When Aggregated Data Means Opportunity in the FIFRA/ESA Emerging Species Assessment and Consultation Process: Definition of Species Range

Authors: Bernalyn McGaughey[1], Perry Beale[2], Richard A. Rupp[3], Tilghman Hall[4]
[1]Compliance Services International, Lakewood, WA; [2]Washington State Department of Agriculture, Yakima, WA; [3]Department of Crop and Soil Sciences, Washington State University, Pullman, WA; [4]Bayer CropScience, Research Triangle Park, NC

ABSTRACT

Since 1997, the FIFRA Endangered Species Task Force (FESTF) has been researching methods to aggregate and collect data that will assist EPA's Office of Pesticide Programs (EPA-OPP) in the conduct of FIFRA risk assessments on endangered species, and which would be acceptable to either the USDI Fish and Wildlife Service or USDC National Marine Fisheries Service (collectively, the Services) when such risk assessments become an element in the Endangered Species Act consultation process. While FESTF has gathered a tremendous amount of data, and has developed methods for their application, quality assurance, management and updating, a clear regulatory pathway to provide an outlet for use of these resources is only now beginning to unfold under guidance given by the National Academy of Sciences (NAS) Panel on FIFRA and Endangered Species Act (ESA) risk assessment. Recently, EPA-OPP and the Services discussed several aspects of their emerging assessment plans, including the use of HUC 2 units as representative assessment areas, the grouping of crops for definition of the consultation action area, and use of FESTF aggregated species location data to define species national pesticide consultation "range." FESTF discusses in this presentation the challenge of applying information to an assessment when such information varies based on the species of interest, the data believed to be authoritative, and the scale of spatial data available. We explain how such data were handled in the recent development of a national level "species location" dataset suitable for pesticide assessment. FESTF will explore how such data change over time, the improvements and refinements that may come for these data over time, and how states might help shape data collection and use in the future.

CHALLENGES IN APPLYING SPATIAL DATA TO NATIONAL LEVEL PESTICIDE RISK ASSESSMENT

One of the challenges in FIFRA/ESA national-level pesticide risk assessment is to manage complexity, and deal with data that are constantly changing over time. This forces the risk assessment process towards increased use of authoritative spatial data and representative models. As modeling science, imagery and geographic information system (GIS) technologies advance, data quality improves and data abundance multiplies seemingly logarithmically. The national scale of a pesticide registration action by EPA-OPP requires that somehow a reasonable snapshot be composed to offer a nationally representative dataset that can be used in risk assessment given the time and resources available to do so. To that end, the agencies have proposed the use of certain aggregated data sets:

- Representative geographic assessment units based on HUC 2 hydrologic units: There are 21 delineated HUC 2 hydrologic units in the United States (**Figure 1**). For most of these HUC 2 units, a tremendous degree of variability exists in both climate and cropping. For example, the Pacific Northwest HUC 2 (**Figure 2**) embraces sea level to alpine ecosystems, temperate rain forest and arid desert lands, minor crops such as mint and cranberries, and major crops such as wheat. According to an EPA-OPP summary there are 21 national HUC 2 units within which are 221 subregions and 2,264 HUC 8 watershed units.[1] Nationally, the average size of a HUC 8 watershed unit is 1,700 square miles.[2]

- Habitat "bins" for aquatic species assessments: EPA-OPP has proposed 9 generic aquatic habitats (3 static water, 3 flowing water, and 3 marine) that will be used in aquatic exposure modeling. EPA-OPP also included another aquatic habitat bin, "aquatic-associated terrestrial habitats." That habitat bin is not parameterized for receiving water modeling inputs (i.e., no depth, width, length, or flow are specified) and it is unclear how this bin will be applied to assessment of aquatic species. At any rate, fully aquatic listed species can be sorted into these bins by using what is known about their habitat characteristics.

- Land cover data to represent pesticide use sites: EPA-OPP and the Services plan to represent pesticide use sites by using either the USGS National Land Cover Database (NLCD) or the USDA NASS Cropland Data Layer (CDL), with the dataset selection based on the nature of the use site. The CDL data will be aggregated over a 10-year period (when 10 years of data become available – currently fewer years are available and compatible across resolutions and image types), which introduces additional challenges related to interpretation of a dataset combined in this manner.

- FESTF species location data, vetted through FWS field offices: The outlet for use of FESTF resources is only now beginning to unfold under guidance given by the NAS Panel on FIFRA and ESA risk assessment.[3] Following the completion of the NAS Panel's report, EPA-OPP, the Services, and the USDA developed an interim plan for implementation of the Panel's recommendations.[4] The interim plan's summary of follow-up tasks included this action related to FESTF data:

  Share information about the FESTF database and ECOS (IPaC, Critical Habitat Portal), and discuss whether/how these tools can be used as part of the interim approach to identify species and define species' ranges and critical habitats[5]

  Initially, FESTF shared information on the types of data available with EPA-OPP and the Services. Shortly thereafter, EPA-OPP requested that FESTF demonstrate the spatial extent of species location data by comparing FESTF's aggregated sources with information available from the Services, and the provision and review of these data are underway.

Figure 1. HUC 2 Hydrologic Units of the United States. Currently, EPA and the Services plan to use these units as the basis for their national-level endangered species pesticide assessment and consultation process.

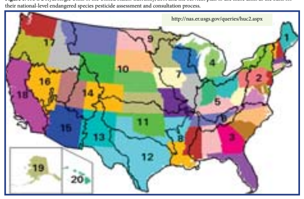


Figure 2. Example of a HUC 2 hydrologic unit, overlaid with USDA Cropland Data Layer (2014) land use categories and HUC 8 watershed units. In this instance, the HUC 2 area encompasses almost all of Oregon, Washington and Idaho, with some HUC 2 subunits extending into California, Nevada, Wyoming, Montana and a small corner of Utah.

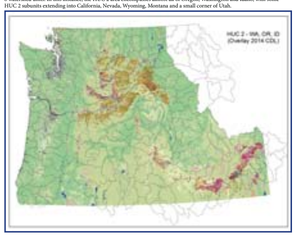
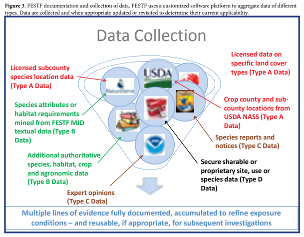


Figure 3. FESTF documentation and collection of data. FESTF uses a customized software platform to aggregate data of different types. Data are collected and when appropriate updated or revisited to determine their current applicability.



LESSONS LEARNED: FESTF EXPERIENCE IN COLLECTING AND AGGREGATING DATA

FESTF has learned what is necessary to create a usable, national set of information based on data from various sources which must be arrayed in a manner that (1) evens out data portrayal across the nation and (2) provides a uniform platform for data use amongst registrants and EPA-OPP. In response to FIFRA data requirements for the registration of crop protection and other pesticide products, FESTF has addressed the collection and management of data for endangered species risk assessment by establishing two data management systems to house and manage species location data and aggregated species and use site attribute data, and collected data through applied assessment activities and access to proprietary or licensed data. This concept of data collection, management and distribution arose from an analytical and regulatory-responsive process that began nearly 20 years ago with input from EPA-OPP, the Services and industry. The documentation and aggregation of data is illustrated in **Figure 3**. Along the way, FESTF developed methods and concepts to address some of the challenges discussed above. Data collected for species assessment can be tabular, narrative or spatially based. As such, there are different ways to aggregate it. Additionally, there are differing ways in which it must be managed. Loosely, data fall into four broad categories:

- Type A Data (Figure 2, red font): These data largely are georeferenced, spatial datasets that can be manipulated in a GIS environment to the degree their resolutions allows. Often, the lowest common denominator for all datasets, when combined, is the county unit, primarily because some types of data are only available at the county level. This is true for location data for many species and for some land cover types, such as certain minor crops for which USDA Census of Agriculture data are more specific or reliable than spatial datasets composed nationally from satellite imagery. Another element common between these datasets is that they are updated on a specific and predictable basis.

- Type B Data (Figure 2, green font): These data are largely textual or tabular, and are collected in support of assumptions used in the national level risk assessment process. Once collected they typically remain useful over time. Unlike Type A data, above, Type B data usually don't require updating, and in any case, are not maintained by underlying sources in a way that allows a predictable or comprehensive update interval. These data provide information important to consider in the risk assessment process such as species attributes related to diet, body weight, migration patterns, and breeding/reproductive cycles, as well as crop attributes related to planting dates, typical blooming and harvest times, and soil or climate requirements. Some of these data may possibly be portrayed spatially, but they most typically are not delivered or reported in that manner.

- Type C Data (Figure 2, brown font): These data are typically point-in-time publications or opinions that yield important information but do not necessarily provide data that can be used repeatedly without validation of their continued usefulness. Thus, once collected, this kind of information needs to be reexamined to determine its relevancy to current conditions and the full body of information that may have grown or changed over time. Such data include a species expert's opinion of the presence of a given species in areas where a certain crop will be grown or treated with a given pesticide when this opinion is given in void of specific species location data, or conclusions in a listing notice that may be modified or enhanced as more is learned about a listed species over time.

- Type D Data (Figure 2, black font): These data are usually highly specific to evaluations conducted for a specific purpose, primarily related to the assessment conducted for a registration of a particular proposed pesticide use. These data are an outcome of using the other types of data discussed here, with refinements and/or additions to them. These data could include site-specific surveys, use-specific pesticide or species monitoring, or highly refined spatial data generated from customized image analysis or acquisition. Typically, only portions of data like these apply to subsequent assessments, and may not be applicable to a multi-species or nationwide setting.

FESTF

MEMBER COMPANIES

Albaugh, Inc	Dow AgroSciences, LLC	MacDermid Agricultural Solutions, Inc.	Nufarm Americas, Inc.
AMVAC Chemical Corp.	DuPont Crop Protection	Monsanto Co.	PBI/Gordon Corp.
BASF Corp.	FMC Corp., Ag. Products	Nichino America, Inc.	Syngenta Crop Protection, Inc.
Bayer CropScience	Gowan Company LLC	Nissan Chemical Industries, Ltd	Valent USA Corp.
Cheminova A/S	ISK Biosciences Corp	Nisso America, Inc.	

Over the years of dealing with national scale concepts for data management, and realizing the importance of local and point-specific data to the endangered species risk assessment, FESTF has developed ways to aggregate and share such details so that multiple lines of evidence can efficiently and reliably be applied to a regulatory assessment. The first example of how these efforts have been accepted in the agencies' emerging endangered species risk assessment process is in the development of species range maps. Additional FESTF data are available to yield support to this kind of collaborative effort.

COLLABORATING TO ESTABLISH "BEST AVAILABLE" NATIONAL-LEVEL SPECIES DISTRIBUTION MAPS FOR USE IN PESTICIDE RISK ASSESSMENT

One of the first steps in implementing the NAS Panel's findings has been to review species location data as aggregated by FESTF. To address the question of species locations, and after EPA-OPP, FWS, NMFS, USDA and FESTF met in data review sessions, the agencies decided to develop a national set of species "ranges" for species under FWS jurisdiction. All parties agreed to a "map kit" concept whereby FESTF would supply data from their aggregated species location database to assist in the determination of species range. All data sources have been through a thorough quality control and validation process prior to being incorporated into FESTF's aggregated dataset. As agreed to by EPA-OPP and FWS, the species maps are being delivered on a regional basis to FWS field offices that have been and will be revising the range data, or providing range data of their own, for the purposes of confirming the "best available" species location data for use in national level pesticide assessments (**see Figure 4**). The more refined the resulting species distribution data, the more accurate a species assessment becomes. This exercise sets precedent on how multiple stakeholders can work together efficiently to improve data applicability, availability and validation.

Phase 1 of this process is underway and includes Services review of aggregated species maps for 461 species. Phase 2 of the process has generated 641 additional species maps that will shortly undergo Services review. A third phase of work will complete maps for all listed species and deal with the unique situation that Hawaii and the Pacific Island Territories represent: hundreds of listed species on a small land mass (**Figure 5**), and regional offices that have very limited resources available. Over 500 maps will be generated in the third phase of this process.

Figure 4. The data sharing process. Through evaluation and discussion of FESTF's species location data, EPA-OPP and the Services decided that a collaborative process would leverage the work that FESTF had already completed in a way that delivers a validated national level data set on FWS-agreed species locations.

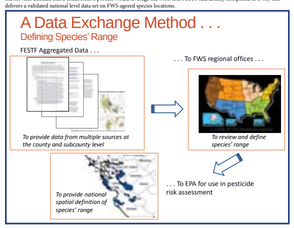
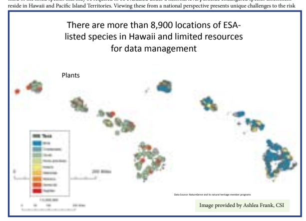


Figure 5. ESA-listed species (including those that are candidates or purposed for listing) of the Hawaiian Islands. Nearly one-third of the listed species that may be required to be evaluated under a national level pesticide endangered species assessment reside in Hawaii and Pacific Island Territories. Viewing these from a national perspective presents unique challenges to the risk



ADDRESSING FURTHER CHALLENGES IN SUPPORT OF NATIONAL LEVEL PESTICIDE RISK ASSESSMENT

Perhaps the most valuable lesson learned in the data collection, aggregation and vetting process is that there are other opportunities to improve national-level data which can be met by collaboration and the leveraging of available resources.

- Refining data within HUC 2 units: Again using the Pacific Northwest as an example, a cooperative exchange program used to "train" the CDL classification process has greatly upgraded the quality of data available for the HUC 2 (and in particular, Washington) area. This is illustrated by comparing a classification of crops in 2007, before state data was used to train national imagery (**Figure 6**) and a classification of crops in 2013, after processing was "trained" by using state data (**Figure 7**). Involving Oregon and Idaho under a USDA-led effort facilitated by FESTF base data for example, in the manner done for species data, could "lift" state data to a HUC 2 platform to provide enhanced yet consistent data within a given EPA-OPP risk assessment unit.

  Even with training of the USDA NASS CDL data by ground-truthed state agency data, differences in the data sets still occur (**Figures 8 and 9**). It is interesting to compare these depictions to the base USDA Farm Service Agency (FSA) and Washington State Department of Agriculture (WSDA) data (**Figures 10 and 11**). And, even upon only visual inspection, it is possible to see that original imagery contains much more information that is yielded by national datasets extracted from it (**Figure 11**). Future opportunities to work collaboratively between states and federal agencies could bring further enhancements to national-level datasets.

Figure 6. A section of Washington state crops depicted by USDA NASS CDL 2007 data. At this time, the CDL used only USDA Farm Service Administration (FSA) data. USDA FSA data includes all of the major program crops such as wheat and field corn and a fair amount of what are big ticket center pivot crops in Washington, such as potatoes, sweet corn, peas, onions, alfalfa, dry beans in the basin.

Figure 7. A section of Washington state crops (same section as Figure 6) depicted by USDA NASS CDL 2013 data. To better represent the many types of crops grown in the Pacific Northwest, the WSDA focus for ground-truthing to support training the USDA NASS CDL data is in the areas not covered by FSA: orchards, vineyards, specialty crops, and intensively farmed and irrigated basins such as the Yakima valley. This results in a perfect marriage of federal and state data sets to capture the majority of agriculture in this state. Washington agriculture is more diverse (300+ crops) than that of most states, excluding California, which makes Washington one of the "the worse case scenarios" for achieving as accurate as possible depiction of crops in the federal USDA aggregated data. The CDL may be inclusive enough in some states, while other states may be able to help improve CDL data in a similar but reduced effort from state partners.

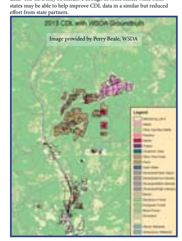


Figure 8. An example of the depiction of landcover using the National Land Cover Database (NLCD 2011) set and general land use categories depicted within it. The National Land Cover dataset is not as finely characterized by crop type as is CDL.

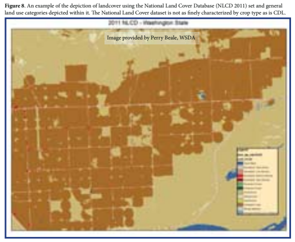


Figure 9. Depiction of Washington crops using USDA CDL 2011 data, as "trained" by WSDA sate-collected data. Even when trained by WSDA data, the national dataset is not as refined as the state dataset.

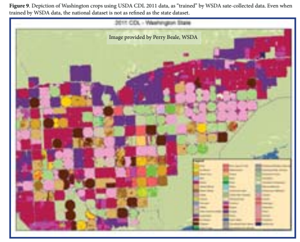


Figure 10. Crops as represented by the WSDA state-collected data. State data are ground truthed to the field level.

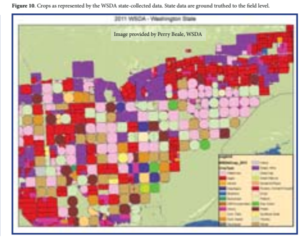
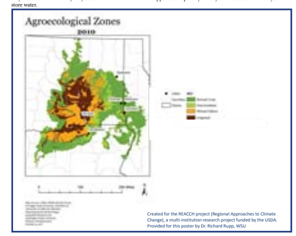


Figure 11. This section of a 2013 National Agricultural Imagery Program (NAIP) image of Benton County hints at the detailed data caught by aerial inventory of all US counties, conducted once every 1 to 3 years, depending on the state and funds available. NAIP imagery is acquired at a one-meter ground sample distance (GSD) with a horizontal accuracy that matches within six meters of photo-identifiable ground control points, which are used during image inspection. Most aggregated data sets that are used for national crop site depiction are aggregated at a 30-meter resolution. The WSDA geospatial crop data is aggregated at higher resolution, but for public release is portrayed at the township level.

- Adjusting portrayal of crop action areas (currently 11 national-level groups) as footprints:

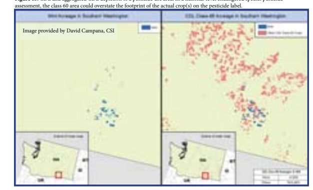
  EPA-OPP will be grouping crops together in an effort to bring efficiency to national level endangered species pesticide risk assessment, and currently plans to have 11 groups of crops as representative units for national assessments. For each of the 11 crop groups envisioned for national risk assessment use, the multi-year aggregation process in and of itself will introduce influence over the size and shape of the crop "action area." In the above discussion, we note that older USDA data, not having the benefit of more specific ground-truthing introduced by WSDA's effort, will negate the better quality data resulting from that exercise (**compare Figures 6 and 7**).

Figure 12. Agroecological zones in a section of eastern Washington developed by the REACCH project. This map shows what REACCH called "agroecological zones" as they existed in 2010. The annual crop zone can support a crop every year. The intermediate zone will incorporate a fallow year into the rotation to save up water during the fallow year, but rotation may not be to fallow on a yearly basis. The wheat-fallow zone supports a crop every other year and uses the fallow year to store water.



Another challenge faced when aggregating these data over time is the impact of trends that change from year to year as influenced by actions such as permanently changed land use, transitioning crop practices, and the adoption of new conservation set-asides in national or state programs. Multi-year aggregation of cropland datasets without some consideration of permanent or long-term changes will likely distort the action area in an unrealistic way. One way of dealing with such variables is to assign the aggregated data attributes that can be used to influence their behavior in the risk assessment process. That is, attributes can indicate how often crop rotation occurs, and under what circumstances it may change. A good example of how this can be done is land use characterization currently in use under the REACCH project (Regional Approaches to Climate Change), a multi-institution research project funded by the USDA. (**Figure 12**).

Figure 13. CDL and aggregated CDL depiction of a pesticide use area. In a national-level endangered species pesticide assessment, the class 60 area could overstate the footprint of the actual crop(s) on the pesticide label.



A final challenge will be maturing risk assessment crop groupings to account for minor crops and specialized pesticide uses. One northwest minor crop, mint, serves as a good illustration of this concept. The WSDA crop footprint for mint is quite limited (**Figure 9**). In the current EPA-OPP concept of using 11 crop groups for national-level assessment, mint is potentially included with other "Class 60" CDL crops (**Figure 13**) which could distort the shape and extent of the action area should mint be the primary use area.

- Keeping data current: FESTF has learned that "Type A" data described at left are important to keep current and, given a strong base platform, annual updating of such data is not overly burdensome. A single federal agency, however, may not have resources to do that. In the FESTF/Agency species map exchange process, FESTF is prepared to provide new maps to the agencies for those species locations that change with the annual updating process. Typically, only about 10% of the data change in a single update cycle. It may be helpful to establish this kind of an interaction process for additional datasets, such as for those species regulated by NMFS. For example, NMFS and EPA-OPP have agreed that the data available from the NMFS species website is authoritative for use in national-level pesticide risk assessment. However, the NMFS data are based on 2005 Washington Department of Fish and Wildlife (WDFW) data and NMFS has not yet been able to update their website with the 2011 WDFW data, which results in the "best available" data being gathered only if both sources are consulted (**Figures 14 and 15**).

Figure 14. Comparison of data from NOAA-NMFS and Washington State Department of Fish and Wildlife (WDFW). Note that the presence of salmon in some stream or river reaches differs between datasets.

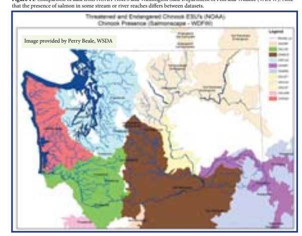
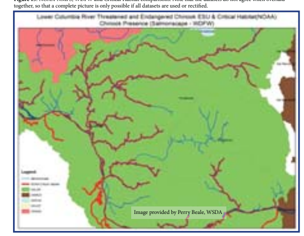


Figure 15. NOAA-NMFS, WDFW and Critical Habitat comparisons. Note that the datasets do not agree when overlaid together, so that a complete picture is only possible if all datasets are used or rectified.



OPPORTUNITIES FOR IMPROVING DATA BY LEVERAGING RESOURCES BETWEEN PRIVATE, STATE AND FEDERAL PROGRAMS

The challenges of a national scale risk assessment for hundreds of listed species requires an efficient use of resources that may already exist but are split between stakeholders, federal agencies and other entities, such as state agencies. Combining agency and other resources for the purpose of agreeing on authoritative data to be used in national scale pesticide endangered species assessment is a mechanism that may achieve relatively rapid agreement on additional authoritative data, some of which may already reside in the data collected and managed by private, federal or state entities such as NatureServe, the Forest Service, USDA field offices and/or state and private data collections. FESTF would like to thank USDA, EPA-OPP and Services staff for their innovative and open approach to leveraging resources to achieve data access and validation, and stands ready to vet its additional data as the national level endangered species assessment process evolves.

---

[1] (http://www.epa.gov/pesticides/science/efed/models/water/archives/pca_adjustment_dwa_9sep2010.html)
[2] EPA's website reports the average size is 700 square miles, but this is most likely a typographical error. By actual calculation, the average area is ~1,700 square miles with a median of 1,350 square miles.
[3] Committee on Ecological Risk Assessment under FIFRA and ESA, Board on Environmental Studies and Toxicology, Division on Earth and Life Studies, National Research Council, 2013
[4] *Interim Approaches for National-Life Pesticide Endangered Species Act Assessments, Based on the Recommendations of the National Academy of Sciences April 2013 Report, November 2013*
[5] Ibid., page 10