NOAA'S Scientific Data Stewardship Program: Connecting Essential Climate Variables To Science Questions And Societal Impacts For Long-Term Information Access

Bruce R. Barkstrom, John J. Bates,
and Jeffrey Privette

NOAA's National Climatic Data Center

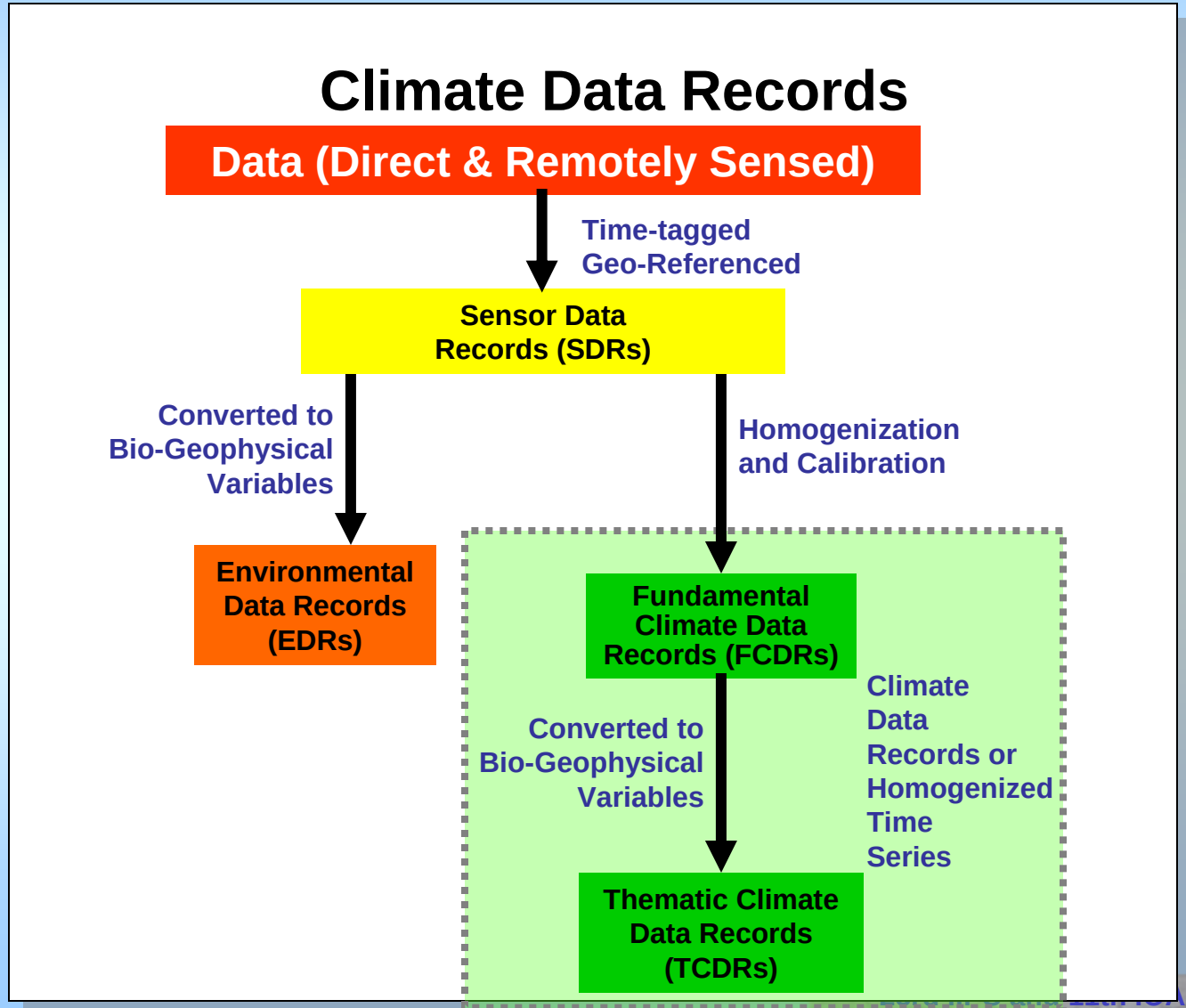
Outline

- Defining Scientific Data Stewardship
  - Characteristics of Climate Data Records
  - A Model for Maturity of Data Sets
  - Defining Roles and Responsibilities – The Data Submission Agreement
  - Assessing and Mitigating Information Survivability
- Prioritizing Observations and Climate Data Records
  - Societal Benefits
  - Science and Policy Benefits
  - Prioritization
  - Engineering Scientific Data Flows
- Conclusions

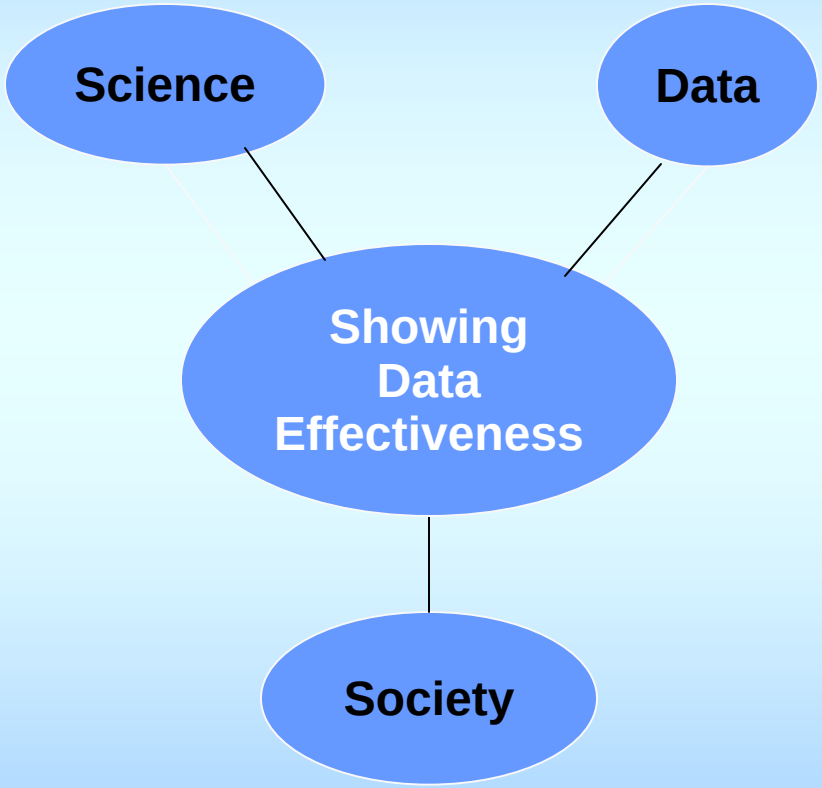
Defining Scientific Data Stewardship - Characteristics of Climate Data Records

- Record Length – Climate variations occur on many time scales, thus requiring long records => this means stitching together many different instrumental records
- Error Structure Homogeneity – Artifacts from data processing can be as large or larger than the signal to be detected => reprocessing will be required
- Assurance of Data Provenance – Enough information must be kept on the heritage and chain of custody of the data so those who were not ever involved with the original data can use them at a future date
- Ability to Retain Usefulness over Long Time Periods – Assurance of data provenance requires context information be preserved in the presence of large changes in information technology and user access methods

Defining Scientific Data Stewardship - Characteristics of Climate Data Records



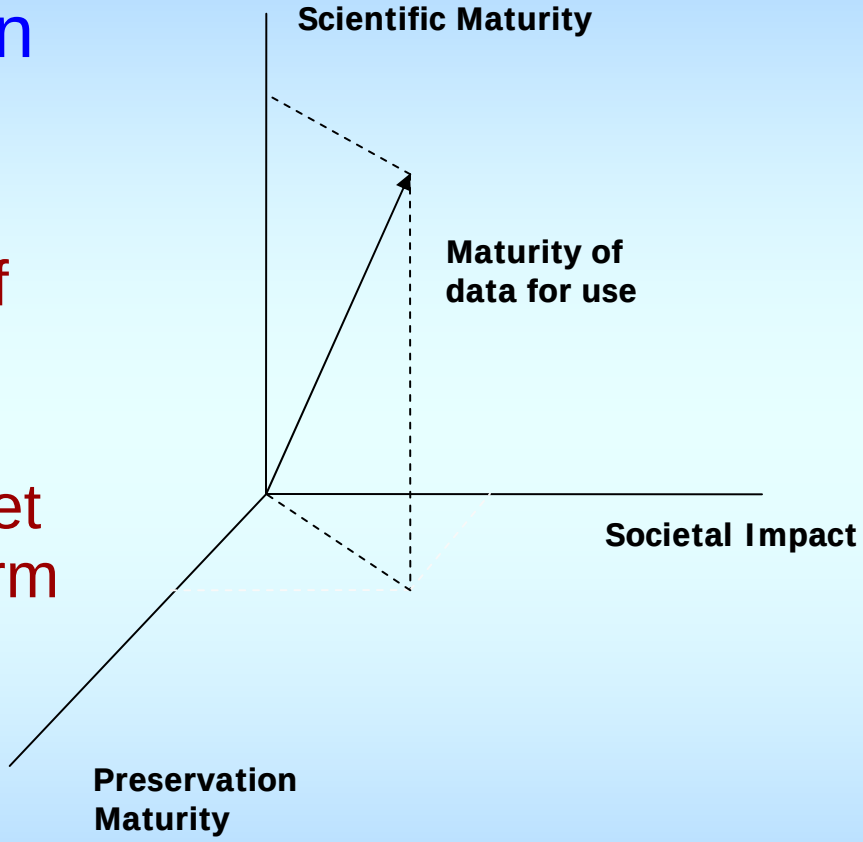
Defining Scientific Data Stewardship
A Model for Maturity of Data Sets



- Differing community views of data maturity and science data stewardship need to reflect wide diversity
  - Vocabulary
  - Experience
  - Background
- Need understandable approach

Defining Scientific Data Stewardship
A Model for Maturity of Data Sets

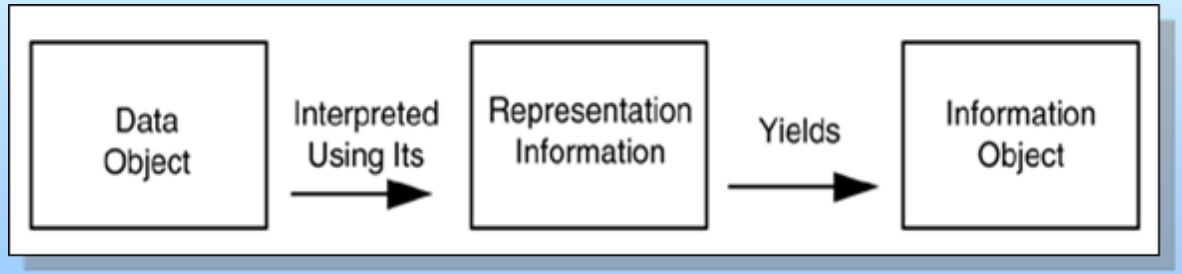
- Represent data maturity in terms of three separate dimensions:
  - Scientific Maturity - a set of measures of the scientific quality of a data product
  - Preservation Maturity - a set of measures of the long-term sustainability of the information in a data set
  - Societal Impact - a set of metrics that assess the potential value of a data set



Defining Scientific Data Stewardship - Roles and Responsibilities: The Data Submission Agreement

- A Climate Data Record producer must negotiation with an archive to ensure their information content is preserved by an archive
- In turn, to ensure information preservation an archive* must:
  - Negotiate and accept information from information producers
  - Obtain sufficient control to ensure long-term preservation
  - Determine which communities (designated) need to be able to understand the preserved information
  - Ensure the information to be preserved is independently understandable to the Designated Communities
  - Follow documented policies and procedures that ensure the information is preserved against all reasonable contingencies
  - Make the preserved information available to the Designated Communities in forms understandable to those communities

*Open Archive Information System – Reference Model

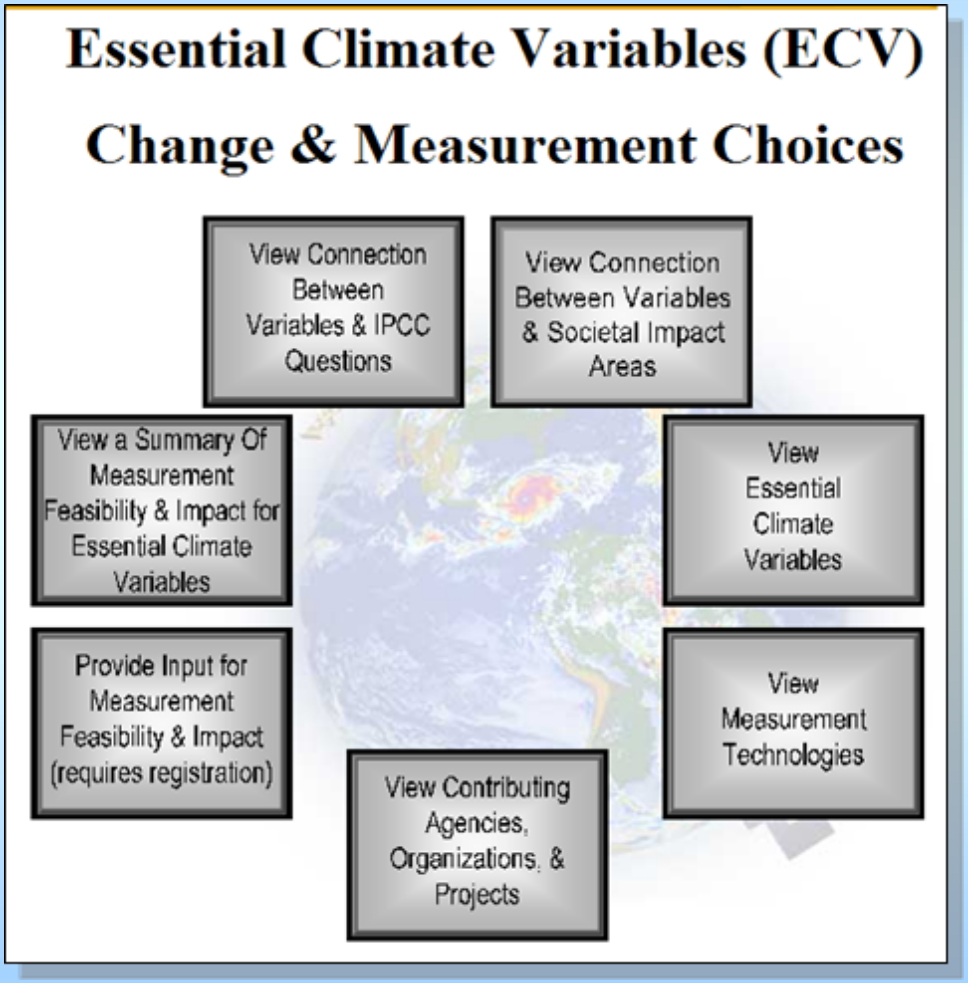


Defining Scientific Data Stewardship – Risk Assessment and Mitigation for Information Survivability

- Risk factors
  - Hardware and software planned obsolescence
  - Migration of data and metadata formats and read software
  - Amount and quality of context data
  - Data production software availability, understandability, and usability
  - Operator error or hardware and software failures
- Mitigation options
  - Keep one copy of critical data sets off line
  - Develop highly reliable, automated data migration and handling procedures
  - Data product preservation conundrum – reprocess from raw data or save all data from all steps

Prioritizing Observations and Climate Data Records – Societal Benefits and Key Science Issues

- Societal benefits of the Global Earth Observations System of Systems (GEOSS) emphasize the tradition of applied climatology but using 21st century systems
- Key Science Intergovernmental Panel on Climate Change (IPCC) Issues address the policy-relevant scientific research topics of highest priority for Climate and Global Change studies



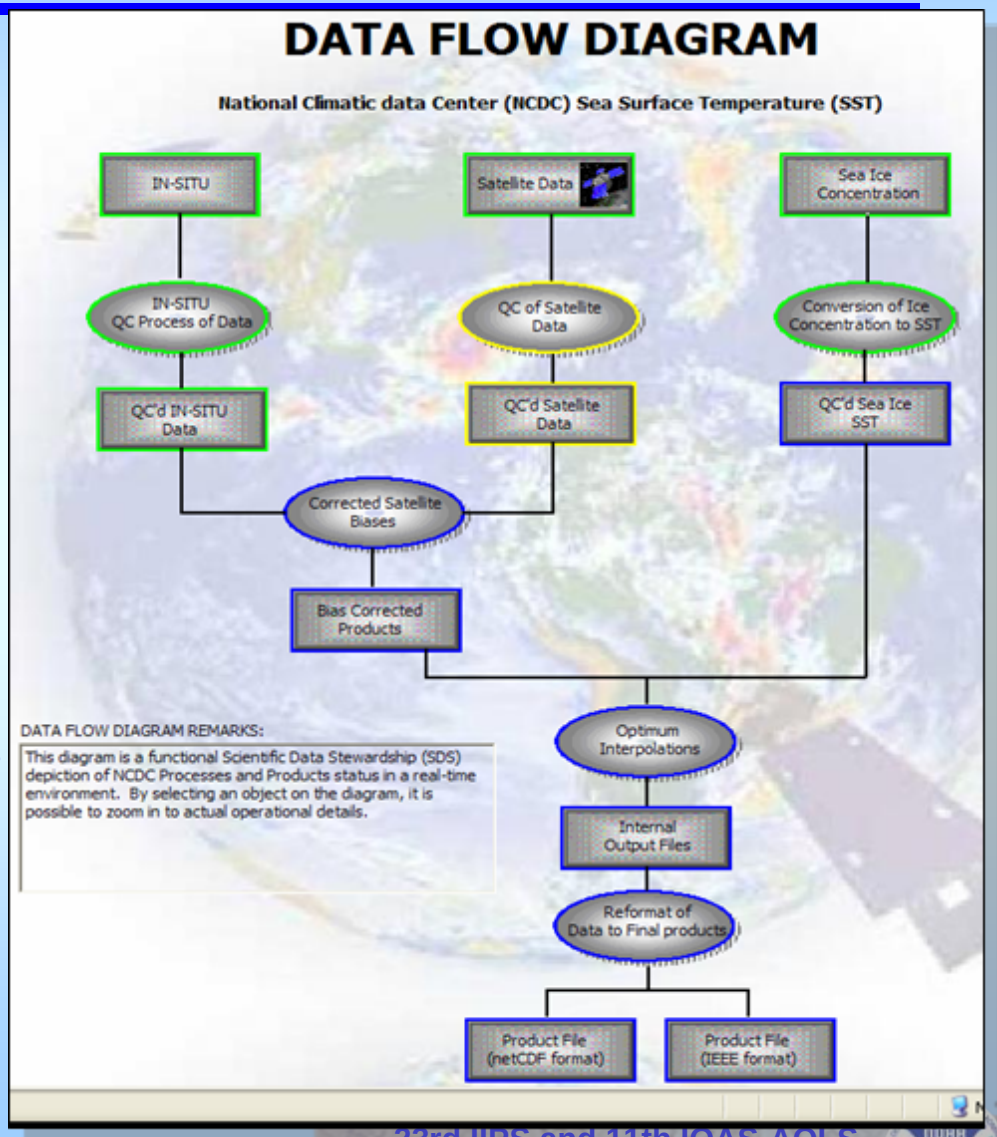
Prioritizing Observations and Climate Data Records – Societal Benefits and Key Science Issues

ECVs:		IPCC IMPACT ISSUES:											
IPCC Matrix ID:	Essential Climate Variable	O3 RECOVERY	EL NINO	ECOSYS & CLIMATE	WARMING POTENTIAL	CO2 SRC&SINKS	UNCERTAINTY	VALIDITY	DETECTION	ATTRIBUTION	HUMAN INTERACTION
1	Ozone Description	S	NA	W	M	NA	NA	NA	NA	NA	S
2	Total Solar Irradiance	NA	NA	NA	NA	NA	S	S	S	S	NA
3	Spectral Solar Irradiance	NA	NA	NA	NA	NA	S	S	S	S	NA
4	TOA Radiation Budget	NA	NA	NA	S	NA	S	S	S	S	NA
5	Cloud Properties	W	S	S	M	NA	S	S	S	S	NA
6	Surface Radiation Budget	NA	NA	NA	M	NA	S	S	S	S	S
7	Surface Albedo	NA	NA	M	M	NA	M	NA	NA	NA	S
8	Sea Level	NA	M	M	NA	NA	S	S	M	S	S
9	Aerosols	NA	NA	M	S	W	S	S	S	S	M
10	Snow Cover	NA	W	M	NA	NA	M	M	W	W	W
11	Glaciers & Ice Caps	NA	NA	M	NA	NA	S	M	M	W	W
12	Ice Sheets	NA	NA	M	NA	NA	S	M	M	W	W
13	Lakes	NA	W	M	NA	NA	NA	W	W	W	W
14	Precipitation	NA	S	S	W	NA	S	S	S	S	W
15	Oceanic Precipitation	NA	S	NA	NA	NA	M	M	M	M	NA
16	SST	NA	S	M	NA	NA	S	S	S	S	W
17	Sea Ice	NA	W	W	NA	NA	M	S	S	S	W
18	Land Cover	NA	NA	S	NA	S	M	NA	S	S	S
19	Evaporation & Evapotranspiration	NA	NA	S	W	W	W	NA	NA	NA	M
20	Soil Moisture	NA	W	S	NA	NA	M	NA	NA	NA	NA
21	Leaf Area Index (LAI)	NA	NA	NA	NA	S	W	NA	NA	NA	NA
22	FAPAR	NA	NA	NA	NA	S	W	NA	NA	NA	NA
23	Biomass	NA	NA	NA	NA	S	M	NA	NA	NA	M
24	Fire Occurrence	NA	M	M	W	S	M	M	S	S	S
25	Active Fire Area	NA	NA	NA	NA	M	NA	NA	NA	NA	M
26	Fire Radiated Power	NA	NA	NA	W	M	NA	NA	W	M	M
27	Volcanic Emissions	NA	NA	NA	W	M	W	NA	NA	NA	NA

ECVs:		BENEFIT IMPACT AREAS:								
Benefit Matrix ID:	Essential Climate Variable	ENERGY	WATER	CLIMATE	AGRICULTURE	DISASTERS	ECOSYSTEMS	WEATHER	HUMAN HEALTH	BIODIVERSITY
1	Ozone Concentration	NA	NA	S	NA	NA	NA	W	W	NA
2	Total Solar Irradiance	S	M	S	NA	NA	NA	NA	NA	NA
3	Spectral Solar Irradiance	S	M	S	NA	NA	NA	NA	NA	NA
4	TOA Radiation Budget	S	S	S	NA	NA	NA	NA	NA	NA
5	Cloud Properties	S	S	S	S	NA	NA	NA	NA	NA
6	Surface Radiation Budget	S	S	S	S	NA	NA	NA	NA	NA
7	Surface Albedo	S	S	S	S	NA	NA	NA	NA	NA
8	Sea Level	NA	NA	S	NA	S	NA	NA	NA	NA
9	Aerosols	S	S	S	NA	NA	NA	NA	S	NA
10	Snow Cover	M	S	S	NA	M	NA	NA	S	NA
11	Glaciers and Ice Caps	M	S	S	NA	M	NA	NA	NA	NA
12	Ice Sheets	M	S	S	NA	M	NA	NA	NA	NA
13	Lakes	M	S	S	NA	NA	NA	NA	NA	NA
14	Precipitation	M	S	S	S	M	S	NA	NA	NA
15	Oceanic Precipitation (fresh	W	M	S	NA	NA	NA	NA	NA	NA
16	SST	M	M	S	NA	NA	NA	NA	NA	NA
17	Sea Ice	NA	NA	S	M	NA	NA	NA	NA	NA
18	Land Cover	NA	NA	M	S	W	S	NA	NA	NA
19	Evaporation and Evapotranspiration	S	S	S	NA	M	M	NA	NA	NA
20	Soil Moisture	NA	M	NA	S	M	M	NA	NA	NA
21	Leaf Area Index (LAI)	NA	NA	NA	S	NA	M	NA	NA	NA
22	FAPAR	NA	NA	NA	S	NA	M	NA	NA	NA
23	Biomass	NA	NA	NA	S	M	S	NA	NA	NA
24	Fire Occurrence	NA	NA	NA	M	S	S	NA	NA	NA
25	Active Fire Area	NA	NA	NA	M	S	M	NA	NA	NA
26	Fire Radiated Power	NA	NA	NA	W	S	W	NA	NA	NA
27	Volcanic Emissions	NA	NA	NA	W	S	NA	NA	NA	NA

Engineering Scientific Data Flows

- ***Essential Climate Variables*** – and the deeper ties to individual parameters and data structures contained in the files that usually contain archived data
- ***Justification and Prioritization*** – which provide the ties between ECVs and the societal benefits as well as scientific issues
- ***Data Sources*** – meaning the platforms and instruments that provide the Level 0 data, including both satellite-based systems and in-situ data sources
- ***Data Set Versions*** – by which we mean data collections that have common contents, common time intervals of data collection, common data sources, and homogeneous error structure
- ***Data Flow Diagrams*** – and the extensions that provide the connectivity between different data sources and data versions
- ***Uncertainty Assessments*** – which we treat like error budgets, with the added assessment of systematic biases and probability distributions
- ***Production Schedules*** – based on both data sources and (re)processing expectations



Conclusions

- Climate products and services are transforming from the research to operations
- Care and deliberations are required to capture and define required roles and responsibilities of all steps involved
- A robust system for operational production of climate data records for essential climate variables is being developed