FP-7 CORE-CLIMAX: HOW TO ENSURE SUSTAINABLE, TRANSPARENT, AND TRACEABLE GENERATION OF CLIMATE DATA RECORDS?

Joerg Schulz, <u>Viju John</u>, Andrea Kaiser-Weiss, Rob Roebeling, Andre Obregon, Else Swinnen, Carolien Tote, Ali Nardir Arslan, Jean-Christophe Calvet, Hilppa Gregow, Terhikki Manninnen, Paul Poli, Bob Su, David Tan, Wim Timmermans, and Yijian Zeng





GCOS-WCRP Letter, 12 May 2010 to many agencies:

"However, there is currently no systematic international approach to ensure transparency, traceability and sound scientific judgement in the generation of climate data records across all fields of climate science and related Earth observations, and there are no dedicated sustained resources in place to support such an objective. For example, there are currently eight sea-ice concentration products produced by different organizations globally that differ significantly in providing an estimate of sea-ice extent and concentrations, mostly due to differences in methodology and not the variability or dynamics of underlying phenomenon. It is very confusing and frustrating for the non experts as to which one of these products they can use in their research and analysis, and the necessary documents to describe their attributes in a comparative manner akin to the global model inter-comparisons do not exist."



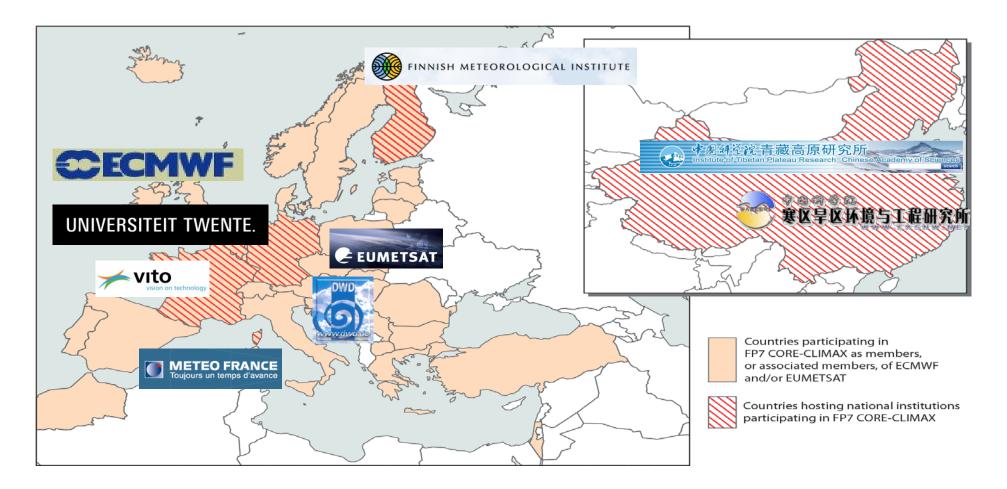






CORE-CLIMAX

COordinating Earth observation data validation for RE-analysis for CLIMAte ServiceS

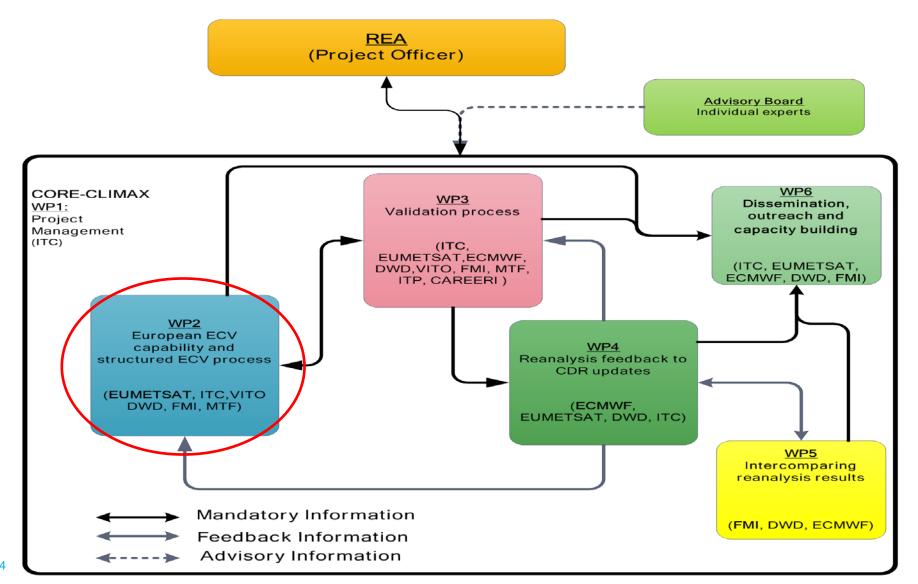








CORE-CLIMAX work packages



EU FP7 CORE-CLIMAX Assessment of European Capacity for CDRs



The capacity is assessed using three support tools developed by the project:

Data Record Descriptions (DRD)

- Contain technical specifications and links to documented information on quality;
- Provides consistent and coherent information about CDRs produced in Europe (serves as input to CMIP-6 obs4mips activities).

System Maturity Matrix (SMM)

- Evaluates if the production of a CDR follows best practices for science and engineering and is assessing if data records are used and feedback mechanisms are implemented;
- The SMM can be used in self assessment mode or in an audit type assessment.

Application Performance Metric (APM)

- Evaluates the performance of a CDR with respect to a specific application;
- Might be implemented as an interactive App that convolves user requirements with product specification information in a database.



Maturity Matrix Concept



	Is the software robust and aaintainable?	Are the data and methods well documented?	Has the uncertainty the data be systematica assessed?	en feedb lly taken ca	d user acks
Software readiness	Metadata	User documentation	Uncertainty Characterisation	Public Access, Feedback and Update	Usage
Are the codes compliant with standards, stable, portable and reproducible?	Do the metadata meet international standards, and allow provenance tracking?	Are the formal documents and peer-reviewed papers up-to-date and public?	Are the uncertainties assessed systematically in a standard manner?	Are the data, source code, and documents publicly available and regularly updated?	Are the data widely used in the scientific, and decision and policy making communities?

A СС star р



Core-Climax: System Maturity Matrix



Maturity	SOFTWARE READINESS	METADATA	USER DOCUMENTATION	UNCERTAINTY CHARACTERISATION	PUBLIC ACCESS, FEEDBACK, UPDATE	USAGE
1	Conceptual development	None	Limited scientific description of the methodology available from PI	None	Restricted availability from PI	None
2	Research grade code	Research grade	Comprehensive scientific description of the methodology, report on limited validation, and limited product user guide available from PI; paper on methodology is sumitted for peer-review	Standard uncertainty nomenclature is idenitified or defined; limited validation done; limited information on uncertainty available	Data avaliable from PI, feedback through scientific exchange, irregular updates by PI	Research: Benefits for applications identified DSS: Potential benefits identified
3	Research code with partially applied standards; code contains header and comments, and a README file; PI affirms portability, numerical reproducibility and no security problems	Standards defined or identified; sufficient to use and understand the data and extract discovery metadata	Score 2 + paper on methodology published; comprehensive validation report available from PI and a paper on validation is submitted; comprehensive user guide is available from PI; Limited description of operations concept available from PI	Score 2 + standard nomenclature applied; validation extended to full product data coverage, comprehensive information on uncertainty available; methods for automated monitoring defined	Data and documentation publically available from PI, feedback through scientifc exchange, irregular updates by PI	Research: Benefits for applications demonstrated. DSS: Use occuring and benefits emerging
4	Score 3 + draft software installation/user manual available; 3rd party affirms portability and numerical reproducibility; passes data providers security review	Score 3 + standards systematically applied; meets international standards for the data set; enhanced discovery metadata; limited location level metadata	Score 3 + comprehensive scientific description available from data provider; report on inter comparison available from PI; paper on validation published; user guide available from data provider; comprehensive description of operations concept available from PI	Score 3 + procedures to establish SI traceability are defined; (inter)comparison against corresponding CDRs (other methods, models, etc); quantitative estimates of uncertainty provided within the product characterising more or less uncertain data points; automated monitoring partially implemented	Data record and documentation available from data provider and under data provider's version control; Data provider establishes feedback mechanism; regular updates by PI	Score 3 + Research: Citations on product usage in occurring DSS: societal and economical benefits discussed
5	Score 4 + operational code following standards, actions to achieve full compliance are defined; software installation/user manual complete; 3rd party installs the code operationally	Score 4+ fully compliant with standards; complete discovery metadata; complete location level metadata	Score 4 + comprehensive scientific description maintained by data provider; report on data assessment results exists; user guide is regularly updated with updates on product and validation; description on practical implementation is available from data provider	Score 4 + SI traceability partly established; data provider participated in one inter-national data assessment; comprehensive validation of the quantitative uncertainty estimates; automated quality monitoring fully implemented (all production levels)	Score 4 + source code archived by Data Provider; feedback mechanism and international data quality assessment are considered in periodic data record updates by Data Provider	Score 4+ Research: product becomes reference for certain applications DSS: Societal and economic benefits are demonstrated
6	Score 5 + fully compliant with standards; Turnkey System	Score 5 + regularly updated	Score 5 + journal papers on product updates are and more comprehensive validation and validation of quantitative uncertainty estimates are published; operations concept regularly updated	Score 5 + SI traceability established; data provider participated in multiple inter-national data assessment and incorporating feedbacks into the product development cycle; temporal and spatial error covariance quantified; Automated monitoring in place with results fed back to other accessible information, e.g. meta data or documentation	Score 5 + source code available to the public and capability for continuous data provisions established (ICDR)	Score 5 + Research: Product and its applications becomes references in multiple research field DSS: Influence on decision and policy making demonstrated





Is the Core-Climax SMM concept generally applicable? (In-situ, Satellite, and Reanalysis CDRs)

Baseline Surface Radiation Network (BSRN)

		Baseli	ne Surface Radiation No	etwork		maturity level as of 0.152004
	rot applicable CORE-CLIMAX System Maturity Matrix					
Maturity	SOFTWARE READINESS	METADATA	USER DOCUMENTATION	UNCERTAINTY CHARACTERISATION	PUBLIC ACCESS, FEEDBACK, UPDATE	URAGE
1	Conceptual development	Nam	Limited scientific description of the methodology available from 71	None	Restricted analability from PL	Nam
2	Revences grade code	Research grade	Competensive scientific description of the methodicing, report or limited middeirs, and limited product user paids available from PC; paper os methodology is manited for percentier	Sandard unortainty someclature is identified or definet; limited validation dore; finited afformation or secretainty available	Data evalutile from PL fieldwale through scientific escharge, angular updates by PL	Result: Bendits for applications identified DIS: Potential bendits identified
3	Research code with partially applied standards, code contains leader and consents, and a #2ADAE file, PI affirms portability, manerical reproducibility and as security problems	Standards defined or identified, sufficient to use and understand the data and estant discovery metadata	Score 2 + paper on methodology published; compodensis validation report available from FT and a paper on validation is submitted; comprehensive user gaids is available from FT. Limited description of operations coupt available from FT	Score 2 – standard somenclature applied, validation extended to full product data coverage, comprohensive information on sometrasity available, methods for automated monitoring defaned	Dats and documentation publically available from PL firedback through scientific exchange, irregular updates by PL	Research: Besefits for applications demonstrated. DSI: Use occaring and benefits enouging
•	Score I + draft software institlation lawe manual available, itsi party affirma portubility and numerical reproducibility; process data providers security review	Seen 3 + standards systematically applied, meets international standards for the data set; submood discovery metadate, limited location level metadate	Scen 3 - comprehensive scientific description available from data provider, report on reter comparison available from PL, paper on validation published, user guide available from data provider, comprehensive description of operations concept available from PL	Score 3 + percedures to establish SI transbility are defined; (attentionsparison against corresponding CDRs (other methods, mobile, (c), quantitative estimates of scoretariary provided within the product characterizing more the scorectaria data points; automated memoring partially implemented	Data record and decomentation multiple from data provider and under data provider's version control; Data provider establishes feedback esterbasism; regular updates by Pl	Score 3 + Research: Gitations on product usage in eccuring DBS societal and eccentratial benefits discussed
5	Scate 4 * operational code following standards, actions to achieve full compliance are defined; software installation twee manual complete, ited party installs the code operationally	Score 4+ fully compliant with standards; complete discovery metadata; complete location lovel metadata	Score 4 + comprehensive scientific description maintained by data provider, report on data summante results exists; user gade in regularly updated with updates on product and validation; description on practical implementation: is available from data provider	Score 4 + SI transibility partly established, data provider participated in one inter-maximal data movement; comprehensive raidations of the quantizative accordancy estimates, automated quality monitoring fully implemented (all preduction levels)	Score 4 + sourcede antived by Data Provide; fordback mechanism and international dras quality assessment are considered in periodic data record speaters by Data Previder	Sorre 4+ Essearch: product becomes reference for pertain applications DSS: Società and economic benefits are demonstrated
6	Scare 5 = fully compliant with standards; Tunkey System	Score 5 + replaty splated	Score 5 + journal papers on product apdates are and more compendentive validations and validation of quartizative sourcitary estimates are published, operations encorpt regularly apdated	Sees 1 + 31 taxahiday andafashad, data persider patricipated in multiple inter-actival data subsesser taki acceptorating fisedata into the product development cyclic temporal and patrice new remaining quantifier, Attaneous data activity in place with smalls fol back to other accessible information, e.g. sets data or documention	Score 5 + source code available to the public and capability for continuous data previouse established (CCR)	Score 5 + Ensuech: Product and its applications because indexes in militple research fold DSS: Influence on decision and policy making dementationd

NKDZ Precipitation time series

NKDZ Precipitation time series (daily station data)				maturity level as of 01.09/2014 (mm/dd/yyyy)			
	CORE-CLIMAX System Maturity Matrix						
Maturity SOFTWARE READDRESS METADATA			USER DOCUMENTATION	UNCERTAINTY CRARACTERISATION	PUBLIC ACCESS, FEEDBACK, UPDATE	USICE	
1	Conceptual development	Nee	Limited scientific description of the methodology available from 71	Nee	Restricted analobility from Pl	Name	
2	Research grade code	Research grade	Compositensive scientific description of the methodology, report or limited validation, and limited product user paide available from PC, paper on methodology is sumitted for pare review	Sandarf unortainty annexistua is identified or definet; limited validation done, finited information on secretarity available	Data evalutile from PI, feedback through scientific endange, angular updates by PI	Resurch Benefits for applications identified DIS: Potential benefits identified	
3	Research code with partially applied standards, code contains leader and comments, and a FEADAE file, PI affirms portability, manerical reproducibility and no security problems	Standards defined or identified, sufficient to use and understand the data and extract discovery metadata	Score 2 + paper on methodology published; compositencies validation report available from FI and a paper on validation is submitted; comprehensive user paile is available from FI. Linated description of operations except available from FI	Score 2 – standard somerclarure applied, validation encoded to full product data coverage, comprohensive information en mentanty available, methods for automand monitoring defand	Data and documentation publically available from PL firstback through scientific escharge, irregular spaties by PL	Rosards Bosefits for applications descentrated. DSS: Use occuring and benefits enorging	
•	Score I + draft software institlicies tow manual available; ind party affirms portability and numerical reproducibility; process data providens security ordere	Scen 3 + standards systematically applied; meets meranical standards for the data set; eshanced discovery netadars; limited location lovel metadars	Scen 3 - comprehensive scientific description available from data provider, report on reter comparison available from Pi, paper os validation published, user guide available from data provider, comprehensive discription of operations concept available from Pi	models, etc), quantitative estimates of uncertainty provided	Data record and documentation available from data provider and nade data provider's vension control; Data provider establishes feedback: mechanism; repair: spidates by Pl	Score 3 + Research: Citations on product stage in eccuring DBS societal and ecoectaical bearfuls discussed	
5	Score 4 = operational code following standards, actions to achieve full compliance are defined, software installation tower manual complete, ited party installs the code operationally	Score 4+ fully compliant with standards; complete discovery metadata; complete location lovel metadata	Score 4 * comprehensive solestific description maintened by data provider, report on data suresment results exists; sare gade in regizally supdated with updates on product and validation, description on practical implementative a available from data provider	Score 4 + SI tracebility partly established, deta provider participated in one inter-eacland data assessment; comprehensive raidation of the quantizative score-tainty estimates, automated quality monitoring fully implemented (all production levels)	Score 4 + source-cole archived by Data Provide; feedback mechanism and international data quality assessment are considered in periodic data record splaters by Data Previder	Scen 4+ Events: produc because inference for ontain applications DSS: Societal and economic benefits an demonstrated	
•	Score 5 + fully compliant with standards; Turskoy System	Score 5 + reptlatly splated	Score 5 + journal papers on product aplates are and more comprehensive validations and validation of quartization security estimates are published, operations encorpt regularly aplated	Soon I + El tacenhility enablasier, deta pervider participated in maltyle intre-enciental data assessment sub acceptantag desduais sins the product development cycle, temporal and guide nere transmouter quartificher, Atasatuat durationer pi pater with results fol Jack to other accentille reformation, e.g. meta data tri documentation	Score 5 + source code available to the public and capability for continues data previouse established (JCCR)	Scen 5 + Essent: Product and in applications because references in miliple research 544 DSS: Influence on decision and policy miking demonstrated	

Providers of SMMs for In-Situ CDRs initially indicated that the **Software Readiness** and **User Documentation** categories are not applicable to their data.

ERA-Interim (ECMWF)

	ECMWF Interim Reanalysis (ERA-Interim) CORE-CLIMAX System Maturity Matrix					
Maturity						USAGE
1	Conceptual development	None	Limited scientific description of the methodology available from PI	None	Restricted availability from PI	None
2	Research grade code	Research grade	Comprehensive scientific description of the methodology, report on limited validation, and limited product user guide available from PI; paper on methodology is sumitted for peer-review	Standard uncertainty nomenclature is idenitified or defined; limited validation done; limited information on uncertainty available	Data avaliable from PI, feedback through scientific exchange, irregular updates by PI	Research: Benefits for applications identified DSS: Potential benefits identified
3	Research code with partially applied standards; code contains header and comments, and a README file; PI affirms portability, numerical reproducibility and no security problems	Standards defined or identified; sufficient to use and understand the data and extract discovery metadata	Score 2 + paper on methodology published; comprehensive validation report available from PI and a paper on validation is submitted; comprehensive user guide is available from PI; Limited description of operations cocept available from PI	Score 2 + standard nomenclature applied; validation extended to full product data coverage, comprehensive information on uncertainty available; methods for automated monitoring defined	Data and documentation publically available from PI, feedback through scientific exchange, irregular updates by PI	Research: Benefits for applications demonstrated. DSS: Use occuring and benefits emerging
4	Score 3 + draft software installation/user manual available; 3rd party affirms portability and numerical reproducibility; passes data providers security review	Score 3 + standards systematically applied; meets international standards for the data set; enhanced discovery metadata; limited location level metadata	Score 3 + comprehensive scientific description available from data provider, report on inter comparison available from PI; paper on validation published; user guide available from data provider; comprehensive description of operations concept available from PI	Score 3 + procedures to establish SI traceability are defined; (inter)comparison against corresponding CDRs (other methods, models, etc); quantitative estimates of uncertainty provided within the product characterising more or less uncertain data points; automated monitoring partially implemented	Data record and documentation available from data provider and under data provider's version control; Data provider establishes feedback mechanism; regular updates by PI	Score 3 + Research: Citations on product usage in occurring DSS: societal and economical benefits discussed
5	Score 4 + operational code following standards, actions to achieve full compliance are defined; software installation/user manual complete; 3rd party installs the code operationally	Score 4+ fully compliant with standards; complete discovery metadata; complete location level metadata	Score 4 + comprehensive scientific description maintained by data provider; report on data assessment results exists; user guide is regularly updated with updates on product and validation; description on practical implementation is available from data provider	Score 4 + SI traceability partly established; data provider participated in one inter-national data assessment; comprehensive validation of the quantitative uncertainty estimates; automated quality monitoring fully implemented (all production levels)	Score 4 + soure code archived by Data Provider; feedback mechanism and international data quality assessment are considered in periodic data record updates by Data Provider	Score 4+ Research: product becomes reference for certain applications DSS: Societal and economic benefits are demonstrated
6	Score 5 + fully compliant with standards; Turnkey System	Score 5 + regularly updated	Score 5 + journal papers on product updates are and more comprehensive validation and validation of quantitative uncertainty estimates are published; operations concept regularly updated	Score 5 + SI traceability established; data provider participated in multiple inter-national data assessment and incorporating feedbacks into the product development cycle; temporal and spatial error covariance quantified; Automated monitoring in place with results fed back to other accessible information, e.g. meta data or documentation	Score 5 + source code available to the public and capability for continuous data provisions established (ICDR)	Score 5 + Research: Product and its applications becomes references in multiple research field DSS: Influence on decision and policy making demonstrated

 (\mathbf{i})

BY

EUMETSAT

 \bigcirc



Fitness for Purpose? Motivation for Application Performance Metric (APM)

- SMM provides assessment of whether the data set can be sustainable in terms of engineering, science, archive, and usage aspects;
- There is no guarantee that a data set with high System Maturity is suitable for all applications!
- For example, data set X with over all System Maturity FIVE/SIX that provides DAILY mean humidity values is NOT suitable for assessing DIURNAL cycle of humidity in climate models;
- How do we assess the performance of a data set for a particular application?





Support User's to Select Data



- User requirements collection exercises show a large variability in the stated requirements of users with nominally similar applications;
- But a core set of typical questions may always be isolated:

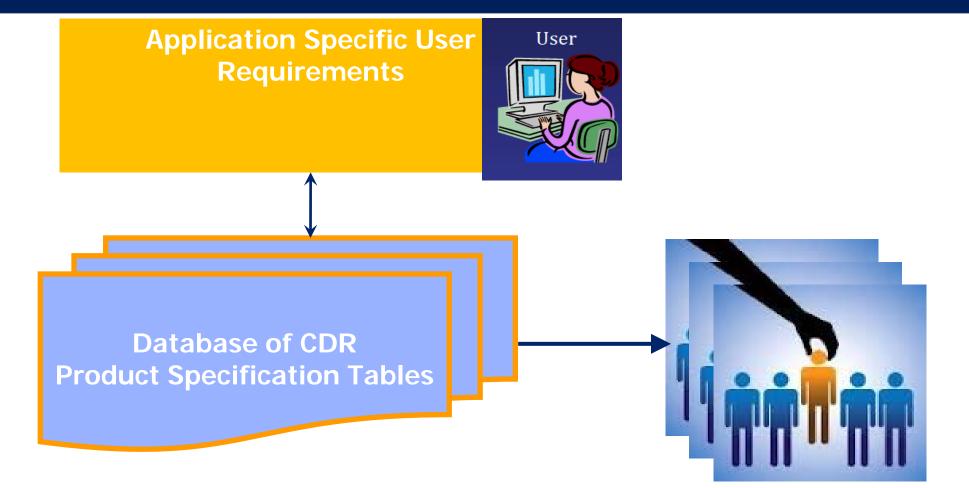


Coverage	Sampling	Uncertainty	Stability
Are the record length and spatial coverage meeting the application's requirements?	Do the spatial and temporal sampling meet the applications requirements?	Do the random and systematic uncertainties meet the requirements?	Do the temporal and spatial stability meet the requirements?



EUMETSAT

General Concept of APM



CDRs matching the URs best



European Capacity Assessment



- Data Assessment Workshop was held at EUMETSAT in January 2014;
- All major dataset developers were present and they endorsed the assessment tools (DRD, SMM, and APM);
- 39 data records were assessed (FCDRs and TCDRs for atmosphere, ocean, and land);
- The assessment:
 - Provides consistent view on strengths and weaknesses of the process to generate, preserve and improve CDRs to each individual CDR producer, agencies, and EC;
 - Provides the status of CDRs for the first time across different observing systems (satellite, in situ, and reanalysis);
 - Increases transparency and openness towards the user;
 - Supports selection of CDRs for services and applications;
 - Supports Europe's contribution to the next Obs4Mips activity by providing consistent information on CDRs produced in Europe;
- An assessment report is being prepared, which will be made publicly available .

