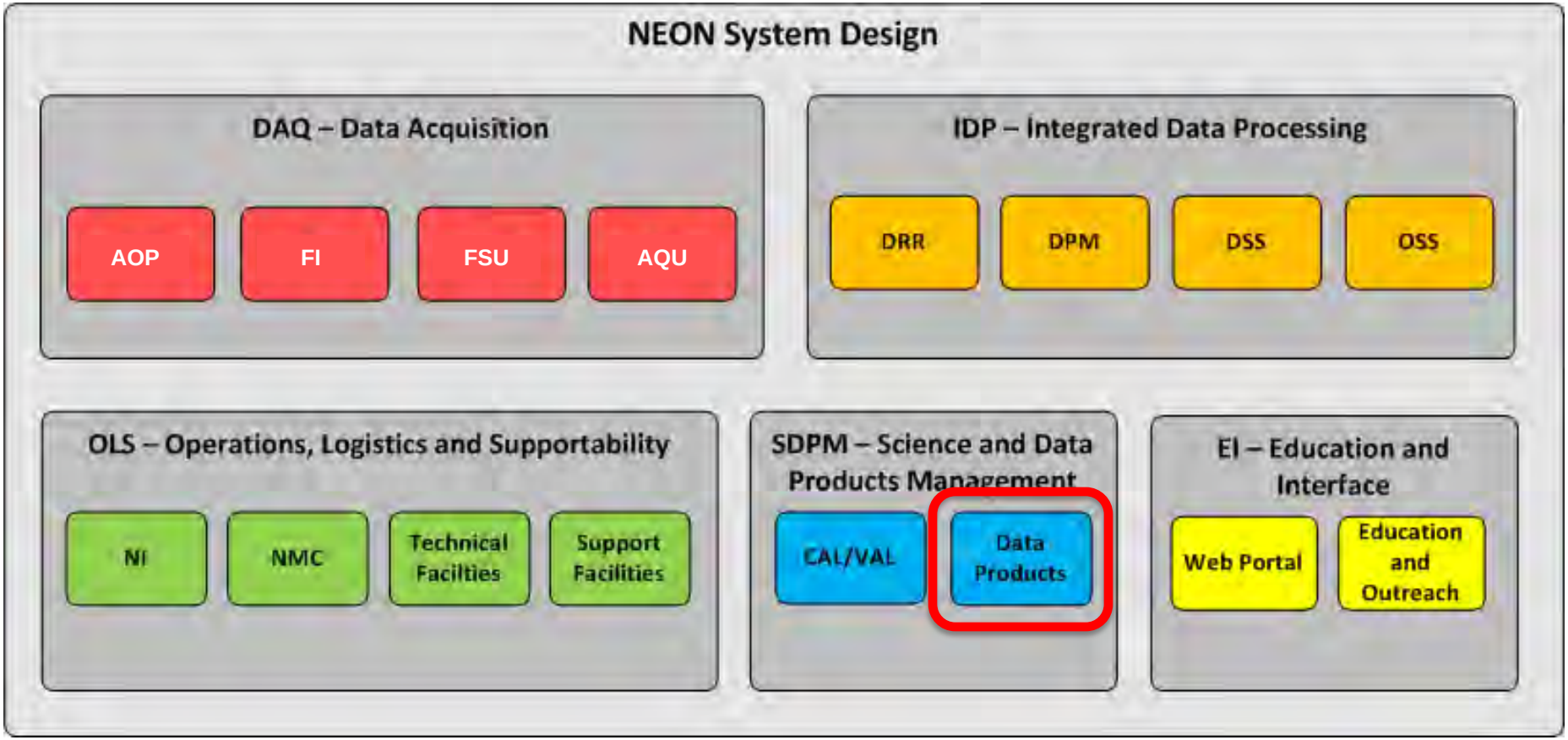
National Ecological Observatory Network

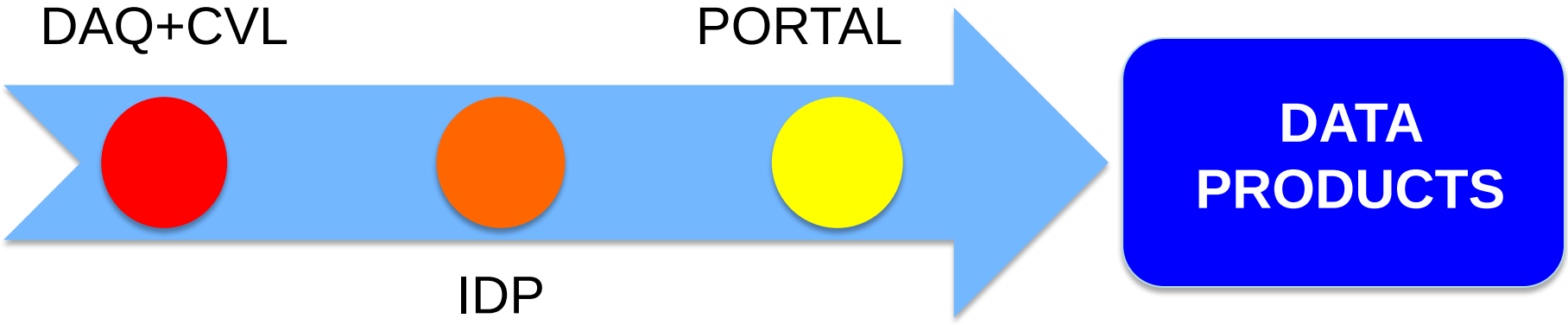
DATA PRODUCTS

Michael Keller & NEON Project Team

DATA PRODUCTS

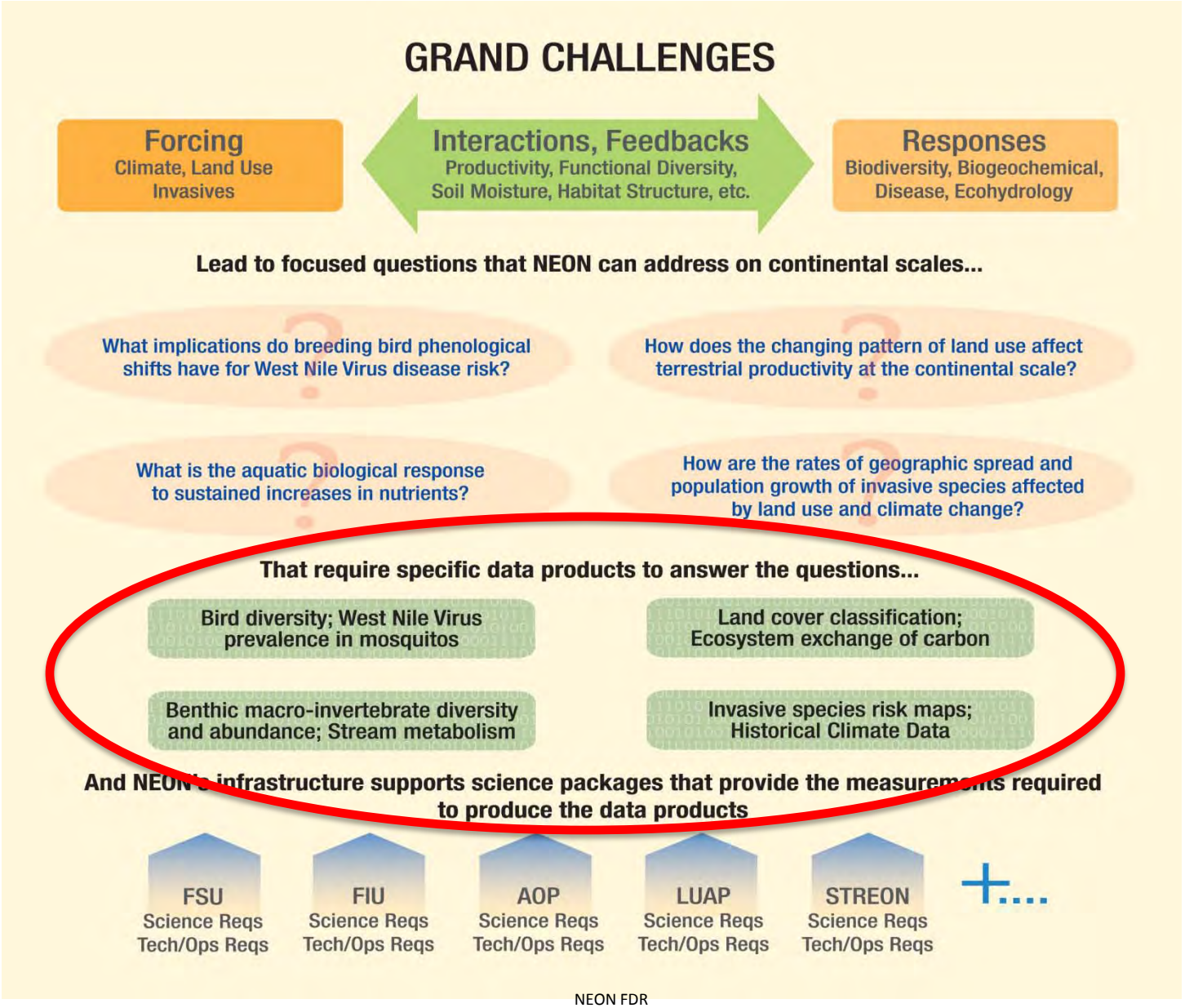


DATA PRODUCTS: FUNDAMENTAL SCIENCE DELIVERABLES



✓The infrastructure of the observatory is designed to support the data products.

Role of Data Products



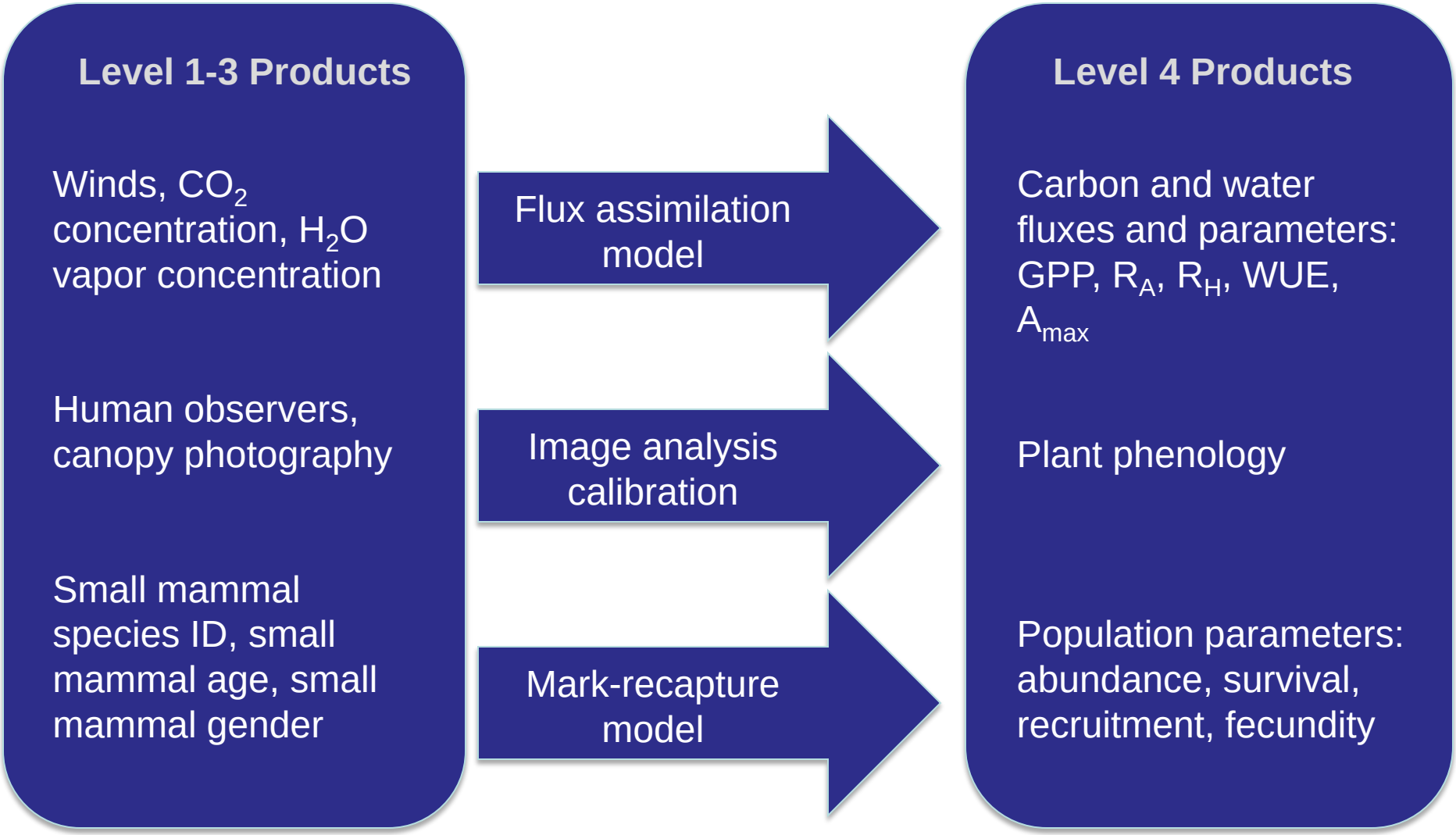
Organization of Data Products

- Data Product Levels for NEON. This data product organization is based on/consistent with the CODMAC standard.

Level 0	Raw data from instrumental or human observations.
Level 1	Calibrated data generally from a single instrument, observer, or field sampling area. These data may include information on data quality.
Level 2	Combinations of level 1 data used to create a gap filled data stream that may replace a level 1 product. Generally, products at this level this will reflect a stream from a single instrument, observer, or field sampling area. Annotations will indicate the gap filling approach employed.
Level 3	Level 1 and /or 2 data mapped on a uniform space-time grid.
Level 4	Derived products using levels 1, 2 and/or 3 data. Products at this level may combine observations from more than one instrument, observer, and/or sampling area.

• Level 0, 1, 2, & 3 data products are associated with a specific science sub-systems (FSU, Aquatic/STREON, FIU, AOP, LUAP)

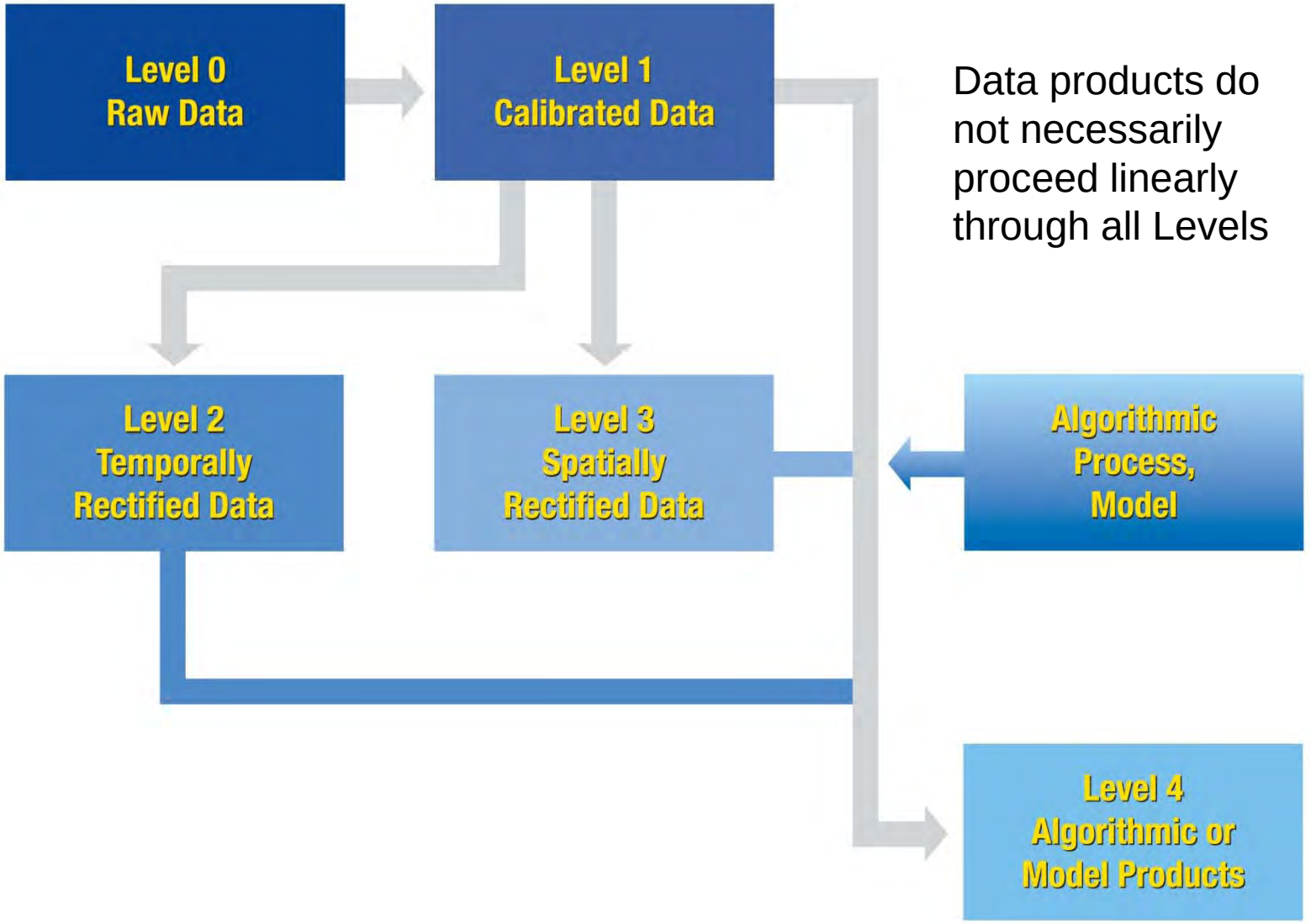
Low and High Level Data Products



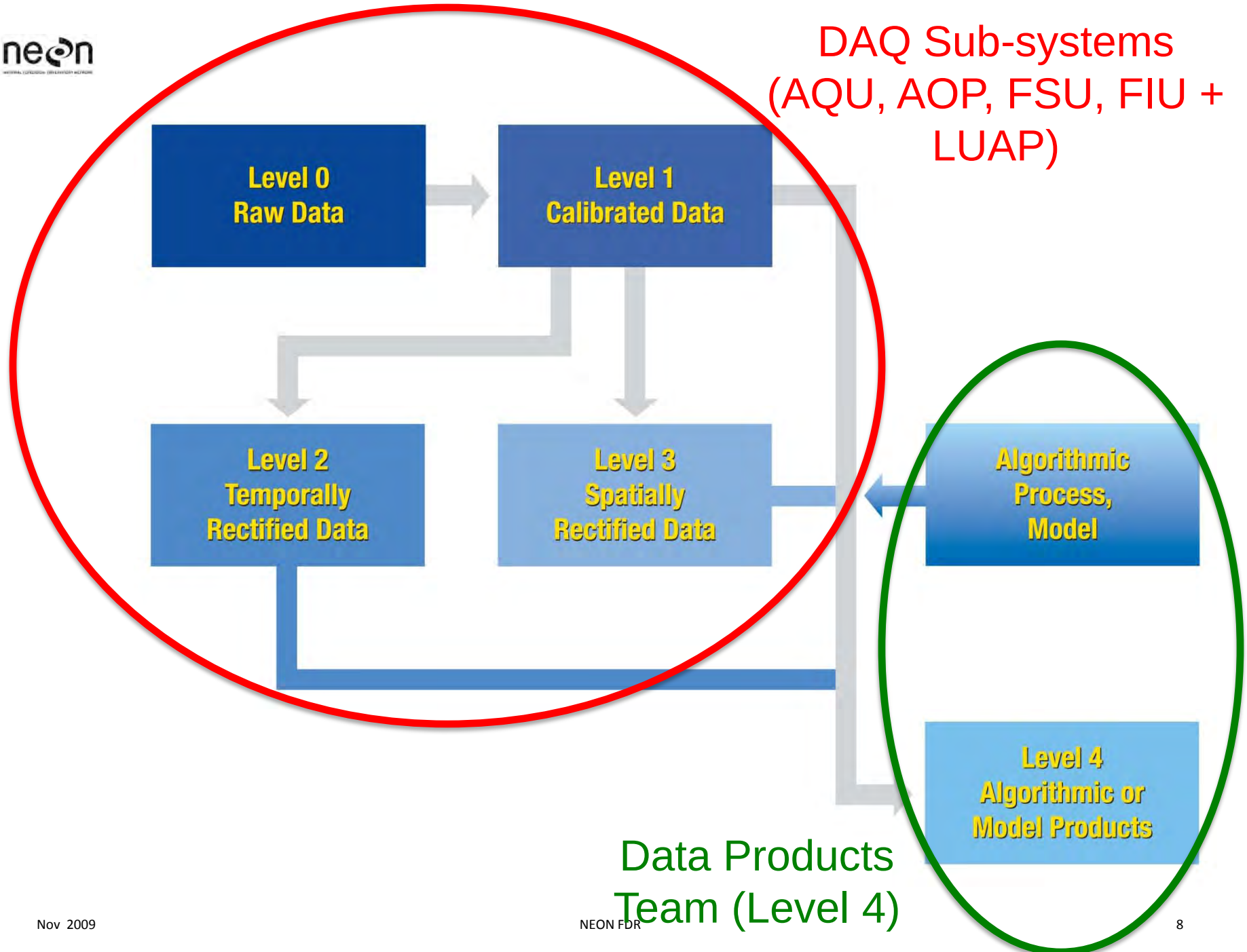
584 L1 products + 35 L2-L3 products

117 L4 products

Data Products Flow



Data products do not necessarily proceed linearly through all Levels


DAQ Sub-systems (AQU, AOP, FSU, FIU + LUAP)
Level 0 Raw Data
Level 1 Calibrated Data
Level 2 Temporally Rectified Data
Level 3 Spatially Rectified Data
Algorithmic Process, Model
Level 4 Algorithmic or Model Products
Data Products Team (Level 4)

Data Product Traceability Completed

From	To	Document
Level 0	Level 1	Level 0 Data Products Catalog
Level 1	Levels 2 or 3	Level 1-3 Data Products Catalog
Level 1	Level 4	NEON L1 to L4 Traceability Matrices (68,328 cell matrix also displayed as a poster)
Level 4	Specific Questions	Informal poster

Science Case 1- Infectious Disease

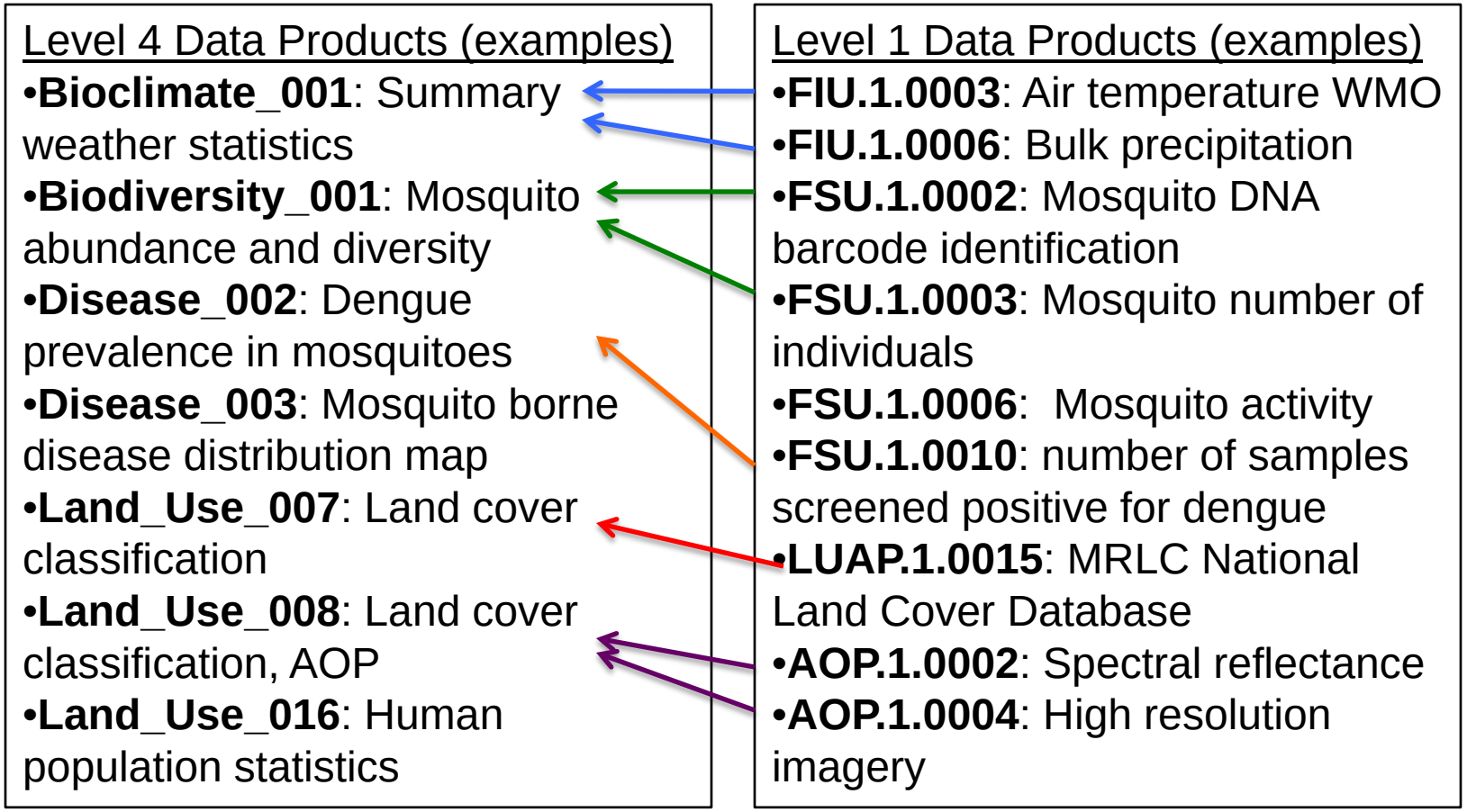
Question: How will climate change affect the spread of mosquito-borne diseases? The specific case of dengue fever.
Approach: Climate envelop model for Aedes aegypti (vector for dengue fever)

Level 4 Data Products (examples)
- **Bioclimate_001**: Summary weather statistics
- **Biodiversity_001**: Mosquito abundance and diversity
- **Disease_002**: Dengue prevalence in mosquitoes
- **Disease_003**: Mosquito borne disease distribution map
- **Land_Use_007**: Land cover classification
- **Land_Use_008**: Land cover classification, AOP
- **Land_Use_016**: Human population statistics

Level 1 Data Products (examples)
- **FIU.1.0003**: Air temperature WMO
- **FIU.1.0006**: Bulk precipitation
- **FSU.1.0002**: Mosquito DNA barcode identification
- **FSU.1.0003**: Mosquito number of individuals
- **FSU.1.0006**: Mosquito activity
- **FSU.1.0010**: number of samples screened positive for dengue
- **LUAP.1.0015**: MRLC National Land Cover Database
- **AOP.1.0002**: Spectral reflectance
- **AOP.1.0004**: High resolution imagery

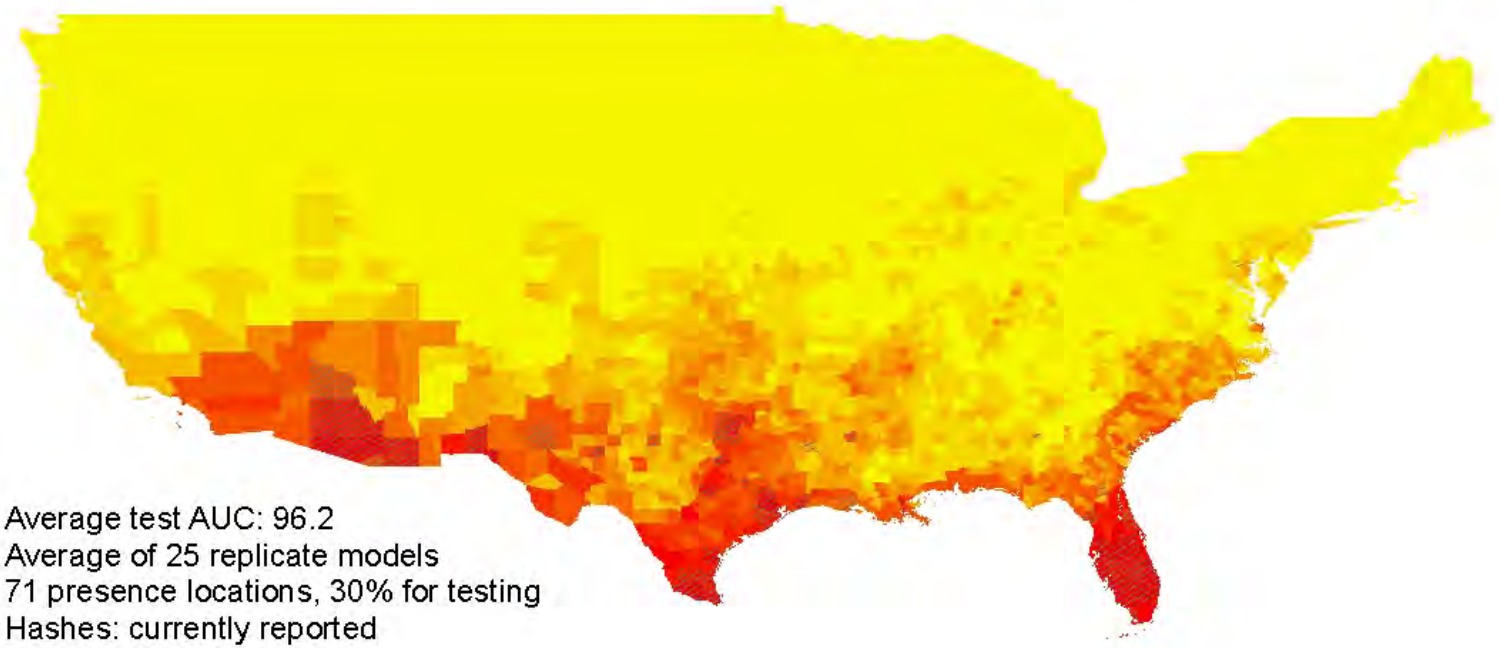
Science Case 1- Infectious Disease

Question: How will climate change affect the spread of mosquito-borne diseases? The specific case of dengue fever.
Approach: Climate envelop model for Aedes aegypti (vector for dengue fever)



Science Case 1 – Infectious Disease

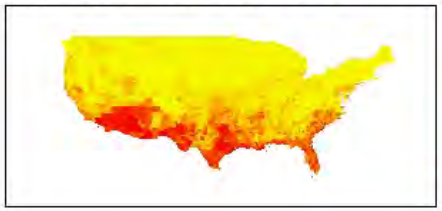
County habitat suitability for Aedes aegypti



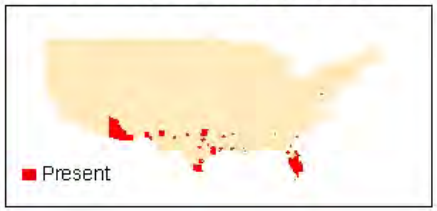
Average test AUC: 96.2
Average of 25 replicate models
71 presence locations, 30% for testing
Hashes: currently reported

Created with: Maxent v 3.2.19

Replicate model deviation



Current reported distribution



Greatest variable contribution (avg):
Annual mean temperature (54.2%)
Urban area (14%)
Mean temperature of wettest quarter (10.9%)
Mean temperature of coldest quarter (10.7%)

T. Stohlgren, unpublished

Science Case 2- Biogeochemistry

Question: How will climate change affect US ecosystem carbon uptake?
Approach: Carbon data-assimilation model

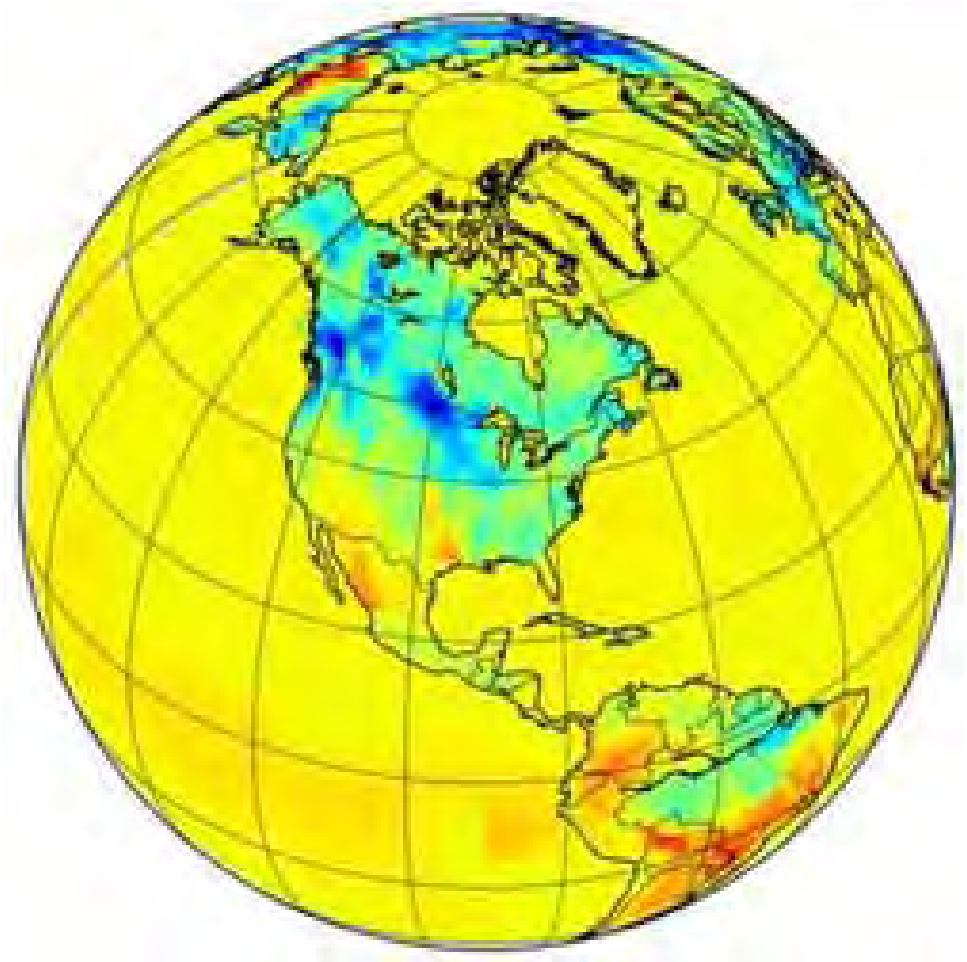
Level 4 Data Products (examples)

- **Bioclimate_001**: Summary weather statistics
- **Biogeochemistry_017**: Ecosystem exchange, tower
- **Biogeochemistry_019**: Ecosystem exchange of carbon, NEON Realm
- **Biogeochemistry_020**: Net Primary Productivity
- **Land_Use_007**: Land cover classification
- **Land_Use_008**: Land cover classification, AOP
- **Land_Use_024**: Historical Climate Data

Level 1 Data Products (examples)

- **FIU.1.0003**: Air temperature WMO
- **FIU.1.0006**: Bulk precipitation
- **FIU.1.0010**: CO_2 concentration
- **FIU.1.0011**: CO_2 profile
- **FSU.1.0112**: DBH live trees
- **FSU.1.0116**: Litter traps, leaves
- **FSU.1.0125**: Coarse root live biomass
- **FSU.1.0142**: Coarse downed woody debris
- **LUAP.1.0015**: MRLC National Land Cover Database
- **AOP.1.0002**: Spectral reflectance
- **AOP.1.0004**: High resolution imagery

Science Case 2- Biogeochemistry



Results from NOAA Carbon Tracker (displayed) present estimates of carbon uptake by ecosystems based only on atmospheric data. NEON will complement this with a data-assimilation model based on terrestrial measurements.

Science Case 3- Biodiversity

Question: How will changes in exurban land use affect US biodiversity?
Approach: Statistical analysis of comparative sites

Level 4 Data Products (examples)

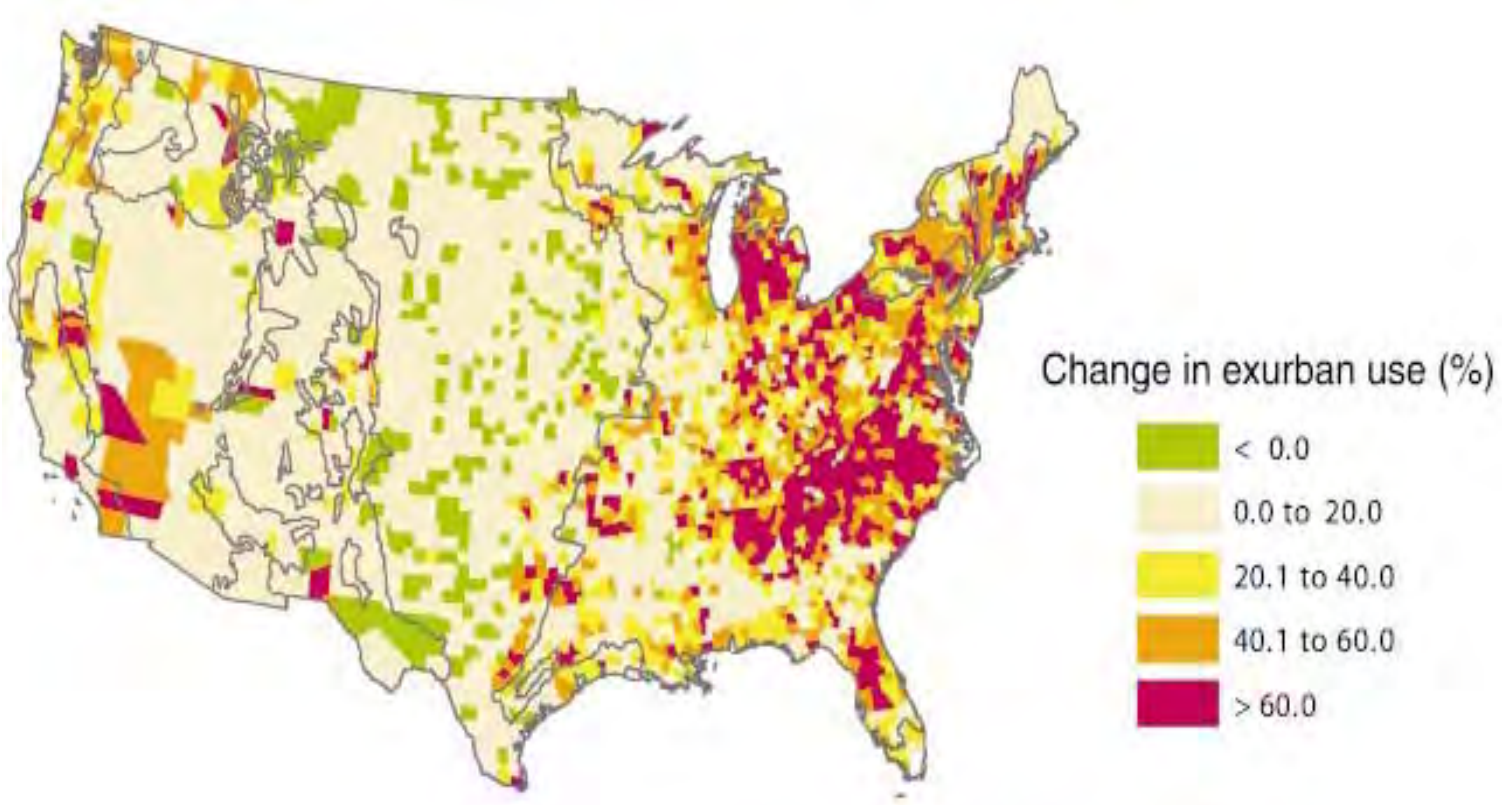
- **Biodiversity_001:** Mosquito abundance and diversity
- **Biodiversity_003**: Ground beetle abundance and diversity
- **Biodiversity_005**: Small mammal abundance, diversity, and density
- **Biodiversity_009**: Bird diversity
- **Biodiversity_013**: Plant abundance, richness, and diversity
- **Biodiversity_027**: Benthic macro-invertebrate abundance and diversity
- **Land Use 007**: Land cover classification

Level 1 Data Products (examples)

- **FSU.1.0013**: Ground dwelling beetle species identification
- **FSU.1.0013**: Ground beetle DNA barcode identification
- **FSU.1.0016**: Small mammal species identification
- **FSU.1.0055**: Bird species identification
- **FSU.1.0065**: Plant species identification
- **AQU.1.0042**: Benthic macro-invertebrate species identification
- **LUAP.1.0015**: MRLC National Land Cover Database

Science Case 3 - Biodiversity

NEON organismal abundance and diversity data from relocatable and core wildland sites will contribute to analysis of biodiversity changes resulting from exurban development.



Change in exurban use by county between 1950 and 2000 (Brown et al. 2005)

Data Product Definitions

Title: NEON Scientific Data Products Catalog	Author: Michael Keller	Date: ####
NEON Doc. #: NEON.MGMT.DPS.005003.REQ		Version: ####

NEON SCIENTIFIC DATA PRODUCTS CATALOG

Michael Keller, Luciana Alves, Steve Aulenbach, Brian Johnson,
Tom Kampe, Rebecca Kao, Michele Kuester,
Henry Loescher, Valerie McKenzie,
Heather Powell, David Schimel

PREPARED BY (Name and Signature)	ORGANIZATION	DATE
Michael Keller	CVL	

APPROVALS (Name and Signature)	ORGANIZATION	DATE
Tony Beasley	CCB Chairman COO	
Tom Cilke	CCB Director of Engineering	
Michael Keller	CCB Chief of Science	
Brian Damiani	CCB Systems Engineer	

RELEASED BY (Name and Signature)	ORGANIZATION	DATE
Dina Cilke	CCB Administrator DCS	

- Source of Level 4 data products
  - Extensive community consultation
  - Tiger teams (2007)
  - Review by NEON staff (2008)
  - Review by STEAC (2008)
  - NSF Science Review (2009)
  - De-scoping prior to PDR resulted in no removals of Level 4 data products although some sub-products were eliminated.
  - The same 117 L4 products are presented for the FDR

NEON Algorithm Theoretical Basis Document Design Specification

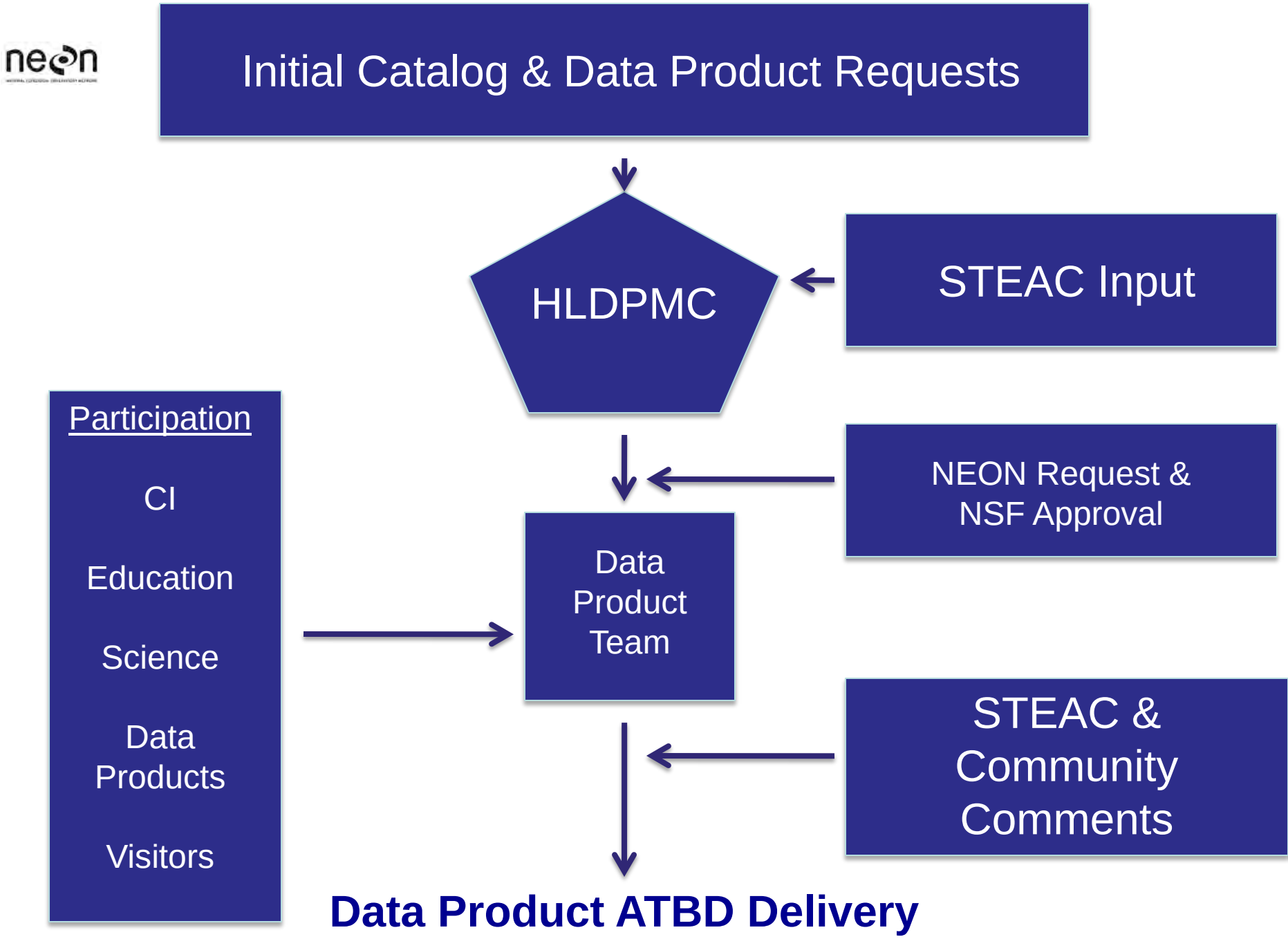
- Scientific motivation
- Scientific basis of the algorithm including research and publications leading to the algorithm
- Data requirements (Level 1-4)
- Mathematical, numerical, statistical and procedural implementation
- Variables reported
- Analysis and reporting of uncertainty
- Calibration and validation

Table 1. Outline of sections for the NEON algorithm theoretical basis document.

Section Number	Section Content	Subsection Content	Required or Optional
1.	Scope and Description		Required
2.	Reference Material		Required
2.1.		Applicable Documents	Required
2.2.		Reference Documents	Required
2.3.		Acronyms	Optional
2.4.		Variables and Symbol Definitions	Required
2.5		Data Product Tree	Required
3.	Data Product Overview		Required
4.	Theory of Measurement and Algorithm		Required
4.1.		Theory of Measurement	Optional
4.2.		Theory of Algorithm	Required
5.	Observations and instrumentation		Required
6.	Algorithm implementation		Optional
7.	Data Product Description		Required
7.1.		Variables reported	Required
7.2.		Product Instances	Required
7.3.		Temporal Scale and Extent	Required
7.4.		Spatial Scale and Extent	Required
8.	Sources of Uncertainty		Required
8.1.		Analysis of uncertainty	Required
8.2.		Reported uncertainty	Required
9.	Calibration and Validation		Required
10.	Scientific and Educational Applications		Optional
11.	Future Modifications and Plans		Optional
12.	References		Required

Data Product Management

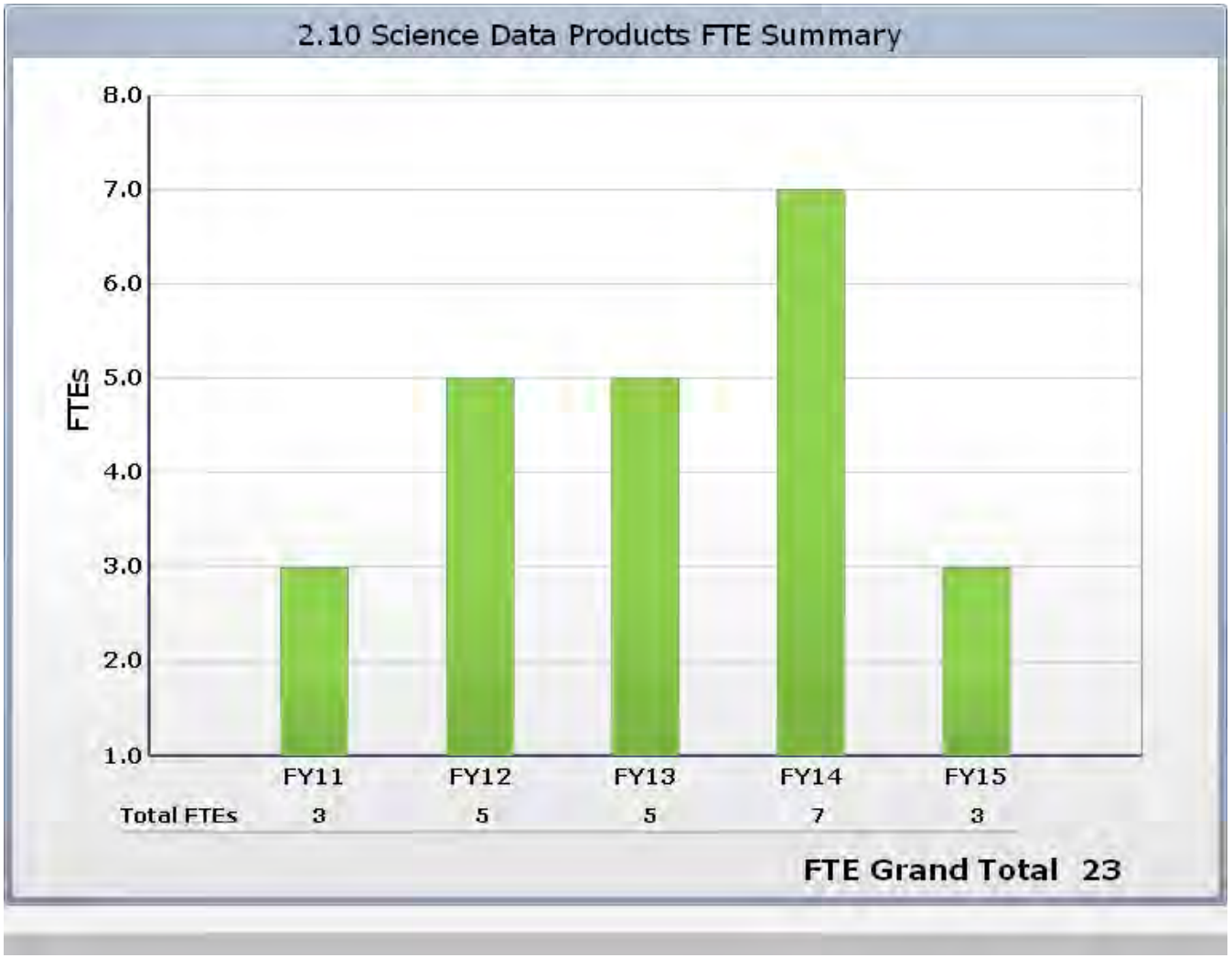
- Produce 117 level 4 data products in the initial catalog during construction.
- 99 data products (complexity A-C) are based on well-known algorithms.
  - Prepare ATBDs primarily with in-house expertise with support of brief consultations by experts (Honoraria)
- 18 data products in the L4 catalog (complexity D & E) require development.
  - Supplement in-house expertise with visiting scientists (~1 per product)
- NEON Science Technology and Education Advisory Committee and community provide input to data products and ATBDs
- During operations upgrade and replace products to reflect changing science


Initial Catalog & Data Product Requests
HLDPMC
STEAC Input
Participation
CI
Education
Science
Data Products
Visitors
NEON Request & NSF Approval
Data Product Team
STEAC & Community Comments
Data Product ATBD Delivery

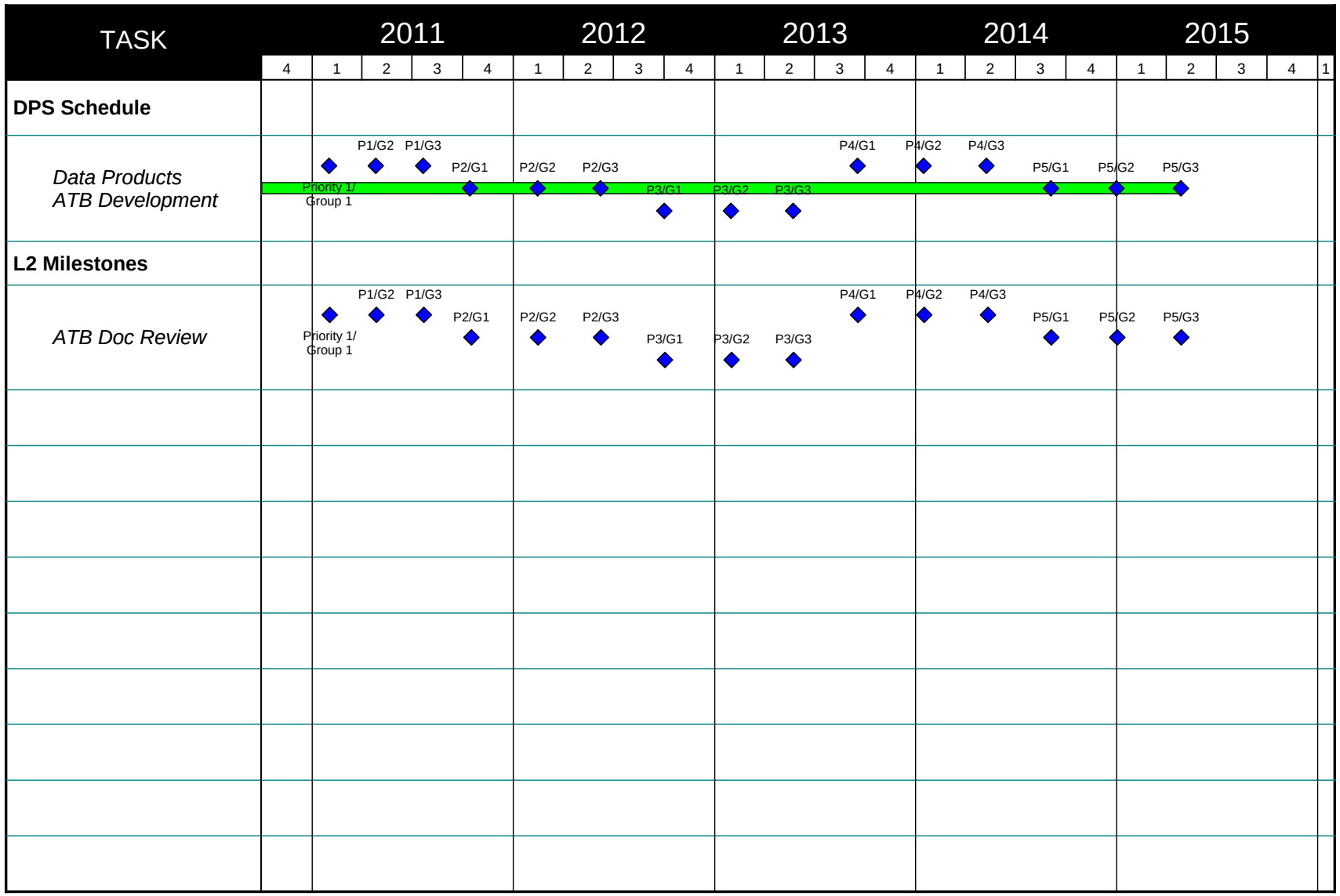
2.10 Science Data Products - WBS

WBS	Title
2.10	Science Data Products
2.10.10	Science Data Products Management
2.10.10.01	Priority 1 Data Products
2.10.10.02	Priority 2 Data Products
2.10.10.03	Priority 3 Data Products
2.10.10.04	Priority 4 Data Products
2.10.10.05	Priority 5 Data Products

2.10 Science Data Products FTE Spread by FY



2.05 Data Products (DPS)
Construction Phase Schedule



Data Products Risk Register

Risk ID	Risk Title	Description	RRS	Risk Exposure	Occurrence Cost	Program Area	Status
41	Pattern Recognition Algorithm (Root Phenology)	Development of pattern recognition algorithm for root and hyphae phenology is delayed or does not meet quality standards. (minirhizotrons)	2.5	High	$ 500,000	DPS	Mitigate
42	Soil Microbial Metagenomes	Development of metagenome algorithm does not meet quality standards.	2.5	High	$ 500,000	DPS	Mitigate
43	Historical Land Cover	Development of historical land cover algorithm is delayed or does not meet quality standards.	1.5	Medium	$ 350,000	DPS	Mitigate
175	Data Product Labor Estimates	Labor estimates associated with data products in general are too low.	1.5	Medium	$ 1,000,000	DPS	Mitigate

- Risk Mitigation beginning with development of a complex data product in the pre-construction period.

- #41 (pattern recognition) and #42 (metagenome) – initiate early start on these algorithms, seek partnerships (NIH, etc.) to leverage existing expertise

Addressing Issues from PDR (1/2)

- *Panel:*
  The Data Products Team is dependent on utilizing small p ortions of time (e.g., 10%) from members of other teams.

- NEON*:* The Data Products Team will not depend upon other team members with the exception of the consultation by PTL's to assure that level 1-3 data products are properly incorporated, etc. Added effort by visiting scientists will supplement data product team activities.

Addressing Issues from PDR (2/2)

- *<u>Panel:</u> Ensure that risks associated with delayed release of ATBDs are well identified and that the work schedule is not front-loaded with just the "easy" ABTDs.*

- <u>NEON</u>*:* The Data Products risks have been revised to include a substantial risk entries associated with insufficient labor and especially complex products. Mitigation will begin pre-construction with work on a complex product. The schedule allows long-lead times for production of complex products.

The Next 12 Months

- Recruiting - starting with a visiting scientist for data assimilation model
- Risk reduction activity – development of a data assimilation version of the Community Land Model (CLM)
- Collaboration with CI on data product prototypes (e.g. porting CLM to NEON)
- Continue observatory simulation and analysis of error budgets
- Initiate ATBDs with existing scientific staff

The National Ecological Observatory Network is a project sponsored by the National Science Foundation and managed under cooperative agreement by NEON Inc.