to expand out the acronym
 - Sort/filter by discipline, more important than by mission area
- Try to include as many things as possible to help someone get started.
- Might want to include model name – brief description of what it does – simulates this, inputs and outputs, timestep. Help people to look at it and get a feel for what is out there, without going to different sites.

Additional notes from other conversations exist if more context or detail is needed. Contact Leslie Hsu at lhsu@usgs.gov.

References and Resources

This is a list of selected resources that would be useful for understanding some of the models and other things mentioned in this summary document.

1. **USGS Water Science Center Hydrologic-Modeling Survey Fiscal Year 2019**, January 2020
Summary Findings (emailed by Paul Barlow)
2. Petersen, M.D and others, 2019, **The 2018 update of the US National Seismic Hazard Model: Overview of model and implications** (description of models making up the US National Seismic Hazard Model)
3. **Ecosystem services model library** (from Ken Bagstad): <https://www.epa.gov/eco-research/ecoservice-models-library> “A searchable database of ecological models for estimating the production of ecosystem goods and services”
4. **Water Mission Area archive policy**: <https://water.usgs.gov/ogw/policy/gw-model/>

Model Catalog Working Group

Description: A short-term working group to inform a USGS Model Catalog effort in FY 2020.

Point Of Contact: Leslie Hsu (lhsu@usgs.gov)

Meetings: TBD - one to three meetings in April 2020

- Overall objective
- Why are we building a USGS Model Catalog?
- Who needs this catalog, and what is it going to do for them?
- What do we know already?
- How can I contribute?

Overall objective

The overall objective of this working group is to get diverse perspectives of what potential users expect to see from a USGS Model Catalog.

We are aware that "model" means different things to different people, so we want to be inclusive of the many different users that might visit such a model catalog.

Why are we building a USGS Model Catalog?

The model catalog will support the USGS Director's vision of modernized, 21st century USGS Science, as detailed in his [USGS-wide blog posts](#). This vision includes the EarthMAP concept, of integrated predictive science capacity.

A USGS-wide model catalog has also been an idea that has been proposed through the CDI previously, for example at the [2017 CDI Workshop and recommendations from the Roadmap to Integrated Science](#) discussions.

"there is a somewhat logical pathway to this from the other things that USGS has done in an attempt to catalog/inventory its wares. We have our Publications Warehouse, our Science Data Catalog, and a code.usgs.gov - all enterprise-scale information systems about certain types of digital artifacts produced by our work. Statistical/mathematical models are another thing we do as a core part of USGS science, and they happen to be one of the types of products we build that has touch points on all of those other catalogs. They are important building blocks for forecasts, which gets into the integrated predictive and actionable intelligence parts of the developing EarthMAP concept, and so there is a somewhat logical connection to that initiative and this particular time period in our history." (Sky Bristol, member of the EarthMAP project team)

Who needs this catalog, and what is it going to do for them?

This is one of the questions that the working group may address further.

But some examples are:

- Researchers who want to find existing models for a specific topic to learn from or link to
- Early career researchers who want to see what models exist in their discipline
- Members of large, integrative projects who may want to discover opportunities for linking different process models

There are surely many more cases not yet listed here.

What do we know already?

Different parts of the USGS have already compiled lists of their models.

For example, the water mission area has a listing of their models.

USGS models are sometimes captured in other systems such as the [Ontosoft USGS listing](#) and the [CSDMS model repository](#).

There are many USGS Digital Object Identifiers that have been tagged as related to models.

ScienceBase hosts many models.

Pubs Warehouse has many publications that describe models for all sorts of USGS science.

There are some "well known" USGS models with name recognition that are displayed/communicated in different ways: [MODFLOW](#) (web page) , [PROSPER](#) (journal publication), [SLAB2](#) (ScienceBase item).

One thing we know for sure is that different people have different ideas of what should be in a model catalog, so we must engage a working group to get more viewpoints.

How can I contribute?

We will be convening sessions to listen to your views on a USGS Model Catalog in Spring 2020.

We want to hear views on

- What are expectations/criteria for a USGS model catalog
- What are major, established models in USGS
- Where are general locations to find the various data/code pieces for USGS models

We will be contacting an initial group of CDI members and asking for recommendations of other participants.

You can also send a note to Leslie Hsu at lhsu@usgs.gov to let her know you'd like to contribute.