United States Department of Agriculture

DYNAMIC SOIL PROPERTIES HUB:

EPIC

Overview

The DSP Hub is fundamentally a flagship, innovative, high-end geospatial data science workbench that builds new data products from a wide variety of existing data sets. The Hub is focused on flexibility and agility to rapidly respond to customer requests for science-based soil property data. This Hub is a critical customer need at the Deputy Chief, Chief, and Under Secretary level.

The Hub expands USDA capacity to model and report on soil properties that change with conservation management on a human time scale. The Hub is focused on supporting the collection, storing and delivery of dynamic soil properties data for USDA Natural Resources Conservation Service, Conservation Innovation Grants (CIG), the Environmental Quality Incentives Program (EQIP), Soil Health Initiatives, and other programs including land use and conservation management information. This involves linking soil and conservation data bases in space and time, providing the ability to assess outcomes in conservation programs, leveraging siloed data and models from across the agency and divisions to maximize usefulness for conservation decision making.

- Dynamic soil properties are soil properties that change rapidly under the influence of land management, with the focus on conservation practice effects on soil (e.g. soil organic carbon).
- Conservation practice effects on soil properties is another way to refer to “environmental benefits” or “outcomes” of conservation practices.
- The DSP Hub is the first to feed the CD/CART “Advanced Benefits Services.”
- The DSH Hub will establish data standards and a peer review process to embed a science-based foundation into conservation program and practice evaluation.
- Science, statistics, and careful data structuring, data stewardship, and model integration will provide high-quality, defensible (i.e. authoritative) “benefits” estimates.

Published in cooperation with
Soil Skunkworks, TSPi & Wisconsin NRCS

Cover Photos: (Clockwise) Soil Colors. Hands together in partnership with stakeholders. Cows in a pasture. Cover Crops field.

Helping People Help the Land

USDA, NRCS, Soil Science
and Resource Assessment

USDA South Building South
Mail Stop 1600
Washington, D.C. 20250-1600

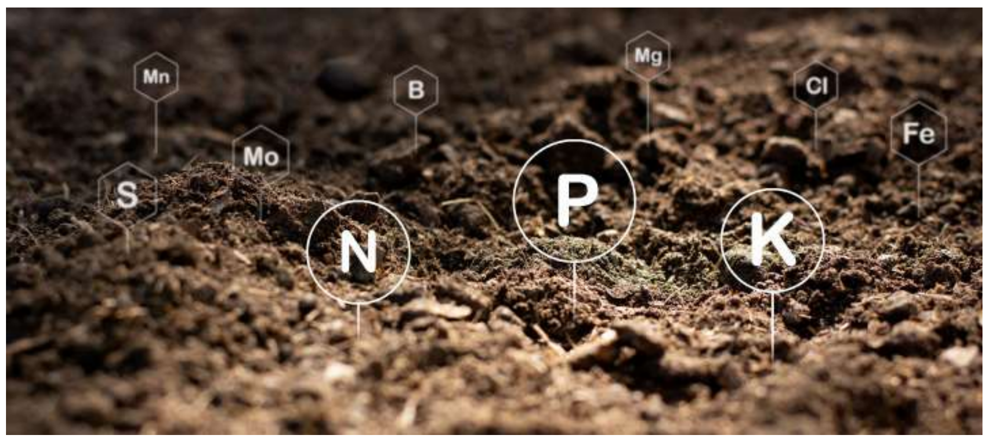
(608) 662-4422

https://www.nrcs.usda.gov/wps/portal/nrcs/main/soils/survey/

Find your local USDA Service Center at
www.farmers.gov/service-center-locator

Follow NRCS on twitter.
https://twitter.com/NCSS1899

Sign up for NRCS GovDelivery emails.
https://public.govdelivery.com/accounts/USDANRCS/subscriber/new



__Photos:__ (Top to Bottom) A waterway for small boats leads into the Chippewa flowage lake region of northern Wisconsin.

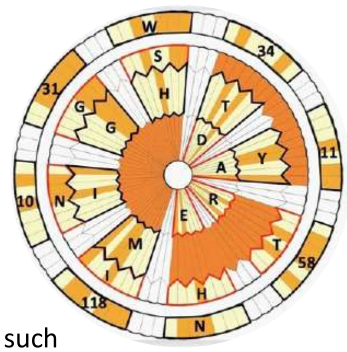
Our agency was born in 1935, during a time of hardship and desperation, when the very soil that put food on our tables was literally blowing in the wind. Wisconsin became the home of the first erosion control demonstration project in the country, the wildly successful Coon Creek Watershed in Vernon County. It was 22 miles long, nine miles wide, 92,000 acres over three counties, with outlet directly to the Mississippi River. There, the science and art of soil conservation to protect our land, water, food and nation, was born.

Dynamic soil properties (DSPs) are soil properties that change with natural and anthropogenic disturbances and stressors including agricultural and wildland management. DSPs are indicators of soil function and soil change over the human time scale (decades to centuries). Soil function describes what soil does, including ecosystem and agricultural services. Since soils typically develop on a geologic time scale, we generally infer soil change by comparing different conditions or management systems in a single type of soil.

Contents

Photo: NRCS conservation practices support wild rice growing on tribal lands.

Background



The Dynamic Soil Properties (DSPs) Hub is a flagship innovation that expands USDA capacity to model and report on soil properties that change with conservation management on a human time scale. President Biden mentioned one example in his Joint Address to Congress on April 28, 2021, when he highlighted the importance of cover crops and carbon sequestration.

The DSP Hub includes high-performance geospatial modeling with specialized software such as PostgreSQL, PostGIS, SAGA and R Studio Server. This high-resolution terrain analysis enables science-based estimates of the impact of conservation practices on soil properties, such as soil carbon and other environmental benefits to prioritize, evaluate, and improve the conservation program delivery.

The approved DSP Hub goal statement is as follows: The DSP Hub will provide an authoritative source for data and interpretations on soil properties that change rapidly due to land uses and conservation management. Specific objectives are:

1. Align to USDA enterprise data governance
2. Ensure dynamic soil property data and interpretations are science-based and authoritative
3. Provide a business tool to support the science-based processes
4. Engage DSP Hub customers and stakeholders

The 2-year strategic and tactical plan is included in Appendix A.

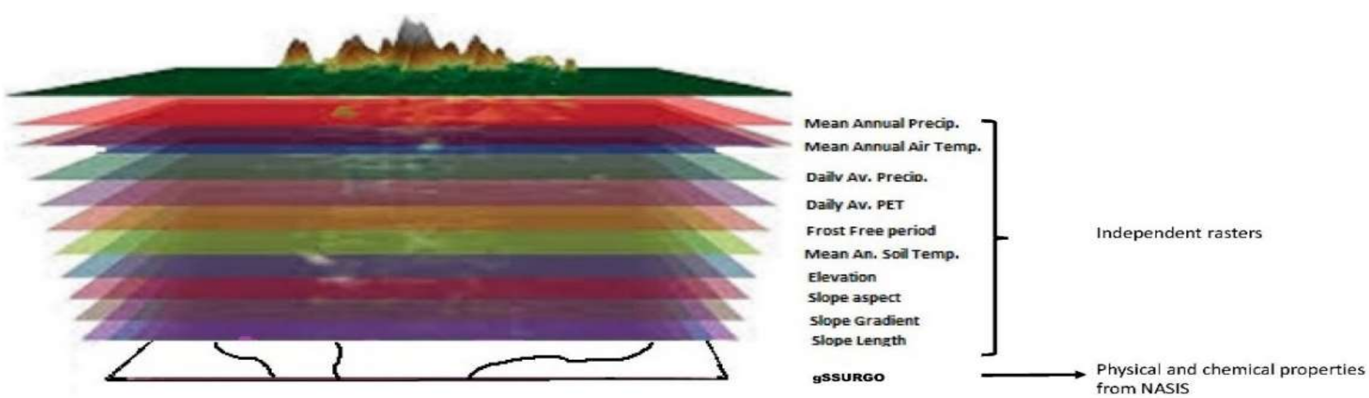
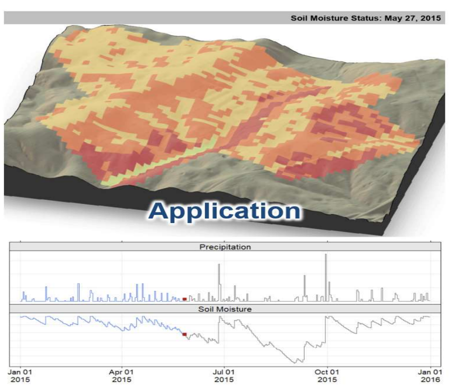
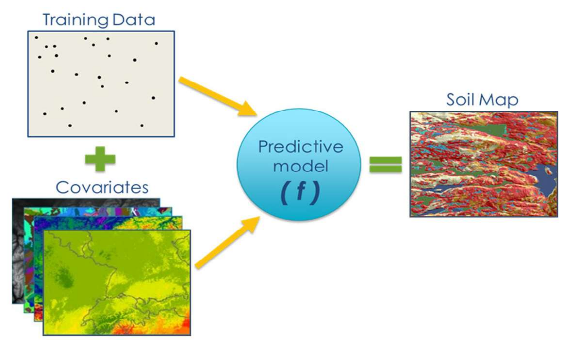


Photo: *Non-soil raster layers used in concert with soils data in a raster format for the development of interpretations.*

Fundamental Principles of DSP Hub Development

- The DSP Hub is a flagship innovation project for the USDA Natural Resources Conservation Service to update many existing methods and data assets with the latest technology.
- The DSP Hub uses SAFe Agile, with the flexibility to rapidly respond to Agency and Departmental priorities as well as iteratively learn and adapt in order to avoid large and expensive rework.
- The DSP Hub focuses on data management and modeling, with iterative architecture and processing design that intentionally starts small to test, learn, and adapt before scaling up.
- The DSP Hub will establish data standards and a peer review process to embed a science-based foundation into conservation program and practice evaluation.
- The DSP Hub will link to the Conservation Assessment and Ranking Tool (CART) to provide estimates of future environmental benefits.





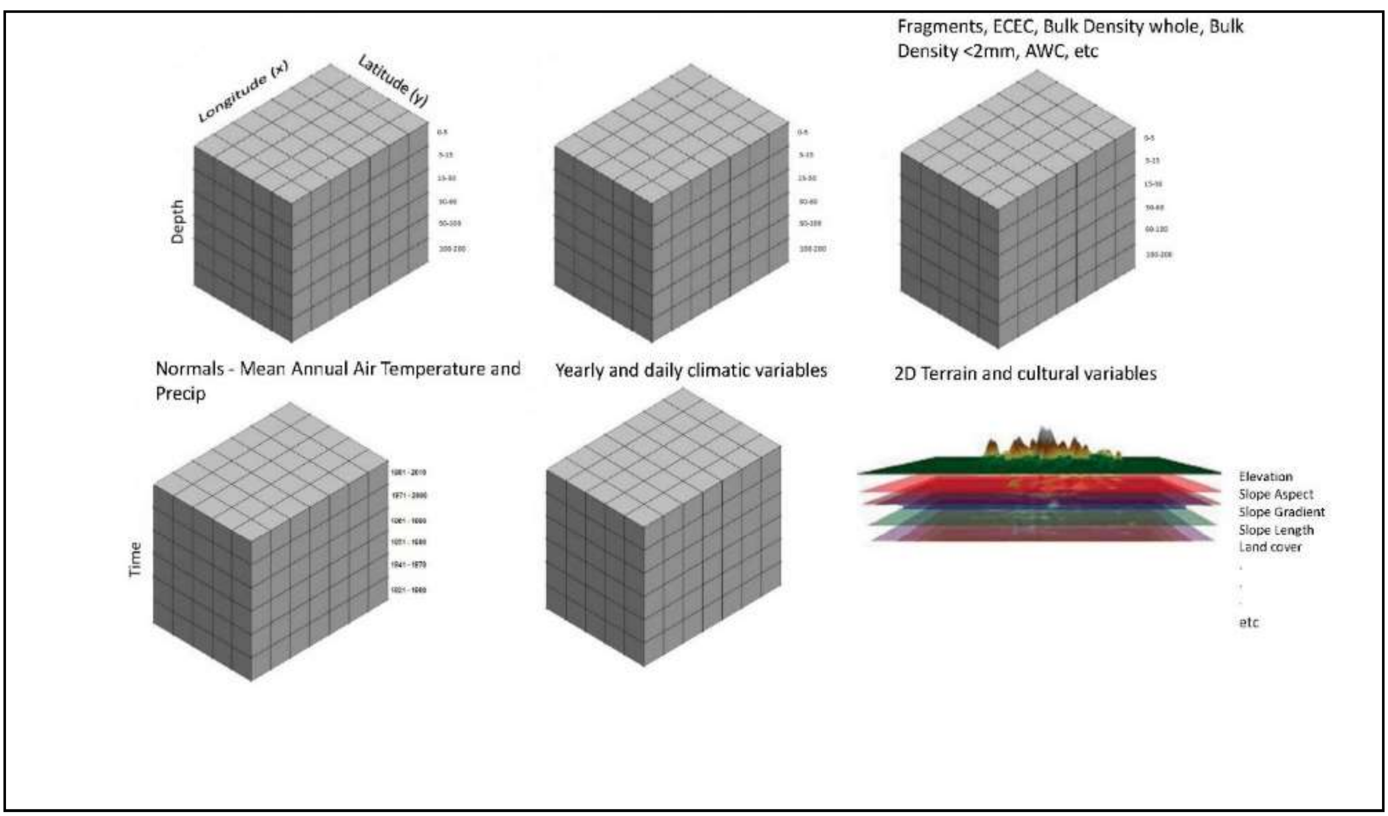


Photo: Multidimensional data: Data over an area varying with location, time, and depth.

Situational Analysis

- The DSP Hub is the processing engine for the Conservation Innovation Grants (CIG) and Conservation Practice Database (CPD) that are working on the 2018 Farm Bill requirements, Section 2307 (d)(2)(A)(i) "a compilation and analysis of effective conservation practices for soil health, nutrient management, and source water protection in varying soil compositions, cropping systems, slopes, and landscapes."
- USDA is under pressure to provide "outcomes" and "environmental benefits" as outlined in the 2018 Farm Bill over 20 times, some of which are dynamic soil properties (soil carbon).
- The treasure trove of soil data is an underutilized asset to be tapped for model development using advanced geospatial processing with imagery, LIDAR, and other terrain analysis covariates.
- The DSP Hub will need multiple flexible tools to transform legacy data into usable, integrated datasets.
- USDA has embraced consistency and governance in managing data across the Mission Area so that data can be more easily used and shared.
- As in other parts of USDA, NRCS has legacy data marts and data stores, making data difficult to access and analyze, with existing models often maintained on outdated technology, siloed databases, application data stores, partners websites, and individual workstations.
- Scientists in NRCS charged with analysis and decision-support for leadership often lack access to raw data and the tools to visualize, validate, and approve information for leadership dashboards.

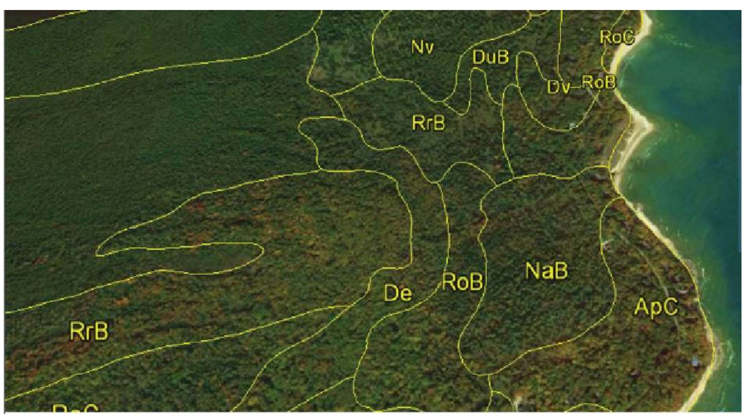
Photos: *Incorporation of spatial, temporal, climatic and biological data into soil survey. Concept is to create a 4-dimensional model of soil encompassing area, depth, time and hydrology, or other temporal soil properties. Dynamic soil modeling is an evolving approach to soil survey that incorporates recent and long-standing advances in pedology, ecology, and hydrology with improvements in digital technology. The goal is to capture and visualize the variability and diversity of the soil landscape, as affected by management and land use.*

Supporting and Expanding Capacity in EDAPT

Through this project, the DSP Hub team and NRCS leadership seek to support and expand the capacity of the Enterprise Data Analytics Platform and Toolset (EDAPT). The goals of USDA and EDAPT align with the goals of the DSP Hub and NRCS to provide consistent, high-quality science-based data through robust data access and advanced tools for the scientists working on urgent Departmental priorities, such as climate change initiatives. Data-driven decisions based on best-available science is an urgent need in the forefront of the DSP Hub project.

To support the USDA data and analytics program, the DSP Hub project has planned for the following methods, strategies, and approaches to partner with the Department on integration with EDAPT:

1. Portability – develop DSP Hub components to be portable to migrate to EDAPT.
2. Data outlet – EDAPT will be the preferred outlet for data from the DSP Hub.
3. Initial design testing – use SAFe agile as an iterative methodology to ensure operational feasibility and successful migration with no loss in functionality.
4. Data management planning – data management needed for a future NRCS conservation data warehouse will be analyzed and established to be leveraged in EDAPT.
5. Capacity for high-end geospatial analysis – the science-based terrain analysis for soil geospatial modeling needed for the DSP Hub will expand the capacity of EDAPT for all USDA agencies.
6. Iterative and targeted migration – at the earliest opportunity, sections of the DSP Hub (see epics below) can be migrated to EDAPT.



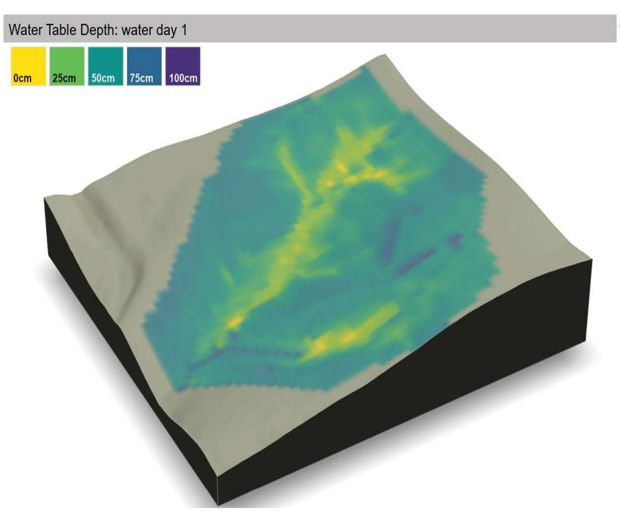
Photos: *Dynamic Soil Properties Hub: The Fundamental Resource. Soil survey data represents the foundation of land-use planning and management The Soil and Plant Science Division has developed and maintains standard operation procedures (laboratory and field) Soil products from the utilization of ecology, hydrology, climate and temporal soil property data will usher in a new era of displaying, querying, integrating and analyzing soils data through digital based models.*

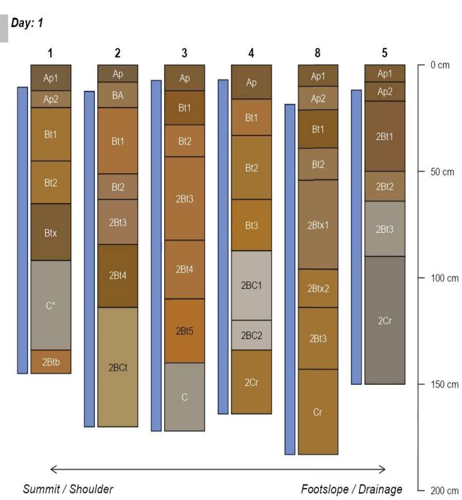
Epics and Roadmap

To deliver customer value in accordance with statutory mandates (2018 Farm Bill requirements), the following epics have been identified as critical components of the DSP Hub environment, technology, and workflows. The epics focus on a 12–18 month roadmap, with minimum viable products (MVPs) identified as technical solutions to prove architectural approaches to meeting business outcomes.

Epics Summary

EPIC	MVP FOR 12-MONTHS	NEXT STEPS
Soil and conservation practice data geospatial modeling research environment	Establish a research environment initial design for testing and evaluation of scaling and costs	Pursue phased EDAPT migration after design and testing of processing
Import and export soil data and models to partners	Automate CIG On-Farm Soil Health Demonstration Trials (OFSHDT) data and DSP4SH data import	Pursue phased EDAPT migration to design and test needed functionality
Curated data pipeline	Core conservation practice activities database for two models	Pursue phased EDAPT migration, align with CART PI14 activities
Single transaction processing of small data sets to provide curated results in live time	Connection from DSP Hub processes to CART and CIG for two models	Pursue phased EDAPT migration, align with CART PI14 activities
Support model/data requests from internal and external customers	Manual process, with few if any external requests. Processes will be internal for 12 months	Analysis of zRoles and process map to implement workflow
Provide data/model publishing governance workflow	Manual process for internally derived data products using the zRoles structure established.	Analysis of zRoles and Open Data and other USDA data governance processes





Photos: *Soil survey is continuously evolving from the original soil survey mapping to the continued development of interpretations. SPSD is continuing to develop new soil survey products to assist landowners, land managers and conservation planners utilizing basic soils data. All land use and management begins with soils data and with the addition of ecological, climatic, hydrologic data and improvements in spatial display. NRCS will be able to address internal and external customer needs through these processes.Example: Look at the changes in technology within the available maps for landowners. During the 1970s, the soil survey map for most landowners was their best planning tool in viewing their property. Today, in a matter of seconds, users can download the latest copy of an aerial flight.*

EPIC 1—SOIL AND CONSERVATION PRACTICE DATA GEOSPATIAL MODELING RESEARCH ENVIRONMENT

The DSP Hub requires high-performance technical computing to develop, test, and improve soil science-based geospatial modeling (terrain analysis with specific software requirements). The agency scientists need to rapidly ingest multiple, large authoritative datasets, transform, reaggregate, and analyze the results to develop new science-based data sets. Modeling and data development will focus on how soil compositions and attributes like slope, landscape, and climate influence soil properties (example: soil carbon) and the effectiveness of conservation practices such as cover crops, crop rotation reduced tillage, and others.

BUSINESS OUTCOMES

- Validated data with science-based estimates of dynamic soil properties.
- NRCS scientists can perform needed geospatial modeling of soil data, tapping the data of the largest natural resource inventory (National Cooperative Soil Survey, SSURGO) for curated state-of-the art soils information.
- Models and data products developed for internal and external customers to predict the impacts of conservation management on soil properties.
- Assist NRCS and USDA with statutory compliance (2018 Farm Bill, 2018 Evidence Act).

FUTURE STATE DIAGRAM

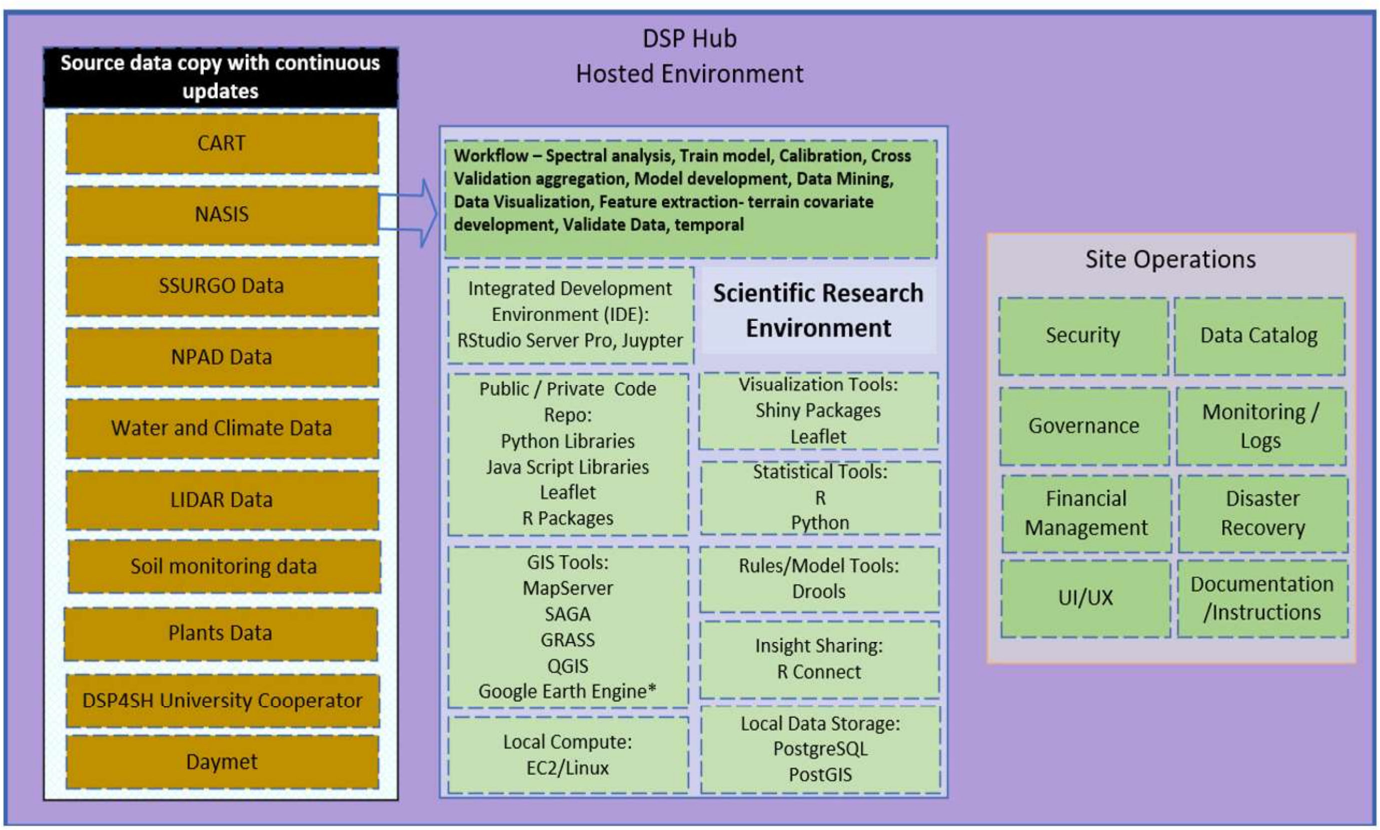


Photo: Scientific Research Environment: Provides a secure environment to safely integrate, explore, and model data using advanced analytic tools to test hypothesis and develop scientific models/calculations. This environment connects directly with the storage/compute environment (aka the curated data pipeline) to access necessary data/storage/compute. Examples of tasks to be conducted in this environment: 1. Data collection and mining. 2. Standards and quality control 3. Feature extraction 4. Region/landscape segmentation model development 6. External validation and integration

MVP

- The first two models identified as priorities to assist other projects in meeting statutory mandates are the Soil Health Assessment Protocol Evaluation (SHAPE) model for soil carbon and COMET models for soil carbon.
- Soil properties modeling (e.g., cover crop impacts on soil carbon, soil moisture, aggregate stability, and biological activity) can be expanded using many of the same data sets without expanding the development footprint.

Leading Indicators (Success Measures of the MVP)

- Large files of geospatial data (imagery, soil data, LIDAR, etc.) are available for running models.
- Processing speed is adequate to run the models.
- Science-based tools/software to develop and run the models are functioning.
- Agency scientists (data stewards) with responsibility for the Federal Geographic Data Committee (FGDC) standards for authoritative soil data have access to the research environment to begin data development.
- Data sources can be loaded quickly to perform analysis and testing (no approval wait time).

Features and Nonfunctional Requirements

- Easy ingestion of multiple, large, authoritative datasets with the ability to transform, reaggregate, model, and create new data sets
- Approximately 40 users during the next 12–18 months
- Analysis of data sources: See Appendix B
- Direct access (manual or API) to servers and applications
- A user interface to reduce manual work interacting with multiple disparate tools
- An adaptable rules engine that can be directly adjusted by DSP Hub owners
- Ability to experiment with DSP data with machine learning/predictive engines/fuzzy logic to explore algorithms and automation
- UI/UX - DSP requires a user interface and automation to reduce manual work interacting with multiple disparate tools.
- Implementation of the define technical stack of software tools supporting machine learning, statistical analysis, advanced spatial capabilities, and data management maturity management. A diagram with the software solutions for DSP Hub is in Appendix G.
- A computational engine with adequate compute capabilities to enable advanced soils modeling. Reasonable compute capacity has been configured for initial and projected hardware sizing considering the defined software stack, large data set sizes, intensive spatial processing, and data visualization so that soil scientist work progresses at a reasonable pace.

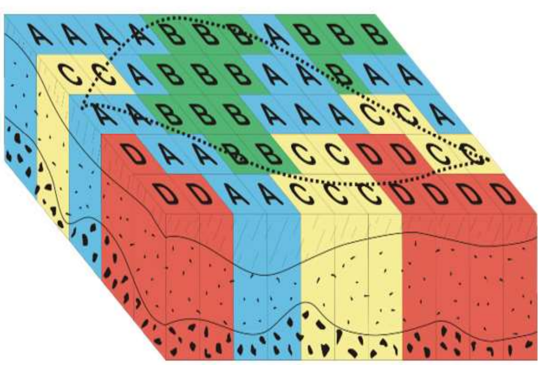


Photo: Specific soil properties or soil type assigned to cells. Polygon map units are viewed mostly by viewers in dominant condition or component. With pixels, changes occurring in soil properties are available with a confidence that allows the user additional information in land management. Statistical confidence is the ability to calculate and communicate deviation from measured property or named soil inclusion of the limiting soils.

12-MONTH ROADMAP

The DSP Hub business and development team has completed the analysis on the technology and architecture requirements for the MVPs. The next 12–18 months will be focused on experimentation and analysis of scaling in an agile design process. The following is planned:

- Stand up Linux servers with the required geospatial processing software and tools
- Add data sets to develop MVPs for Farm Bill requirements
- Test recommended software and tools
- Build a UI to provide basic functionality to internal scientists
- Test APIs needed for serving data internally
- Use manual workflow and approvals processes for data release until the workflow and approval process is established
- Obtain an Authority to Operate (ATO) for the MVP

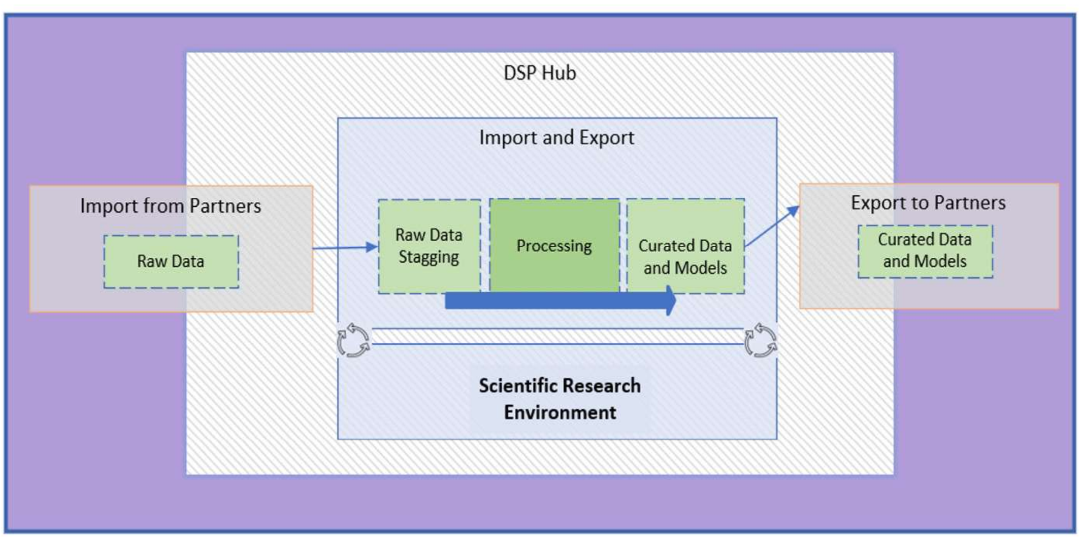
EPIC 2—IMPORT AND EXPORT SOIL DATA AND MODELS TO PARTNERS

Data must be imported from partners to process in the geospatial-soils models. Partners include the Conservation Innovation Grant (CIG) On-Farm Soil Health Demonstration Trials, Dynamic Properties for Soil Health (DSP4SH) partners, pXRF urban assessment datasets, and others. Data and models will be provided to internal and external customers to support the 2018 Farm Bill Requirements. Data sets and models will be made available through existing or planned assets, such as EDAPT, the CIG website, the Conservation Practice Database, or Agricultural Data Commons.

BUSINESS OUTCOMES

- Integration of data currently stored on laptops, siloed data stores or in partner databases into DSP Hub for modeling
- University and other partners can import data for processing in DSP geospatial models to obtain results (example, CIG On-farm Soil Health Demonstration Trials)
- University and other partners can export a model or dataset for their own use and further research

FUTURE STATE DIAGRAM OF THIS EPIC



Photos: *Ability to export and import large amounts of data to partners such as universities and other agencies outside of USDA.*

MVP

- Two small internal projects have been identified as MVPs to test and further refine the methods (CIG On-Farm Soil Health Demonstration Trials project and DSP4SH projects)
- Use the assets of the Agricultural Data Commons and EDAPT to provide data externally

Leading Indicators (Success Measures of the MVP)

- Administrative controls are aligned to the zRoles for the DSP Hub for important and exporting data sets
- Data quality and type standards are enforced
- Alignment to other existing assets to avoid duplication

Features and Nonfunctional Requirements

- There are aligned requirements for this Epic in the CIG project working on the On-Farm Soil Health Demonstration Trial participants (CIG Team #2)
- Ability to import or export large data sets outside of USDA
- Ability to set data standards for quality and type
- Integration of zRoles structure and workflow in Epic 5 (Customer intake) and Epic 6 (Governance workflow)

12-Month Roadmap

- The road map for this Epic is deeply integrated with CIG Team #2 development work.
- Identify and leverage CIG Team #2 technology and development
- Load data sets manually
- Develop data standards and data typology and ontology in a modular way
- Integrate with existing export options such as Agricultural Data Commons and EDAPT

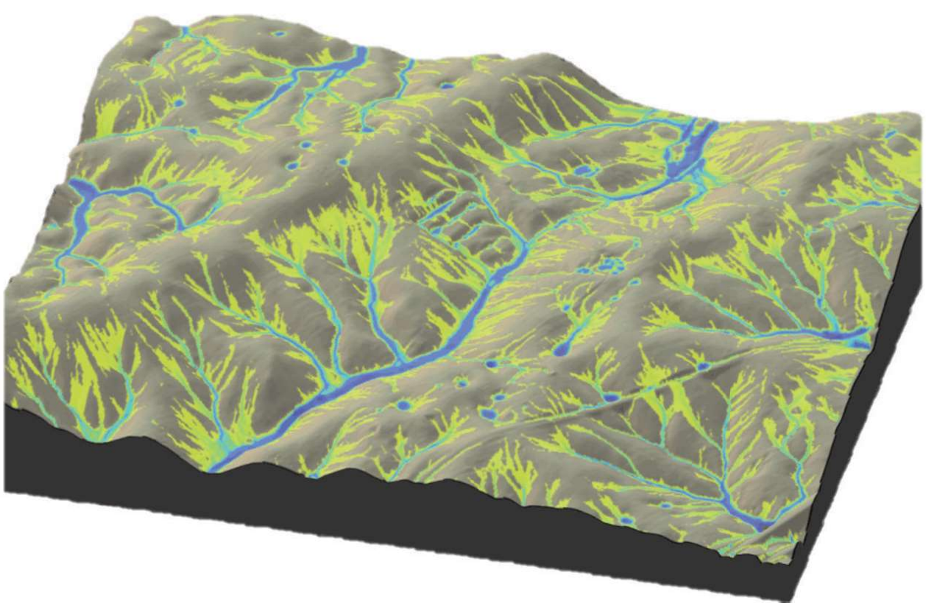


Photo: Dynamic Soil Survey: Real-time modeling of soil moisture, runoff, and dynamic soil properties. This is an uncalibrated simulation of runoff during a hypothetical rainfall event lasting 1 hour, with 1 inch accumulation. This simulation cannot account for vertical/horizontal movement of water through the soil. This simulation does not integrate local knowledge of geology (e.g., fractured limestone) which allows for rapid, deep percolation in otherwise closed depressions.

Without calibration (e.g., details outlined above) and validation (e.g., on-site point measurements), the results are little more than a demonstration.

EPIC 3—CURATED DATA PIPELINE

The DSP Hub will have three main types of data—data provided by internal and external customers for processing (Epic 2, import data from partners), covariate data to run models on customer's data (Epic 1, Soil and Conservation Practice Data Geospatial Modeling Research Environment), and results data generated from the models (Epic 2, export data to partners).

To develop authoritative data, the DSP Hub scientists will upload multiple NRCS and external data sets, combine them, and run combined data sets through transformations (models/business rules/etc.) ultimately resulting in new, 'curated' authoritative data sets. Any transformations and aggregations by the scientists will maintain and document data lineages as needed for supporting authoritative and science-based products.

BUSINESS OUTCOMES

- A stamp of approval as authoritative DSP Hub data for dynamic soil properties
- Repeatable reports on the data (historical records)
- Transparent data standards and data management protocols

FUTURE STATE DIAGRAM OF THIS EPIC

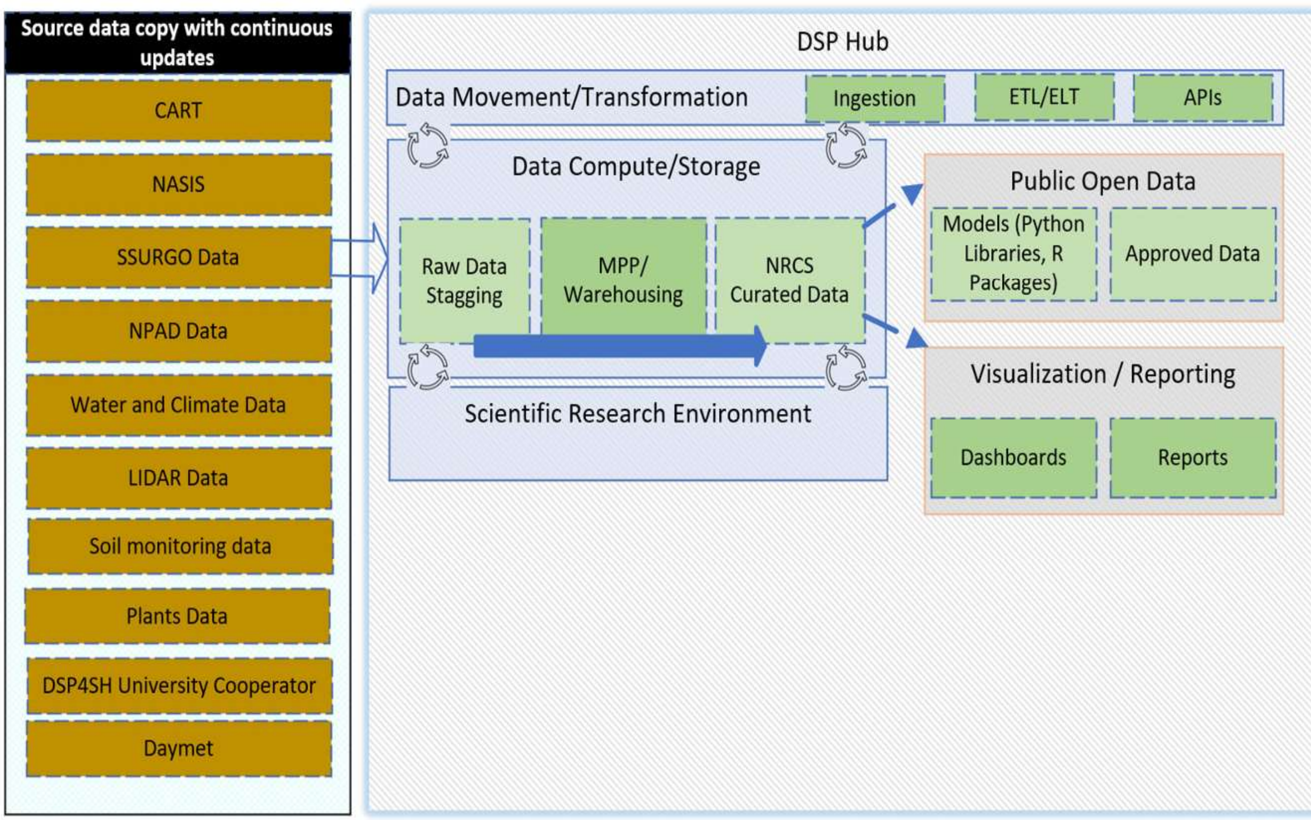


Photo: Photo: Data Compute/Storage: Data storage and compute infrastructure/software required for storing and processing DSP Data. This environment is also known as the 'curated data pipeline'. This is where rules/models/etc., are applied to transform data into a curated data set. This functional area will support DSP's large processing needs. When hypotheses have been developed/proven within the Scientific Research Environment then the relevant rules/models/etc., would be applied to the curated data pipeline. NRCS Curated Data is the purpose-built, read-only data sets designed to meet specific work group, business function, or activity set. This curated data can then be leveraged by internal and external stakeholders.

FEATURES AND NONFUNCTIONAL REQUIREMENTS

- Ability to import data from internal and external partners (Epic 2, Import/export data from partners)
- Authoritative data with robust metadata that includes data lineage, manipulations, and computations
- Data management protocols in place that could support audits (statutory mandates on Congressional or other reporting, or a potential “carbon bank”)
- Ability to aggregate and store raw data from disparate data sources
- Ability to incorporate the data development and approval workflows from Epic 5 (customer intake workflow) and Epic 6 (governance workflow)
- Ability to transform data into analytic structures including capturing time series data
- Ability to export data to external customers (Epic 2, export data to partners)
- Interactive relational database (reporting focused) storing soil properties, location, field management information, and metadata over time with integrated version control for the data
- Ability to develop an authoritative data set from multiple NRCS transitional systems data (ex. CIG/CD/NASIS) and other reference data sources
- Data organized in libraries with appropriate aggregations for the customer
- Ability to capture spatial and temporal data
- Outputs to EDAPT would include a variety of geospatial maps with predicted properties under variable land management scenarios

MVP

- Use small initial MVPs above to analyze and develop a data management structure and process
- Establish data standards for the small data sets from the CIG OFSHDT projects, the SHAPE curve model, and the COMET carbon model
- Identify data patterns for initial data mart design(s) for DSPs and conservation practices (this can be used as an asset in EDAPT when migrated and will use Kimball iterative design methodology)
- Establish data processing protocols for the main internal data sources and types (NPAD, CIG, DSP4SH, etc.) that can be automated as extract-transform-load processes (ETLs) for a future data warehouse (this can be used as an asset in EDAPT when migrated)
- Initial identification and development of processes required for data management maturity

LEADING INDICATORS

- Selection of a small core initial data set to build a solid backbone
- Iterative design with stakeholders engaged in curating high-value data
- Data standards, agreed-upon schemas, and ontologies across conservation management/practice data for statistical validation
- Integrate Epic 5 (customer intake workflow) and Epic 6 (governance workflow) for clear approval workflow, release of authoritative data with metadata, and fine print on use limitations

12-MONTH ROADMAP

The MVPs will establish the high-value-core data structure and processes for a curation pipeline.

- The CIG OFSHDT data schema has been reviewed by internal NRCS scientists and external partner scientists for alignment to the cover crop reference data and ontology and conventions alignment
- Use existing National Planning and Agreements Database (NPAD) structures for practice and resource inventory tables and geometries to align requirements
- Use a manual/email process for approvals until the workflow and zRoles is established
- Evaluate the data structures and patterns in the CIG schema, NPAD tables, and alignment to authoritative reference data sets to establish a small MVP data mart with manually updates
- Determine administrative requirements for auditable system of records for a “carbon bank”

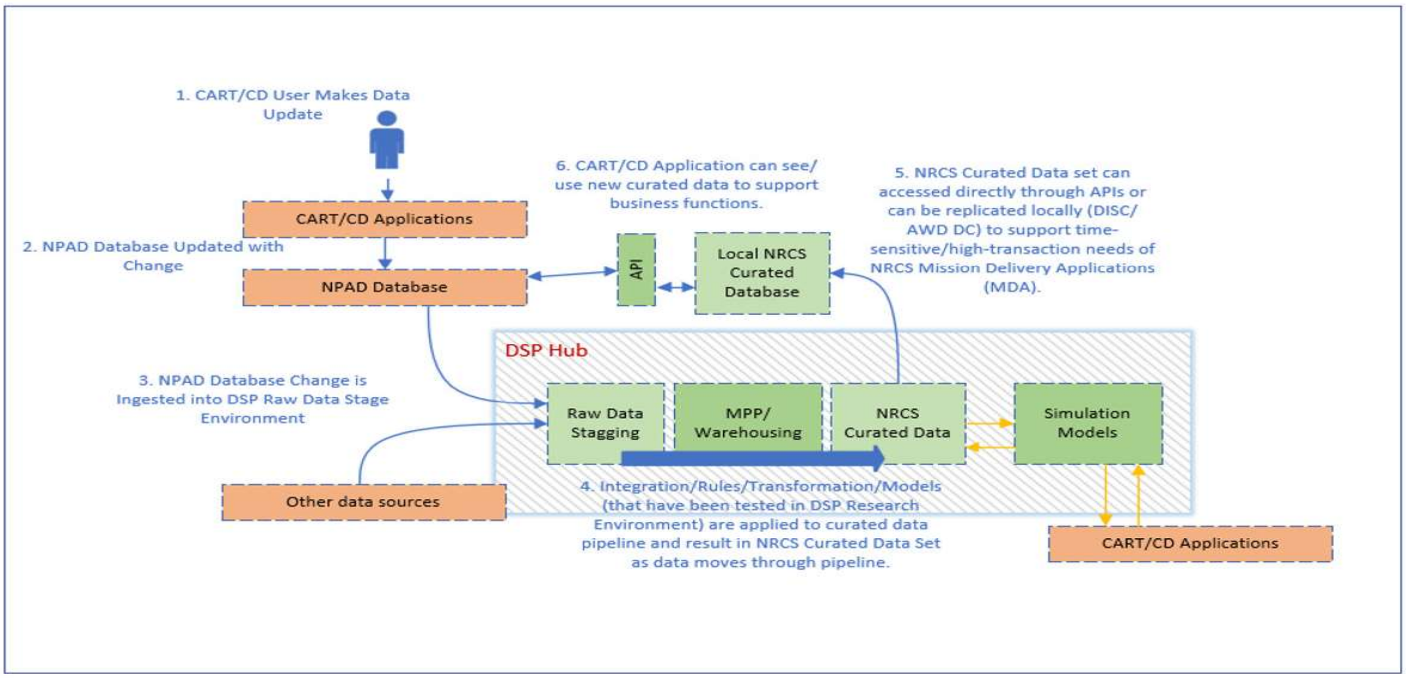
EPIC 4—SINGLE TRANSACTION PROCESSING OF SMALL DATA SETS TO PROVIDE CURATED RESULTS IN LIVE TIME

The DSP Hub will provide on-demand processing for single-transactions to support conservation program delivery and the Farm Bill requirements on other projects. For example, CIG and other soil health projects will submit data and obtain modeled results for that data (example SHAPE curve results for a laboratory result). Conservation Desktop and CART is building the "Benefits Module" which will depend on the DSP Hub for soil-related environmental benefits. There is a pilot project with CART on the COMET-Planner method, the MVP model for the DSP Hub.

BUSINESS OUTCOMES

- Modeled SHAPE results for soil health projects and customers
- Improvement of the science in CART processes to connect to environmental benefits

FUTURE STATE DIAGRAM OF THIS EPIC



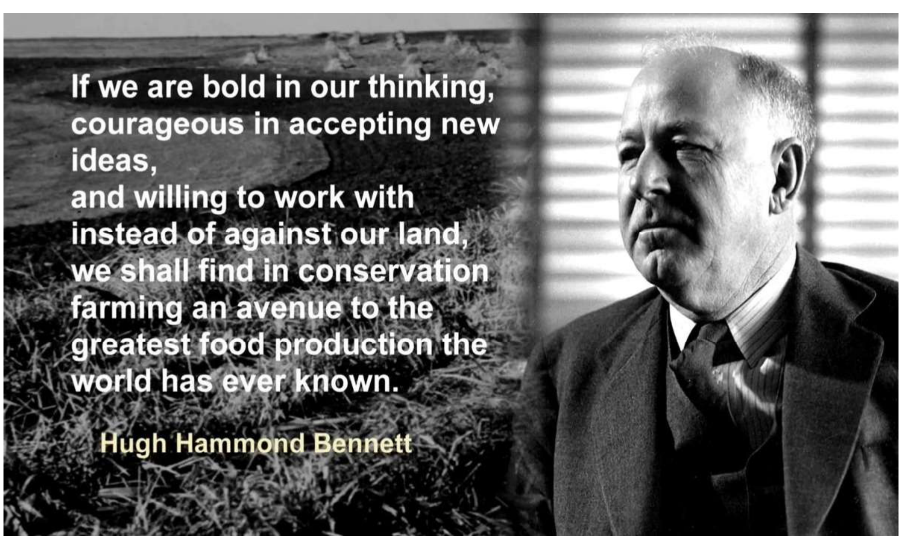


Photo above: *We will have developed simulation models for dynamic soil properties to predict soil properties and interpretations under different soil management systems. This is an old problem and our biggest challenge. It provides the relevancy for soil information. To effectively address this issue, we need to address the other "be" issues list below. Currently we recognize ranges in NASIS, but not how ranges are affected by management. The future will necessitate understanding how dynamic soil and vegetation properties change (i.e., the temporal dimension) and can be managed sustainably*

FEATURES AND NONFUNCTIONAL REQUIREMENTS

- Live-time interactions between transactional systems and the DSP Hub
- Processing of data sets from CIG and CART
- Retain results of computations as needed in DSP Hub storage
- Use of DSP Hub modeling tools including R and Python
- Support live-time results with a frequency initially supporting 2,000 to 10,000 requests per day
- Use of FPAC standard REST API

MVP

The two initial MVPs are the same for testing this Epic:
- The COMET model is being analyzed and planned to integrate with CD/CART
- The SHAPE results are being analyzed and planned to deliver data to the CIG data store

LEADING INDICATORS

- Selection of a small core initial data set to build a cornerstone
- Transactional connection to the DSP Hub and the internal systems needing curated results
- Transactional processing that works in live time

12-MONTH ROADMAP

By the end of 12 months, the SHAPE and COMET methods should be functioning with a live-time connection to the source systems.

- The CART Benefits Module should have a DSP Hub-driven results that feed the benefits report customers obtain in USDA Service Center Field Offices.
- Automated SHAPE results for the soil properties selected by the Soil Health Division are being fed into the CIG database

Photo: *NRCS works with agricultural producers across the country to deliver voluntary climate solutions through climate-smart production decisions and practice applications. NRCS has established conservation practice standards to guide the design, installation, and maintenance of conservation practices. There are approximately 35 NRCS conservation practice standards that deliver quantifiable carbon sequestration and/or greenhouse gas reductions.*

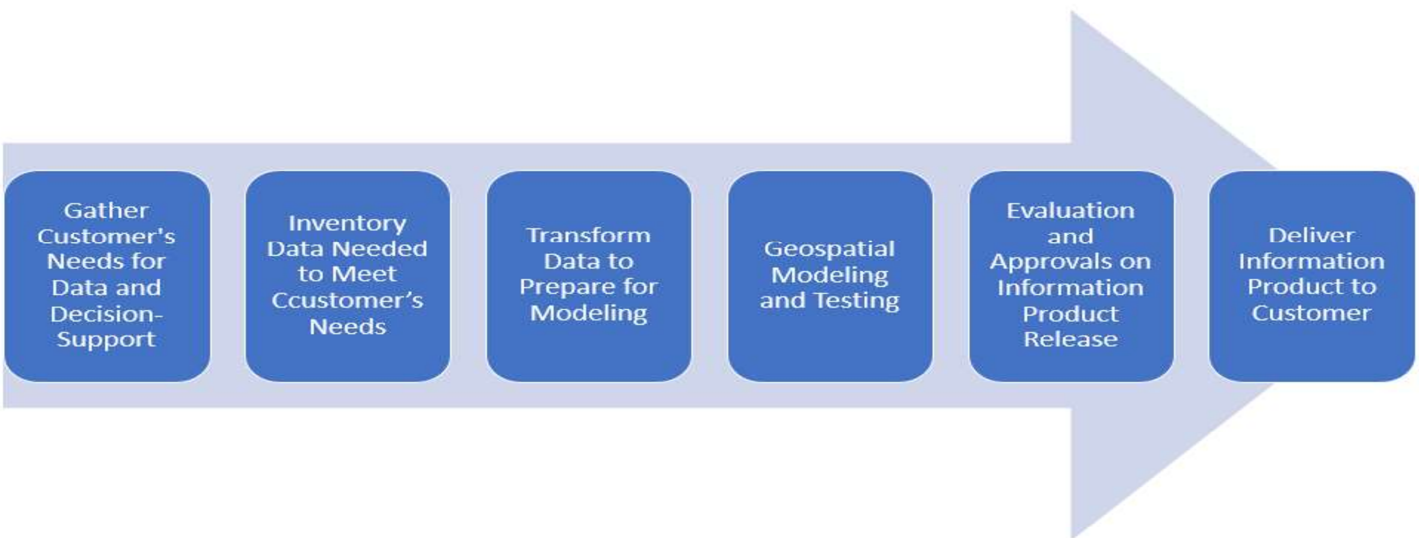
EPIC 5—SUPPORT MODEL/DATA REQUESTS FROM INTERNAL AND EXTERNAL CUSTOMERS

The DSP Hub will need workflow management to support customer requests for soil information, data planning, transformations, modeling, and developing a data delivery method. The workflow has a draft zRoles structure and is illustrated below.

BUSINESS OUTCOMES

- Efficient, transparent process for meeting customer soil information requirements
- Technical assistance from experts in the data to ensure their product is developed correctly
- A process that includes data validation and approvals if they require authoritative data (Epic 6, governance workflow)

FUTURE STATE DIAGRAM OF THIS EPIC



DSP Hub = data and modeling to evaluate 340 effectiveness = temporal geospatial data on soil properties such as soil carbon, soil temp, soil moisture, soil biology, and other dynamic soil properties.

FEATURES AND NONFUNCTIONAL REQUIREMENTS

- Customers need to be able to submit a request for a soil data product and track their request
- DSP Hub owners need to be able to interact with customers and update status of products
- DSP Hub owners can track, prioritize, and tag customer requests
- Workflow is integrated with a UI and zRoles structure for permissions and approvals

MVP

- Start as a manual/policy-based process
- The zRoles structure is designed for the DSP Hub, scalable to SSRA and NRCS-wide data stewardship structure as needed

LEADING INDICATORS

- Customer-friendly intake and data planning process with technical assistance from DSP Hub staff
- User interface to track workflow statuses
- ZRoles structure integrated into workflow

12-MONTH ROADMAP

During the first 12 months, the DSP Hub will not be able to advertise products, as the SAFe agile plan is to establish foundational structures and processes before scaling up. This epic is therefore likely to remain a manual process for urgent requests, and focused on the initial 40 internal users and internally developed data products.

Photos: *Environmental benefits, as outlined in the 2018 Farm Bill, direct NRCS to prioritize Conservation Planning and Program Delivery (e.g., the most conservation for least cost.) Outcomes, as outlined in the 2018 Farm Bill, direct NRCS to articulate more than conservation outputs (i.e. , number of contracts, acres treated, and dollars invested). Both environmental benefits and outcomes can be represented similarly (e.g., tons of soil saved, expected nutrient load reductions, energy savings, etc.) Environmental benefits and outcomes are both referenced repeatedly in the 2018 Farm Bill.*

Epic 6—Provide Data/Model Publishing Governance Workflow

The DSP Hub process workflow includes the ability for data stewards to coordinate the scientific peer review and approval process to provide authoritative data to the public. This Epic is integrated into the workflow with Epic #5 for a seamless customer intake and approval process.

Business Outcomes

- Authoritative soil information approved by the agency and scientists that oversee the Federal Geographic Data Committee (FGDC) standards for authoritative soil data
- Science-based review process to provide validated results
- Publicly available validated data provided through EDAPT and other outlets to meet Farm Bill requirements

Future State Diagram of This Epic

This workflow is integrated with the customer intake workflow.

MVP

- This can be started as a manual/policy-based process.
- The zRoles structure is designed for the DSP Hub, scalable to SSRA and NRCS-wide data stewardship structure.

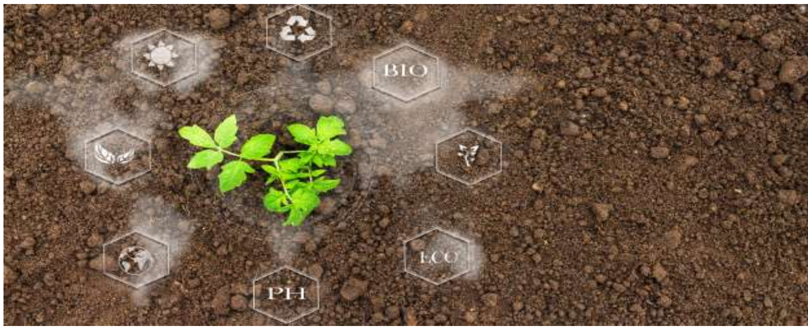
Leading Indicators

- A clear approval process and supporting policies
- Identified data stewards with assigned roles and scopes
- A zRoles structure that aligns to the approval process and permissions

12-Month Roadmap

The initial plan is for a manual, policy-based framework while the infrastructure is built out:

- Manual/email options until zRoles and automation
- Minimal integration with data curation pipeline solutions at first
- Complete the analysis of the zRoles plan with more automation if possible, within 12–18 months
- Demand will initially be for internal customers (CIG/CD/CART) and current partners (no advertising until environment, workflow, and staffing is in place).



Photos: *The Dynamic Soil Properties team is focused on supporting the collection, storage and delivery of the next generation of dynamic soil properties models for Conservation Innovation Grants, Environmental Quality Incentives Program, Soil Health and other programs including land use and conservation management information. This involves linking soil and conservation databases in space and time, providing the ability to assess outcomes in conservation programs, leveraging siloed data and models from across the agency, and divisions to maximize usefulness for conservation decision making.*

APPENDIX A—DSP HUB APPROVED 2-YEAR STRATEGIC/TACTICAL PLAN

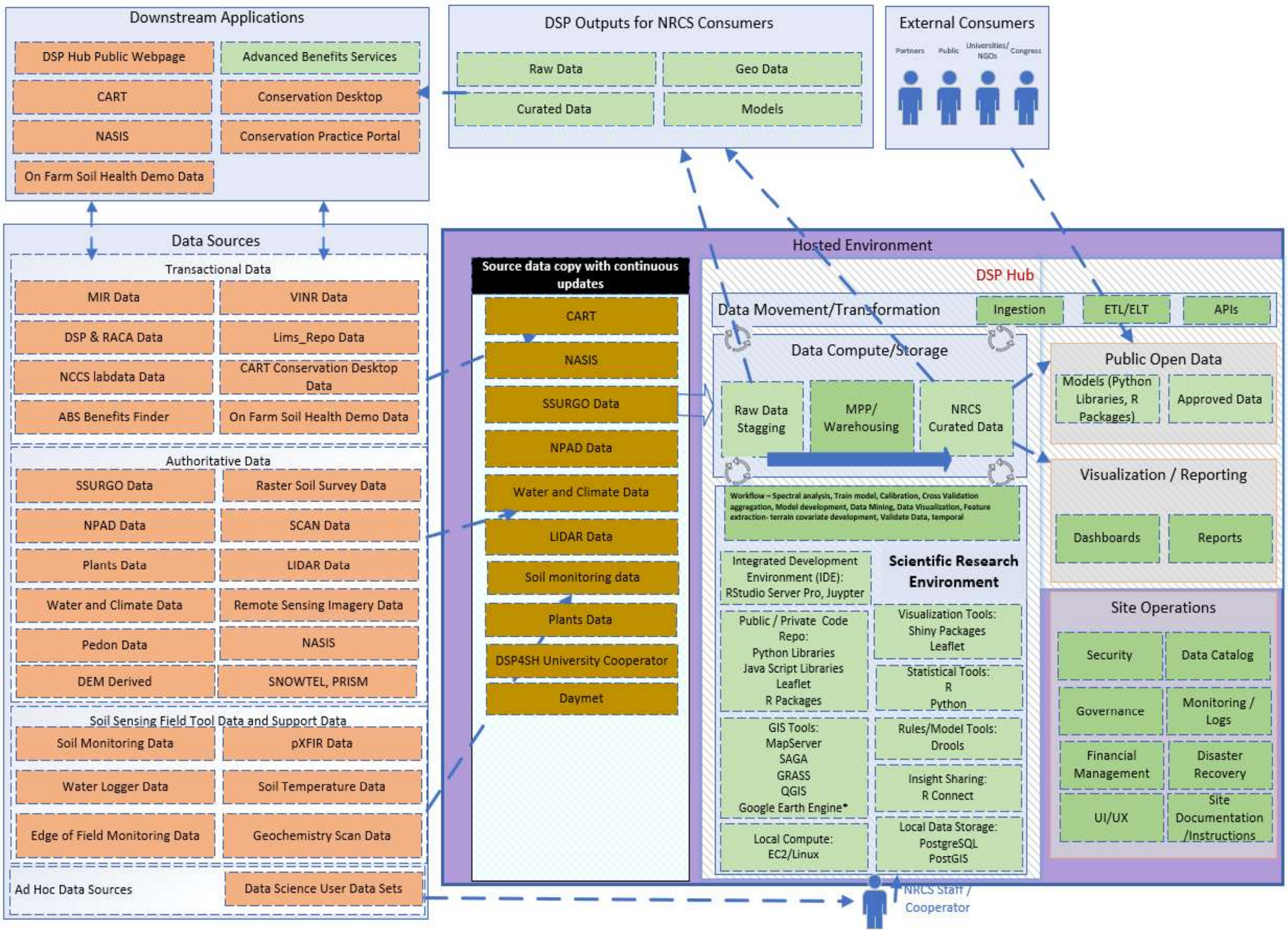
Dynamic Soil Properties (DSP) Hub Goal Statement

The DSP Hub will provide an authoritative source for data and interpretations on soil properties that change rapidly due to land uses and conservation management.

Objective Align to USDA enterprise data governance	**Objective** Ensure dynamic soil property data and interpretations are science-based and authoritative.			**Objective** Provide a business tool to support the science-based processes and data stewardship		**Objective** Engage DSP Hub customers and stakeholders
Strategy Include data stewardship and authoritative data concept (single source of truth)	**Strategy** Provide a scientific framework and decision-making process that supports and enforces the DSP Hub business process.		**Strategy** Adopt advisory structure for DSP Hub (proposed DSPAC)	**Strategy** Prioritize the use of authoritative data sets for DSP Hub processes	**Strategy** Develop the DSP Hub technology	**Strategy** Identify customers, service providers, and stakeholders
Tactic Develop policies for data stewardship and authoritative methods	**Tactic** Adopt a data quality rating framework to evaluate existing data products.	**Tactic** Document data processing and approval process for data products	**Tactic** Identify scopes, roles, and areas of expertise needed.	**Tactic** Link, use, and develop authoritative data	**Tactic** Provide raw data repository for priority initiatives (#1 = Soil Health)	**Tactic** Prioritize intial customers and stakeholders for the MVP
Action Steps Draft policy structure based on DSP Hub Strategic and Tactical Plan	**Action Steps** Develop data standards for use by other teams (e.g. Soil Health Database-OFSHDT)	**Action Steps** Draft zRoles framework to support DSP Hub business process	**Action Steps** Propose candidates and draft policy for discussions with Deputy Chief	**Action Steps** Evaluate data assets: SHDB, PEDON, and COMET reference data	**Action Steps** Work with TSPi on a Minimally-Viable Product initial design	**Action Steps** Focus initial design on SHAPE interpretation and carbon outcomes.

APPENDIX B—DIAGRAM OF FUTURE STATE DSP HUB

The components of this diagram are split out in the document to represent each use case.



APPENDIX C—LIST OF PROPOSED RANK 1 DATA SETS

Agency	File Type	Data Name/Source	Interval	Tool / Integration Method	Source Size (GB)	DISC Hosted (Y/N)
NRCS	geotif	Added - DSP Computations		GDAL, FME	10,000	y
USDA	geotif	1 and 2 Meter LiDAR Bare Earth and Hill-shade	As Updated	GDAL, FME	17,600	y
NRCS	SQL Server Database	CART/CD (NPAD)	Near Real Time	GDAL, FME	381	y
NRCS	SQL Server Database	SSURGO (SDMOnline)	Yearly	GDAL, FME	371	y
USDA	geotif	geotif	As Updated	GDAL, FME	229	y
NRCS	SQL Server Database	NASIS	Near Real Time	GDAL, FME	133	y
USDA	geotif	100m covariates 1	As Updated	GDAL, FME	37	y
NRCS	SQL Server Database	ncsslabdata	Near Real Time	GDAL, FME	31	y
PRISM	geotif	Air Temperature	As Updated	GDAL, FME	25	y
USDA	file geodatabase per state	1961-1990 Monthly Average Precipitation by State	As Updated	GDAL, FME	0.05	y
USDA	geotif	3, 10, and 30 Meter National Elevation Dataset	As Updated	GDAL, FME	17600	y
LANDSAT	geotif	Band 1 – blue	As Updated	SAGA/Post-GIS: Landsat Derived	34	y
LANDSAT	geotif	Band 2 – green	As Updated	SAGA/Post-GIS: Landsat Derived	34	y
LANDSAT	geotif	Band 3 – red	As Updated	SAGA/Post-GIS: Landsat Derived	34	y
LANDSAT	geotif	Band 4 – near infrared	As Updated	SAGA/Post-GIS: Landsat Derived	34	y
LANDSAT	geotif	Band 5 – short wave infrared (MIR)	As Updated	SAGA/Post-GIS: Landsat Derived	34	y
LANDSAT	geotif	Band 6 – thermal infrared	As Updated	SAGA/Post-GIS: Landsat Derived	34	y
LANDSAT	geotif	Band 7 – short wave infrared (MIR)	As Updated	SAGA/Post-GIS: Landsat Derived	34	y
LIDAR	geotif	Canopy Density	As Updated	GDAL, FME	24	y
LIDAR	geotif	Canopy Height	As Updated	GDAL, FME	24	y
Landsat	geotif	Carbonate Difference Ratio	As Updated	SAGA/Post-GIS: Landsat Derived	34	y

Agency	File Type	Data Name/Source	Interval	Tool / Integration Method	Source Size (GB)	DISC Hosted (Y/N)
Landsat	geotif	Clay Difference Ratio	As Updated	SAGA/Post-GIS: Landsat Derived	34	y
USDA	SQL Server Database	Conservation Innovation Grants Database	As Updated	GDAL, FME: potential API in the future	14	y
ELEVA-TION	geotif	Convergence Index	As Updated	SAGA/Post-GIS: Elevation Derived	200	12000
USDA	File Geodatabase	County & State	As Updated	GDAL, FME	14	y
USDA	File Geodatabase	Cropland Data Layer by State	As Updated	GDAL, FME	24	y
USDA	geotif	DEM Derived	As Updated	GDAL, FME	200	y
ELEVA-TION	geotif	Depression Cost Surface	As Updated	GDAL, FME	200	y
USDA	File geodatabase/ shapefile	DSP4SH University Cooperator	As Updated	GDAL, FME	24	y
NRCS	API: json	EDIT	Near Real Time	GDAL, FME	34	n
USDA	geotif	ErMapper Ortho Mosaic by NRCS	As Updated	GDAL, FME	24	y
PRISM	geotif	Evapotranspiration	As Updated	GDAL, FME	24	y
ELEVA-TION	geotif	Ferrous Minerals Difference Ratio	As Updated	SAGA/PostGIS	200	y
ELEVA-TION	geotif	Fuzzy Landform Element Classification	As Updated	SAGA/PostGIS	200	y
LIDAR	geotif	geographic Wetness Index aka Wetness Index aka Compound Topographic Index	As Updated	GDAL, FME	24	y
LANDSAT	geotif	Geology	As Updated	SAGA/Post-GIS: Landsat Derived	34	y
ELEVA-TION	geotif	Geomorphons	As Updated	SAGA/PostGIS	200	y
NRCS	BOX: file geoda-tabase, shapefile, geotif	gSSURGO & gNATSGO	yearly	GDAL, FME	96.5	y
ELEVA-TION	geotif	Heat Load Index	As Updated	SAGA/PostGIS	200	y
LANDSAT	geotif	Iron Difference Ratio	As Updated	SAGA/Post-GIS: Landsat Derived	34	y
USGS/For-est Service	geotif	Landfire (multiple datasets)	As Updated	SAGA/PostGIS	14	n
landsat	geotif	Landsat GeoCover 2000	As Updated	SAGA/Post-GIS: Landsat Derived	34	n
USDA	geotif	Monthly Average Precipitation and Tem-perature by State	As Updated	SAGA/PostGIS		y

Agency	File Type	Data Name/Source	Interval	Tool / Integration Method	Source Size (GB)	DISC Hosted (Y/N)
USDA	geotif	Monthly Average Precipitation and Temperature by State	As Updated	SAGA/PostGIS		y
ELEVATION	geotif	Morphometric Protection Index	As Updated	SAGA/PostGIS	200	y
ELEVATION	geotif	Multiresolution Index of Ridge Top Flatness	As Updated	SAGA/PostGIS	200	y
ELEVATION	geotif	Multiresolution Index of Valley Bottom Flatness	As Updated	SAGA/PostGIS	200	y
NRCS	geotif	Normalized Difference Vegetation Index (NDVI)	As Updated	Post Processing	24	y
ELEVATION	geotif	Normalized Height	As Updated	SAGA/PostGIS	200	y
ELEVATION	geotif	Overland Flow Distance to Network Channel	As Updated	SAGA/PostGIS	200	y
USDA	file geodatabase	PLSS	As Updated	GDAL, FME		y
ELEVATION	geotif	Potential Drainage Density	As Updated	SAGA/PostGIS	200	y
PRISM	geotif	Precipitation	As Updated	SAGA/PostGIS	24	y
NACSE	geotif	Prism	weekly	SAGA/PostGIS	34	y
ELEVATION	geotif	Profile Curvature	As Updated	SAGA/PostGIS	200	y
NRCS	shapefile	Proximal Sensing	As Updated	GDAL	24	y
ELEVATION	geotif	Relief	As Updated	SAGA/PostGIS	200	y
ELEVATION	geotif	Relief Ratio	As Updated	SAGA/PostGIS	200	y
ELEVATION	geotif	Roughness by Relief and Aspect	As Updated	SAGA/PostGIS	200	y
ELEVATION	geotif	Roughness byStandard Deviation of Relief	As Updated	SAGA/PostGIS	200	y
ELEVATION	geotif	SAGA Wetness Index	As Updated	SAGA/PostGIS	200	y
ELEVATION	geotif	Slope Aspect	As Updated	SAGA/PostGIS	200	y
ELEVATION	geotif	Slope Gradient	As Updated	SAGA/PostGIS	200	y
ELEVATION	geotif	Slope Height	As Updated	SAGA/PostGIS	200	y
ELEVATION	geotif	Slope Heterogeneity	As Updated	SAGA/PostGIS	200	y
ELEVATION	geotif	Slope Length	As Updated	SAGA/PostGIS	200	y
NRCS	API: json	Soil Data Access	Near Real Time	GDAL, FME	120	y
ELEVATION	geotif	Solar Radiation	As Updated	SAGA/PostGIS	200	y

Agency	File Type	Data Name/Source	Interval	Tool / Integration Method	Source Size (GB)	DISC Hosted (Y/N)
ELEVA-TION	geotif	Stream Power Index	As Updated	SAGA/PostGIS	200	y
NRCS	file geodatabase	Subcounty	Quarterly	GDAL, FME	14	y
ELEVA-TION	geotif	Tangential Curvature	As Updated	SAGA/PostGIS	200	y
ELEVA-TION	geotif	Terrain Surface Classification	As Updated	SAGA/PostGIS	200	y
ELEVA-TION	geotif	Terrain Surface Texture	As Updated	SAGA/PostGIS	200	y
USDA	geotif	US Pacific Basin including Hawaii imagery	As Updated	SAGA/PostGIS		y
ELEVA-TION	geotif	Valley Depth	As Updated	SAGA/PostGIS	200	y
ELEVA-TION	geotif	Vertical Distance to Channel	As Updated	SAGA/PostGIS	200	y
PRISM	geotif	Water Balance aka Effective Precipitation	As Updated	SAGA/PostGIS	24	y
USDA	geotif, file geodata-bases, API	Models and data derived from DSP HUB		GDAL, FME		y
USDA	SQL Server Database	Pedon Data Mart		GDAL, FME	133	y
NRCS	SQL Server Database	plants	Yearly	GDAL, FME	9	y
NRCS	File geodatabase	NCSS_LDM & SSL_Repo	Near Real Time	GDAL, FME	3	y
USDA	geotif	Digital Ortho County Mosaic of 7.5' quads by NRCS	As Updated	SAGA/PostGIS		y
USDA	geotif	Digital Ortho HiRes Mosaic-Color Infrared - Most Recent	As Updated	SAGA/PostGIS		y
USDA	geotif	Digital Ortho HiRes Mosaic-Natural Color - Most Recent	As Updated	SAGA/PostGIS		y
USDA	geotif	Digital Ortho Quad County Mosaic-Color Infrared	As Updated	SAGA/PostGIS		y
USDA	geotif	Digital Ortho Quad County Mosaic-Natural Color	As Updated	SAGA/PostGIS		y
USDA	geotif	Digital Raster Graphic County Mosaic by NRCS	As Updated	SAGA/PostGIS		y
USDA	geotif	DOQ Multi-County Mosaic by NRCS	As Updated	SAGA/PostGIS		y
USDA	geotif, shapefile, file geodatabase, API	Ad Hoc Data Sources	As Updated	SAGA/PostGIS/ GDAL		y
EARTH	geotif	Consensus Land Cover	As Updated	SAGA/PostGIS		n
USDA	geotif, shapefile, file geodatabase, API	Data Science User Data Sets	As Updated	SAGA/PostGIS/ GDAL		y
USDA	geotif	Daymet	As Updated	SAGA/PostGIS	5000	n
ELEVA-TION	geotif	Downslope Distance Gradient	As Updated	SAGA/PostGIS	200	y
USDA	SQL Server database	Economic Datasets	As Updated	GDAL, FME		y
ELEVA-TION	geotif	Elevation	As Updated	SAGA/PostGIS	200	n

Agency	File Type	Data Name/Source	Interval	Tool / Integration Method	Source Size (GB)	DISC Hosted (Y/N)
USDA	geotif	Enhance Digital Raster Graphic Quadrangles	As Updated	SAGA/PostGIS		y
USDA	SQL Server Database	Geochemistry Scan Data	As Updated	GDAL, FME		y
USDA	geotif	Geocue Lidar	As Updated	SAGA/PostGIS/ GDAL		y
USDA	file gerodatabase	Henry Mount Database	As Updated	GDAL, FME	24	y
USGS	geotif	Lidar	As Updated	SAGA/PostGIS/ GDAL	34	n
USDA	geotif/database	MIR Data	As Updated	SAGA/PostGIS/ GDAL	10	y
USDA	geotif	NAIP Imagery	As Updated	GDAL, FME		y
USDA	geotif	National Ag. Imagery Program County Mosaic	As Updated	GDAL, FME		y
NASS	geotif, shapefile	National Agricultural Statistics Service	As Updated	SAGA/PostGIS/ GDAL		y
USDA	json	National Historic Places	As Updated	GDAL, FME	34	y
USDA	geotif	National Land Cover Dataset by State	As Updated	GDAL, FME		y
USDA	json	National Wetland Inventory Vers2	As Updated	GDAL, FME		y
USDA	geotif, file geodatabase	NRCS Portal Online	As Updated	GDAL, FME		y
USDA	SQL Server Database	On Farm Soil Health Demo Data	As Updated	GDAL, FME		y
USDA	shapefile	Plant Hardiness Zones	As Updated	GDAL, FME		y
USDA	geotif, file geodatabase	pXFIR Data	As Updated	GDAL, FME		y
IEM	geotif	Radar	Monthly	GDAL, FME	24	n
USDA	geotif	Raster Soil Surveys	As Updated	GDAL, FME		y
USDA	geotif	Remote Sensing Imagery Data	As Updated	GDAL, FME		y
LANDSAT	geotif	Rock Outcrop Difference Ratio	As Updated	GDAL, FME		y
NRCS	geotif, file geodatabase	SNOWTEL	Near Real Time	GDAL, FME		y
NRCS	API: geotif, json	Soil Climate Analysis Network (SCAN)	Near Real Time	GDAL, FME		y
USDA	shapefile	Soil Monitoring Data	As Updated	GDAL, FME		y
USDA	shapefile	Soil Temperature Data	As Updated	GDAL, FME		y
USGS	gerotif, file geodatabase	The National Map - Data Delivery	As Updated	GDAL, FME	34	n
USDA	file geodatabase	TIGER County Social, Econ, Housing Stats by State	As Updated	GDAL, FME		y
LANDSAT	geotif	Topographic Position Index	As Updated	Post Processing		y
EARTH	geotif	Topography	As Updated	GDAL, FME	104	n
USDA	file geodatabase	USGS National Hydrography Database (NHD)	As Updated	GDAL, FME		y
USDA	file geodatabase	Watershed Boundary Dataset and Lines for HUC2-12	As Updated	GDAL, FME		y

Agency	File Type	Data Name/Source	Interval	Tool / Integration Method	Source Size (GB)	DISC Hosted (Y/N)
NOAA	geotif	(NOAA's Climate Data)	Near Real Time	Raster2PGSQL, GDAL	0	n
USDA	SQL Server Database	ABS Benefits Finder	As Updated	Post Processing, Need to provide API for PI13		y
USDA	File geodatabase: geotif	ArcGIS Online	As Updated	API, Crunchy Bridge, Raster2PGSQL, GDAL, Post Processing	34	y
USDA	SQL Server geodatabase	Edge of Field Monitoring Data	As Updated	GDAL, FME		y
USDA	API:json	EPA Mandatory Class 1 Federal A	As Updated	GDAL, FME		y
EARTH	geotif	Global Forest Watch	As Updated	Raster2PGSQL, GDAL	34	n
EARTH	geotif	Habitat Heterogeneity	As Updated	Raster2PGSQL, GDAL	34	n
USDA	API:json	Habitat Threatened/Endangered Species USFWS ECOS	As Updated	GDAL, FME		y
USDA	API:json	Protected Areas Database	As Updated	GDAL, FME		y
USDA	API: geotif	Radar Web Service	As Updated	Raster2PGSQL, GDAL		y
USDA	SQL Server Database	VINR Data	As Updated	API, Crunchy Bridge, Raster2PGSQL, GDAL, Post Processing		y
USDA	geotif	Water and Climate Data	As Updated	API, Crunchy Bridge, Raster2PGSQL, GDAL, Post Processing		y
NRCS		Baflog & baflogrpt	Daily	GDAL, FME	197	y
EARTH	geotif	Cloud cover climatology	As Updated	Raster2PGSQL, GDAL	15	n
EARTH	geotif	90m digital elevation model	As Updated	Raster2PGSQL, GDAL	34	n
USDA	API: feature server	Coastal Zone Management Counties	As Updated	GDAL, FME		y
USDA	API: feature server	Addressing Easement Ranking National Questions	As Updated	GDAL, FME		y
USDA	geotif	Alaska IFSAR (Interferometric Synthetic Aperture Radar)	As Updated	Raster2PGSQL, GDAL		y
USCB	json	Congressional District	As Updated	GDAL, FME	0	y
USDA	API: feature server	Coral Reefs	As Updated	GDAL, FME		y
USDA	API: feature server	EPA Ozone	As Updated	GDAL, FME		y
USDA	json	Flood Hazards	As Updated	GDAL, FME		y

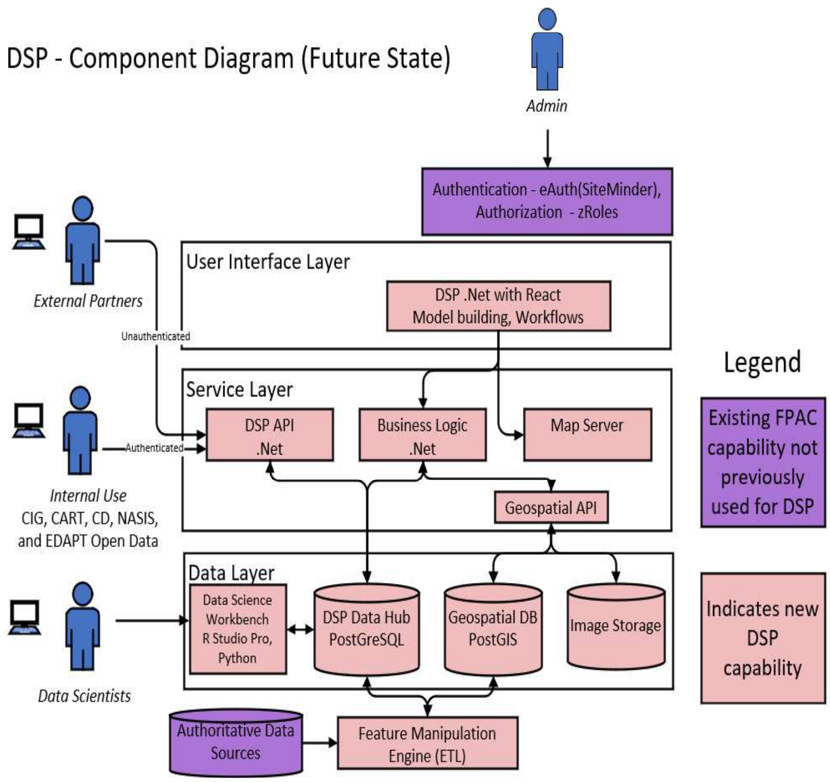
Agency	File Type	Data Name/Source	Interval	Tool / Integration Method	Source Size (GB)	DISC Hosted (Y/N)
EARTH	ascii text	Freshwater environmental variables	As Updated	Raster2PGSQL, GDAL		y
USDA	file geodatabase	IGER State and County Demographic Statistics by State	As Updated	GDAL, FME		y
MRLC	geotif	Multi-Resolution Land Characteristics Consortium	As Updated	GDAL, FME	404	n
USDA	geotif	Soil Color Raster Maps	As Updated	GDAL, FME		y
USGS	geotif	USGS Moderate Resolution Imaging Spectroradiometer (MODIS)	As Updated	GDAL, FME		y
WC	geotif	WorldClim	As Updated	GDAL, FME	0	n

APPENDIX D—RISKS AND MITIGATION STRATEGY

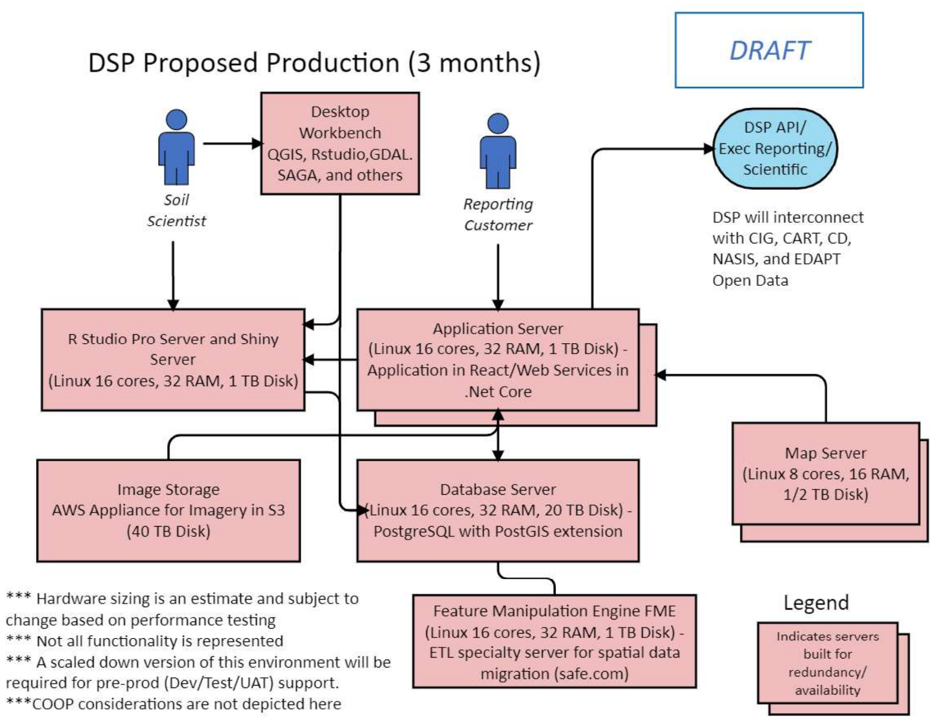
Risk	Description	Risk Mitigation Strategy
Culture shift to SAFe Agile	FPAC is early in the transition to an iterative, agile approach in software development.	Trainings and open communication, along with highlighting successes in the project in embracing agile design principles.
Limited budget and scope creep	The vision for the DSP Hub is a flagship innovation project which can create unrealistic timelines and expectations. Scope creep can be a concern.	Adopt key pillars of scaled agile design such as fast-fail, pivot without mercy, cross-team collaboration, and relentless improvement. These principles help to fail-proof and grow the project on a strong foundation.
Alignment to USDA data and analytics program	EDAPT is the required platform for data analytics and data warehouses	Design and test the environment in small MVP to develop required, specific functionality for EDAPT and migrate at earliest opportunity. Align the ontologies and data schemas for all USDA conservation program data during development activities for data integration and interoperability.
Latency and performance	Time or ability to process multiple large files	Keep the research environment near the important data sources such as NASIS, CD, Lidar and migrate the DSP Hub in tandem when going to a Cloud Service Provider (CSP).
Size of the data sets needed for modeling	The data needed for modeling are large geospatial files. need work to derive working sets more applicable to model building and refinement.	Prioritize the data sets needed for modeling to justifiable list. Justification and approval of larger original and working data sets will be required.
Limited experience in FPAC with geospatial modeling software and design	To perform the science-based geospatial modeling needed, the DSP Hub will need PostgreSQL and PostGIS.	External support services are proposed to assist with architectural analysis (Crunchy Data and other specialized services as needed).
Software that can perform science-based terrain analysis and other tools	There are specific and non-standard needs for software to perform the scientific analysis needed (terrain analysis).	Many of the non-standard software options are open source. Approvals will be required with demonstrated needs.
Time to procure and stand up software and hardware	Often, a significant lead time is needed to establish a working environment with all the licenses and approvals.	There are many open source options that are in testing on DISC. Additional approvals and assistance from ISD can be provided to expedite these efforts.
Proper data governance	Advanced data governance will be needed to support the DSP potential relationship to a carbon bank.	The National Cooperative Soil Survey has a robust data governance program and FGDC standards that can be used as a foundation. Start small with an MVP to establish the structure and process for data governance.
Technical debt during development	Since the EDAPT environment cannot currently support this documents use cases, there will be investments in on-premise development work that need to take place to meet the Farm Bill mandate deadline.	The investment to build on-premise with the intention to migrate will need to be evaluated by NRCS leadership with transparent costs and timelines. The value of the development work on-premise is not a sunk cost since the design and testing of both the platform and the scientific models will add value to USDA EDAPT.
USDA analytics platform is not currently able to meet the DSP Hub needs	EDAPT currently cannot meet the needs of the DSP Hub.	Start the agile design and testing on premise to finalize the architecture (structures and processes) to provide certainty when EDAPT is ready to start development on DSP Hub needs.
Time series data growing beyond estimated capacity	Developing time-series (quarterly or monthly repeatable reports) could grow quickly	Identify the size requirements and potential growth and obtain approvals for additional space as needed. Keep architecture open to include a Data Historian (data historians a commonly used databases excelling in storage and reporting of time-series data).
Lack of data management protocols	There are currently no data management and curation protocols in place for conservation practice data, however there are extensive data management protocols in place for soil data.	Start very small and create scalable protocols that are linked to the zRoles and workflow. Use existing best practices, and at a minimum maintain documentation for data lineages to support audits.

Risk	Description	Risk Mitigation Strategy
Lack of expertise in Development Support Operations (DSO) and Production Support Operations (PSO)	Some of the software needed for science-based terrain analysis and modeling are outside the expertise of DSO/PSO.	Seeking external expertise to expand capacity. DSO/PSO has expressed interest in partnership. Procurement of additional support is a very minor expense ($50K/yr).
Wait time or inability to load new datasets	The EDAPT environment has limited permissions that can cripple the ability to rapidly load and analyze new data sets.	Provide tools such as Safe Software FME, Red Gate Toolbelt, and GDAL (Geospatial Data Abstraction Library - open source) for ingestion of datasets into the DSP Hub.
Processing feasibility between the DSP Hub and transactional systems	Results are returned to client at the Field Office in real time (real time is “fast enough”).	Design and test processes in a small scalable way to address limitations iteratively.
APIs accessibility and functionality	The DSP Hub owners will need access to internal servers to develop data services for customers.	Use a zRoles permissions-based system instead of a manual system of administrative controls.
Cloud Service Provider Egress charges	There are egress charges for data leaving cloud service providers which could become prohibitive.	Start development on premise to ensure the design and processes are established to ensure the egress charges with the cloud environment are budgeted correctly.
The final architectural plan is not complete	There is concern that the DSP Hub does not have a complete architectural roadmap, partially due to a culture of waterfall methods.	DSP Hub is using a SAFe agile approach with architecture established for the MVPs. Scaling up can be done in partnership with key stakeholders with a transparent approach to budgeting and operations so that data-driven decisions can be made by accountable officials.
Cost overruns	The current budget is not adequate to realize the current vision.	DSP Hub will be built out in an agile way to ensure that value is returned, and scaling is strategic.

APPENDIX E—DSP HUB COMPONENT DIAGRAM



APPENDIX F—DSP HUB PROPOSED PRODUCTION DIAGRAM



APPENDIX G—DSP HUB TECHNICAL STACK (SOFTWARE LISTING)

Tech Stack	Description
R	R is a language and environment for statistical computing and graphics. ... R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible.
GRASS	Geographic Resources Analysis Support System - GRASS Open Source GIS
QGIS	QGIS is a free and open-source cross-platform desktop geographic information system application that supports viewing, editing, and analysis of geospatial data.
Putty	PuTTY is an SSH and telnet client
pgAdmin	Open Source administration and development platform for PostgreSQL
Dbeaver	Free universal DB manager
MobaXterm	SSH Windows client
FME Pro	Spatial ETL application
React	A JavaScript Library for building user interfaces
.NET Core	Cross-platform .NET framework
Red Hat Enterprise Linux 8	Operating System
PostgreSQL	PostgreSQL is a powerful, open source object-relational database
MapServer	MapServer is an Open Source platform for publishing spatial data and interactive mapping applications to the web.
Leaflet	Open Source Web Mapping
Google Earth Engine	Google Earth Engine combines a multi-petabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities and makes it available for scientists, researchers, and developers to detect changes, map trends, and quantify differences on the Earth's surface.
Python3	Python is a programing language that contains libraries for storing , manipulating and gaining insight from data.
Drools	Drools is a business-rule management system with a forward-chaining and backward-chaining inference-based rules engine, allowing fast and reliable evaluation of business rules and complex event processing.
Informatica	Informatica offers a suite of tools for uses including data governance, data quality, and Master Data Management.
GDAL	Geospatial Data Abstraction Library is a computer software library for reading and writing raster and vector geospatial data formats, and is released under the permissive X/MIT style free software license by the Open Source Geospatial Foundation.
PostGIS	PostGIS is a spatial database extender for PostgreSQL object-relational database. It adds support for geographic objects allowing location queries to be run in SQL.
SAGA	System for automated geoscientific analysis
RStudio Pro & Rstudio Connect Server	Data analysis and integrated development environment
Shiny Server	Works with Rstudio to create shiny web application

APPENDIX H—RISK ANALYSIS MODEL

Key		
Risk	Neutral	Benefit
25	50	100

Categories		DSP Data Hub Risks and Opportunities Matrix					
		Develop DSP Data Hub MVP in DISC - available immediately		Develop DSP Data Hub MVP in EDAPT, projected availability FY23		Develop DSP Data Hub in scaled agile way with phased/selected migration from DISC to AWS	
Timeline	Impact Weight	Immediate	Impact & Benefit	FY23	Impact & Benefit	Phasing begins immediately	Impact & Benefit
Policy and Governance value 45							
Congressional Statutory Mandate (2018 Farm Bill)	100	This alternative works towards meeting the 2018 Farm Bill mandate.	13.6	Delay in implementing the 2018 Farm Bill mandates	3.4	This alternative works towards meeting the 2018 Farm Bill mandate.	13.6
Presidential and Adminstrative priorities	80	This alternative works towards meeting requirements as currently known	10.9	This alternative DOES NOT work towards meeting requirements as currently known	2.7	This alternative works towards meeting requirements as currently known	10.9
Governance and Data Lineage	70	Governance processes and data lineage are developed during implementation. Soil scientists are aware of and follow governance and auditing requirements.	4.8	More time is afforded to design data governance processes and the lineage to support auditability. While the planning alleviates risk, this does detract from time necessary to meet the congressional mandate.	4.8	Governance processes and data lineage are developed during implementation. Soil scientists are aware of and follow governance and auditing requirements.	4.8
Hardware procurements, ATO, and Security	80	Hardware procurements, ATO, and Security authorizations are proceding interatively and in the same manner as a number of current solution train projects.The DSP solution stack is well defined and similar to EDAPT.	5.5	More time is afforded to perform procurements, ATO's, and understand security implementation before further development continues. While the planning alleviates risk, this does detract from time necessary to meet the congressional mandate.	5.5	Hardware procurements, ATO, and Security authorizations are proceding interatively and in the same manner as a number of current solution train projects.The DSP solution stack is well defined and similar to EDAPT.	5.5
Policy and Governance Sum	**330**		**35**		**16**		**35**
Customer Value 25							
Ability to provide DSP data (i.e. soil carbon, advanced benefits conservation results) to external customers	100	This alternative works towards meeting the 2018 Farm Bill mandate.	8.3	Delay in providing valuable information to external customers	2.1	This alternative works towards meeting the 2018 Farm Bill mandate.	8.3
Increased efficiency	100	In the past 10 years the number of soil scientist have gone down. The limited amount of people could have a huge impact and cut down by 40-50 percent	8.3	Delay in increased	2.1	In the past 10 years the number of soil scientist have gone down. The limited amount of people could have a huge impact and cut down by 40-50 percent	8.3
Ability for staff to access data to to their job	100	Existing USDA data sets are often on outdated technology, siloed databases and applications, and even on individual desktops.	8.3	Would be a roadbloack in staff doing their job	2.1	This alternative works towards meeting the 2018 Farm Bill mandate.	8.3
Customer Sum	**300**		**25**		**6**		**25**
Costs Value 5							
Estimated solution costs	80	Limited initial expenditures already incurred for DISC services, labor and development work may be leveraged if development continues at DISC otherwise it could not be leveraged, the rest of the solution costs would be on hold.	1.3	The cost to implement the alternative would be significant based on current vision. Limited initial expenditures already incurred for DISC services, labor and development work could be leveraged.	0.6	The cost to implement the alternative would be significant based on current vision. Limited initial expenditures already incurred for DISC services, labor and development work could not be leveraged.	1.3
Hidden charges (e.g. internet)	40	N/A	1.3	Internet Fees coming out of disc	0.3	Inernet fees coming out of DISC are comparable to EDAPT if data was still in DISC	0.6
Egress and Ingress fees	40	N/A	1.3	Egress fees coming out of DISC are comparable to EDAPT.	0.3	Egress fees coming out of DISC are comparable to EDAPT if data was still in DISC	0.6
Cost Sum	**160**		**4**		**1**		**3**
Architecture Value 5							
Architectural Roadmap	100	Time is afforded to decide on an architecture roadmap to mitigate potential risks related to RPO, RTO, availability, performance, patching, support, and CI/CD integration.	1.1	Integrating into CI/CD pipeline would be the responsibilities of the developers.	0.5	Time is afforded to decide on an architecture roadmap to mitigate potential risks related to RPO, RTO, availability, performance, patching, support, and CI/CD integration.	1.1
Architectural Guardrails	90	Affords time to continue work on Architectural Guardrails.	1.0	Architectural Guardrails for the project are incomplete so the project could over-investwithout meeting objectives.	0.2	Affords time to continue work on Architectural Guardrails.	1.0
Technical Debt	100	More time is afforded to architectural design to mitigate technical debt if a potential challenge. While the planning alleviates risk, this does detract from time necessary to meet the congressional mandate.	0.5	Technology developed for EDAPT not be innately designed for optimization in the cloud and would accrue technical debt.	0.3	More time is afforded to architectural design to mitigate technical debt if a potential challenge. While the planning alleviates risk, this does detract from time necessary to meet the congressional mandate.	0.5
Scalability	80	Hardware would not be the bottleneck for results. The scalability of the system would come down to architectural design and cost.	0.4	Hardware would not be the bottleneck for results. The scalability of the system would come down to architectural design and cost.	0.4	Hardware would not be the bottleneck for results. The scalability of the system would come down to architectural design and cost.	0.4
SAFe Adherence	100	Adhere to SAFe protocols. Agile	1.1	Waterfall Method	0.3	Adhere to SAFe protocols. Agile	1.1
Architecture Value Sum	**470**		**4**		**2**		**4**
Operational Feasibility 20							
Latency for geospatial processing	100	Little latency since data is on premise.	3.7	Increased latency.	0.9	Little latency sice data is on premise.	3.7
Administrative controls on adding data	80	Administrative controls for DSP Hub owners requires rapid development of new data products from a variety of partial and legacy data sources	3.0	Administrative controls hamper the ability to be flexible	0.7	Administrative controls for DSP Hub owners requires rapid development of new data products from a variety of partial and legacy data sources	3.0
Live-time transactional processing	60	Transactional workflow: ability to interact in real time with data coming in from various connections (CART, CD, NASIS)	2.2	Transaction workflow	0.6	Transactional workflow: ability to interact in real time with data coming in from various connections (CART, CD, NASIS) if all data is in AWS	1.1
Import/export to partners	50	Ability to export and import large amounts of data to partners such as universities and other agencies outside of USDA	0.9	Controls might limits large amounts of data being imported and exported	0.5	Ability to export and import large amounts of data to partners such as universities and other agencies outside of USDA	1.9
Geospatial analysis and functions (science-based tools for terrain analysis)	80	Advanced geospatial functions for analysis and modeling for terrain analysis: ability to develop, test, and maintain landscape-based statistical and scientific models that interact with a variety of existing and new transactional data with a powerful and scalable geospatial engine (massive data sets, complex models, quick processing, big data analytics)	3.0	No plan for highend geopsatial	0.7	Advanced geospatial functions for analysis and modeling for terrain analysis: ability to develop, test, and maintain landscape-based statistical and scientific models that interact with a variety of existing and new transactional data with a powerful and scalable geospatial engine (massive data sets, complex models, quick processing, big data analytics)	3.0
Rework	70	medium risk of reworking development	1.3	High risk of reworking development	0.6	Low risk of reworking development	2.6
Compatible with PostGRES SQL/SAGA	100	Linux server compatible with open source version Postgres	3.7	Not compatible with	0.9	Aurora pretending to be PostgreSQL but using a different database engine behind the scenes.	1.9
Operational Feasibility Sum	**540**		**18**		**5**		**17**
Totals			**57**		**23**		**56**

APPROVAL PAGE STATUS

Darren Ash Darren Ash (Jun 7, 2021 09:15 EDT)	Jun 7, 2021
Darren Ash, Assistant Chief Information Officer	Date
Paul Chevalier	Jun 1, 2021
Paul Chevalier, Assistant Chief Data Officer	Date
Jimmy Bramblett Jimmy Bramblett (Jun 1, 2021 04:53 EDT)	Jun 1, 2021
Jimmy Bramblett, NRCS, Deputy Chief for Programs, NHQ	Date
Luis M. Tupas Luis M. Tupas (Jun 1, 2021 07:28 EDT)	Jun 1, 2021
Luis Tupas, NRCS, Deputy Chief for Soil Science & Resource Assessment, NHQ	Date