FIDUCEO has received funding from the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no. 638822

WP. 3
Data Tools & Services

Project Achievements

T.Block

WP 3 Aims

WP 3 covers all tasks related to software tools and data formats used in the FIDUCEO project.

Three groups of tasks

- Multisensor Matchup System (MMS)
- CDR and FCDR data format and tools
- Data handling and user support

Overview

1. **Multisensor Matchup System**
   - System description
   - Data processed
2. **Data Format and Tools**
   - Format description
   - Supportive tools
3. **Data Handling and User Support**

Task 3.1 MMS

- MMS evolved from a first idea to a mature, scalable and high-performance processing framework.
- Multiple plugin-points allow extensions
- Highly configurable
- Massively parallel execution
- Self-contained output data
- Exhaustive manual
- GPL open software

Task 3.1 MMS (2)

- Three components
  - Ingestion tool
  - Matchup tool
  - Post-Processing tool
- Supportive components
  - GIS-Database system
  - Glue-code to JASMIN LSF

Task 3.1 MMS (3)

- Input data
  - 2.2 million metadata records
  - 400 TB input products
  - Covering 17 satellite sensors and 3 in-situ data types
    - AIRS, AMSR-2, AMSR-E, AMSU-B, (A)ATSR, AVHRR (GAC/FRAC/FCDR), CALIOP (Clay/VFM), HIRS (reproc./FCDR), IASI, MHS, MODIS MxD06, SLSTR, SSM-T2
    - SST-CCI, OceanRain, GRUAN

Task 3.1 MMS (4)

- Output data
  - 21 MMDs calculated
  - Total volume of MMDs: 4.75 TB
  - Total processing time: 730 hours using ~ 100 CPUs
  - Which sums up to 8.3 CPU years

- Thanks to JASMIN to provide parallel processing capabilities!

Task 3.1 MMS (5)

- Problems:
  - Availability of processing resources
- Lessons learned:
  - New technique to detect matchups quickly
  - Use of spherical geometry library
  - Switchable databases
  - Configuration of an MMD processing is a tedious task
- Missing:
  - Validation tool / plotting tool for MMDs

Task 3.2 FCDR Data Format

- FCDR format specification iteratively developed
- Close collaboration with sensor teams
- 15 releases during project time
- Final version confirmed by all teams
- Covers all sensor FCDRs

Task 3.2 FCDR Data Format (2)

- Format specification covers
  - File naming conventions
  - Global metadata content
  - Definition of correlation and probability density functions
  - Definition of virtual variable syntax
  - Definition of common variables and dimension names
  - Sensor specific variables for EASY and FULL format

Task 3.3 FCDR I/O Software

- Supportive software module
  - FCDRWriter
  - FCDRReader
- Python library
  - Tested with 2.7, 3.5 and 3.6
  - Test-driven development, test-coverage > 98%
  - Build server support
- Integrated into sensor teams processing code (except MW FCDR)

Task 3.3 FCDR I/O Software (2)

- FCDRWriter
  - Implements FCDR product specification
  - Ensures format fidelity of data
  - Simplifies processing tools:
    - Request product template
    - Fill in data
    - Generate product file name
    - Write
  - Automated checks
    - Missing data
    - Numerical over/underflows

Task 3.3 FCDR I/O Software (3)

- FCDRReader
  - Generally not required, data is NetCDF self descriptive
  - Reader module treats virtual variables
    - Expand the mathematical expression to data
    - On-the-fly calculation when variable is touched

Task 3.4 CDR Data Format

- CDR format specification iteratively developed
- Close collaboration with CDR teams
- Final version confirmed by teams
- Covers all CDRs
  - Global metadata
  - Common data
  - CDR specific variables
  - File naming convention

Task 3.5 CDR I/O Software

- Supportive software module for writing CDRs
- Implements CDR format specification
- Similar approach as FCDR tools
- Shared code-base
- Re-use of common functionality
  - Over/underflow detection
  - File name generation
  - Many more …
- Reader module not necessary

Task 3.6 Tools

- Tool for uncertainty propagation for L2 retrieval algorithms
  - Python cmd-line tool
  - Framework and user-plugin points
  - Supports all easy FCDRs
- Example implementations for
  - UTH, SST and NDVI
- Implements full „CURUC“ math, taking inter-channel correlations into account
- Estimate uncertainty budget of user retrieval algorithms

Task 3.7 Final FCDR and CDR

Task 3.8 Archiving

Will be covered by Esther Conway

WP7 - Data archiving and availability

WP 3 Deliverables

- PD3.1 - FIDUCEO Multi-sensor Match up system Implementation Plan - done
- PD3.2 - Multi-sensor matchup system implementing identified use cases - done
- D3.1 - Description of easy FCDR format - done
- D3.2 - Description of CDR format - done
- D3.3 - Software for easy FCDR format - done
- D3.4 - Software for CDR format - done
- PD3.3 - Software regridding tool – re-defined: uncertainty propagation tool - done

WP 3 Deliverables (2)

- D3.5 - FIDUCEO data converted to new formats – done
- PD3.4 - Dataset: FIDUCEO publically available from archives – ongoing activity

Thank you!