STANDARDS FOR AIRBORNE HYPERSPECTRAL IMAGE DATA

S. Holzwarth [a, *], M. Bachmann [a], M. Freer [b], M. Hofmann [a]

[a] Deutsches Zentrum für Luft- und Raumfahrt (DLR), Deutsches Fernerkundungsdatenzentrum (DFD), 82234 Oberpfaffenhofen, Germany
[b] Météo-France CNRM/GMEI, 42 avenue G. Coriolis, 31057 Toulouse, France

KEYWORDS: EUFAR, Standards and Protocols, Metadata, INSPIRE, Toolbox, Quality Control.

ABSTRACT:

A key activity of the FP7/EUFAR project (http://www.eufar.net) is providing transnational access to different infrastructures for airborne research. The variety of aircraft and instruments for airborne measurements and the huge number of institutions involved in the EUFAR project introduces a heterogeneous pool of data and many different ways of handling the processes involved in airborne measurements. Therefore, there is a need to introduce standards and best practices within EUFAR to ensure integration and interoperability.

Thus, the main goal of EUFAR's Networking Activity "Standards and Protocols" is to develop these standards and best practices to ensure harmonization, integration and interoperability, to assist new/inexperienced users and to enable easier exchange and comparison of data.

For airborne hyperspectral remote sensing, common protocols have been agreed upon for flight campaign planning, metadata and data distribution and cataloguing. The existing specifications for metadata within the INSPIRE regulations have been implemented and expanded with additional metadata for quality indicators which were requested by EUFAR's Joint Research Activity "HYQUAPRO".

In this paper, the current recommendations of the Networking Activity "Standards and Protocols" will be described, with a special emphasis on the metadata and data structure for airborne hyperspectral data. For this purpose, an INSPIRE and ISO 19115 conforming example based on the metadata file as implemented at DLR's archiving system DSDA (Deutsches SatellitenDaten Archiv) is shown. In addition, hyperspectral image data with all corresponding files (e.g. quality layers and metadata) is organised using the possibilities of HDF5.

1. INTRODUCTION

Within the FP6 project "HYRESSA" (HYperspectral REmote Sensing in Europe – specific Support Actions) the European-wide Strengths-Weaknesses-Opportunities-Threats (SWOT) in the field of hyperspectral remote sensing and its community user needs were addressed. The main user needs identified were standardization of data processing and calibration, a more transparent calibration processes including quality measures, a joint-European platform for hyperspectral remote sensing, software tools for extracting information from hyperspectral data, education and training and a push to increase the awareness in the added value of hyperspectral remote sensing. EUFAR, as a follow-up project to HYRESSA within FP7, intends to meet these user needs.

EUFAR aims to create an European infrastructure for a diverse set of airborne data, ranging from remote sensing to in-situ measurements. Introduction of standards and protocols facilitates the combination of data from different sources, thus EUFAR's Networking Activity "Standards and Protocols" exists to promote standardization within the project to

- ensure harmonization, integration and inter-operability
- assist new/inexperienced users
- facilitate inter-calibration experiments
- allow comparison of results
- enable easier exchange of data

One way to meet these goals is the introduction of common protocols, which includes regulation for metadata. With well-developed metadata elements, one can find what data is available, how it was processed, and under which conditions it was acquired. Another goal of the harmonization process is the development of recommendations on best practices for data processing, which will help new and inexperienced users. Finally, to meet the request for transparency on calibration and processing actions, the Quality Assurance requirements defined within EUFAR's Joint Research Activity group "HYQUAPRO" will be integrated into the proposed standards.

This paper describes the work towards standards for airborne hyperspectral data, performed within the EUFAR project.

* Corresponding author. Stefanie.Holzwarth@dlr.de

2. EUROPEAN FACILITY FOR AIRBORNE RESEARCH (EUFAR)

2.1 General Overview

EUFAR – the European Facility for Airborne Research - is a 4-year (2008-2012) Integrating Activity of the 7th Framework Program of the European Commission. It is the third framework program EUFAR is taking part in, and during this time, EUFAR has greatly grown in impact and visibility within the airborne science community. While EUFAR was formerly focused primarily on atmospheric research, with HYRESSA (a Specific Support Action supported by the European Commission under the 6th Framework Programme) joining in the latest funding period, the remote sensing community specialized in imaging spectroscopy is constantly expanding within EUFAR.
The EUFAR project aims to integrate operators of instrumented aircraft and experts in airborne measurements in the field of environmental research in the atmospheric, marine, terrestrial and Earth sciences. The current project pools 32 European institutions and companies involved in airborne research, plus 21 instrumented aircraft, including several imaging spectrometers (AHS, AISA, APEX, CASI, ROSIS). These aircraft and sensors are operated through an access scheme known as "Transnational Access". The main objectives of EUFAR are to:

- develop trans-national access to national infrastructures;
- strengthen expertise through exchange of knowledge;
- advance joint instrumental research activities;
- promote the use of research infrastructure, i.e. to young scientists from countries without adequate facilities.

The Integrating Activity is threefold, consisting of Networking Activities (NA), Transnational Access Activities (TA) and Joint Research Activities (JRA). There are 9 NAs, namely

1. Scientific Advisory Committee (N1SAC): Provides the EUFAR consortium with independent strategic recommendations on EUFAR objectives and long term developments.
2. Transnational Access Coordination (N2TAC): The management structure behind Transnational Access. TA is offered for all aircraft and instruments within EUFAR, and can include scientific and engineering support for integration of new instruments, planning of airborne campaigns, and data analysis.
3. Future of the Fleet (N3FF): Evaluation of the existing fleet, identification of gaps, and elaboration of further developments.
4. Expert Working Groups (N4EWG): Exchange and transfer of expert knowledge through various working groups and the compilation of a book on airborne physical measurements.
5. Education and Training (N5ET): Educating and training for new early-stage researchers through training courses and participation in campaigns.
6. Standards and Protocols (N6SP): Development of common protocols in airborne research and supporting researchers with recommendations on best practices.
7. Data Base (N7DB): Centralised gateway to data acquired within EUFAR.
8. E-Communication (N8EC): Provision of webpage to disseminate the EUFAR information.
9. Sustainable STructure (N9SST): Framework for a sustainable EUFAR structure and international cooperation.

The current EUFAR project also includes three JRAs. Two deal with development of new instruments for airborne measurements of atmospheric humidity of cloud drop sizing. The third JRA, related to imaging spectroscopy, is called "HYQUAPRO" and aims at developing quality layers for hyperspectral imagery and data products.

The major focus in the paper will be on N6SP and its interaction with "HYQUAPRO". More information about the EUFAR project can be found at www.eufar.net.

2.2 Networking Activity Standards and Protocols (N6SP)

The Networking Activity "Standards and Protocols" (N6SP) aims to harmonize the different processes and documentation concerning the acquired data within EUFAR to provide a better quality of service. The need for standardization, addressed during the review of the EUFAR project under FP6, is due to the trans-European dimension of EUFAR, with its wide array of instruments and various ways of handling data. The N6SP activity covers the overall suite of EUFAR's airborne data, ranging from atmospheric research to hyperspectral remote sensing.

The work of N6SP is divided into 4 different tasks. The first task is harmonization, accomplished by the development of common protocols. An important issue is to introduce common semantics and terminologies. Thus, in order to prevent ambiguous definitions, a basic glossary on airborne measurements has been created. This glossary, which is continually updated, is available through the EUFAR webpage.
The second task is to address new and inexperienced users of airborne data. To assist them, N6SP has developed recommendations on best practices and state-of-the-art software for airborne data pre-processing. These recommendations are documented on the EUFAR website in a report on software performance, software availability and adaptability and a catalogue of best practices for the different steps of data processing.
The third task is development of software toolboxes for further processing of the acquired data and to improve the accessibility of higher-level data products. These toolboxes are available for scientists, operators and users in the European and international community. Accurate and robust algorithms for information extraction are implemented in the processing toolboxes. Users can access the source codes, executables and manuals via the EUFAR webpage.

The fourth task of N6SP is the development of standards for real-time data exchange and data links. Real-time data exchange between aircraft and ground stations over satellite communication links is in its early stages of adoption by many airborne research institutions. Real-time communications, position, meteorological and camera data sent over these links provide essential situational awareness for scientists observing the flight from ground stations. These systems will become more and more critical to airborne research in the near-future, and to ensure the interoperability of these systems between operators, N6SP makes recommendations for the formats of real-time data transmissions, both within the aircraft, but also between aircraft and ground. Since this topic is mainly of interest for the atmospheric scientific community, it will not be further discussed within this paper.

The introduction of common protocols, recommendations on best practices and the provision of software toolboxes shall promote the comparison of results and enable an easier exchange of data as well as facilitate the different processes within EUFAR for operators and users. In the following sections, the design and implementation of tasks 1-3 will be described in more detail.

3. COMMON PROTOCOLS

After an extensive review by the EUFAR community, the first version of common protocols has been published. These protocols are applied within the fields of flight campaign planning, data processing and quality measures, data distribution and cataloguing and real-time data transfer. Wherever possible, N6SP tries to include existing protocols, e.g. concerning the flight campaign planning existing FP6 EUFAR and FP6 HYRESSA documents have been extended and harmonized. A further example is the adoption of the standards of the US-based Interagency Working Group for Airborne Data and Telemetry Systems (IWGADTS) for N6SP´s real-time data transfer recommendations.

Since the procedures for data processing are instrument/system dependent, binding guidelines and protocols are not feasible. One proposed way to implement a common framework for data processing –apart from the best practice approach- is the definition of common metadata. These standards for metadata include descriptions of the different image and instrument processing steps, which enables a direct comparison between different data sets. In this context, a standardized metadata set is already given by the INSPIRE directive (Commission regulation No 1205/2008) with a minimum set of information for spatial data.
Quality reporting is also achieved by the implementation of corresponding metadata. Within EUFAR JRA2 “HYQUAPRO”, common harmonized data descriptors and quality layers for hyperspectral image data have been selected (see also “Towards agreed data quality layers for airborne hyperspectral imagery” by Martin Bachmann et al.). It has been decided that the data descriptors should be part of the common set of metadata for imaging spectrometry data within EUFAR.
The specific metadata components are further described in 3.1.
For data distribution and cataloguing, EUFAR’s N7DB activity, which is coordinated by the British Atmospheric Data Centre (BADC), provides a gateway to data acquired within EUFAR. Their main purpose is to organize and disseminate the data in a common, easily searchable format. For maximum compatibility between datasets and amongst operators, the intention is to converge towards two common data formats: NetCDF for atmospheric data and HDF for remote sensing data, as both formats are widely used in the airborne atmospheric and remote sensing science community, respectively. Thus, the common protocol for data distribution and catalogue is given by the specifications of N7DB, with N6SP supporting this by providing guidelines for the data structure, e.g. for the design of the HDF-file (see 3.2).

3.1 Metadata

The metadata component of the common-data processing framework ensures accurate quality indicators (as defined in “HYQUAPRO”), aids in data storage, retrieval and processing and helps users understand datasets. The common protocols for data processing and quality measures have been merged in a common set of metadata for hyperspectral image data, which should conform to the INSPIRE directive. Since this is only a minimum set of metadata components, it has been expanded by additional quality indicators (data descriptors), as agreed on within “HYQUAPRO”. The INSPIRE Metadata regulation is mainly based on ISO 19155-2, and includes basic information on the data set or series. All metadata elements as agreed on within the common protocols for EUFAR are described in Table 1, including a link to the corresponding reference, whereby e.g. INSPIRE 1.1 is synonymous with Commission regulation No 1205/2008, Part B 1.1 and “HYQUAPRO” DD 6 is synonymous with Data Descriptor 6 as defined in “HYQUAPRO” (see also “Towards agreed data quality layers for airborne hyperspectral imagery” by Martin Bachmann et al.).

Metadata Element	**Description**	**Reference**
Resource Title	Project/survey acronym	INSPIRE 1.1
Resource abstract	Project/survey abstract	INSPIRE 1.2
Resource type	Dataset or series	INSPIRE 1.3
Resource locator	Web link to data / DB	INSPIRE 1.4
Unique resource identifier	File name (unique)	INSPIRE 1.5
Resource language	Language used (usually eng)	INSPIRE 1.6
Topic category	Main scientific field (coarse description)	INSPIRE 2.1
Keyword value	Subject (more detailed description)	INSPIRE 3.1
Originating controlled vocabulary	If the keyword value originates from a controlled vocabulary	INSPIRE 3.2
Geographic bounding box	Geographic extend	INSPIRE 4.1
Temporal extend	Date/interval of data acquisition	INSPIRE 5.1
Date of publication	Date of data publication (e.g. entry into DB)	INSPIRE 5.2
Date of last revision	Only valid, if data has been revised	INSPIRE 5.3

Date of creation	Date of data processing	INSPIRE 5.4
Lineage	Statement on process history and/or overall quality	INSPIRE 6.1
Spatial resolution	Ground sampling distance	INSPIRE 6.2
Conformity – specification	Citation of specification to which resource conforms	INSPIRE 7.1
Conformity – degree	Degree of conformity	INSPIRE 7.2
Conditions applying to access and use	Conditions for access and use of spatial data set	INSPIRE 8.1
Limitations on public access	Information on access limitations and the reasons for them	INSPIRE 8.2
Responsible party	Contact information of the organisation responsible for data	INSPIRE 9.1
Responsible party role	Role of the responsible organisation (e.g. data provider)	INSPIRE 9.2
Metadata point of contact	Contact information of the organisation responsible for metadata	INSPIRE 10.1
Metadata date	Date of metadata creation	INSPIRE 10.2
Metadata language	Metadata language (usually eng)	INSPIRE 10.3
Basic sensor characteristics:		HYQUAPRO DD 6
Scan principle	e.g. whiskbroom	
Spectral range	e.g. 450-2480nm	
Spectral bandwidth	e.g. 15 nm	
No. of bands / binning (if applicable)	e.g. 125	
Total Field of View (FOV)	e.g. 61.3°	
Inst. Field of View (IFOV)	e.g. 2mrad	
Pixels per scanline	e.g. 512	
Radiometric resolution / quantization	e.g. 14bit	
File name – raw data	Raw data name (might be different from unique resource identifier)	HYQUAPRO DD 7
File name - quality layers	File names of quality layers	HYQUAPRO DD 8
Calibration laboratory	e.g. CHB, DLR Oberpfaffenhofen, Germany	HYQUAPRO DD 9
Date of radiometric calibration	DD.MM.YYYY	HYQUAPRO DD 10
Date of spectral calibration	DD.MM.YYYY	HYQUAPRO DD 11
Radiometric calibration file used	Filename of radiometric calibration file	HYQUAPRO DD 12
Radiance unit + scaling	e.g. W / m² sr µm	HYQUAPRO DD 13
Platform	Aircraft call sign	HYQUAPRO DD 15
Sensor	e.g. APEX	HYQUAPRO DD 16
GPS/IMU	e.g. Applanix POS AV 410, DLR Oberpfaffenhofen, Germany	HYQUAPRO DD 17
Spectral mode	e.g. mode 1	HYQUAPRO DD 18
Frame rate / integration time	Statement in Hz	HYQUAPRO DD 21
Overall heading	Statement in degree (range 0-360, west = 270°)	HYQUAPRO DD 22
Overall altitude ASL	Flying altitude above sea level in meter	HYQUAPRO DD 23
Solar zenith during acquisition	Solar Zenith: range 0-90, sunrise = 90°	HYQUAPRO DD 24
Solar azimuth during acquisition	Solar azimuth: range 0-360, North = 0°, East = 90°,…	HYQUAPRO DD 24
Report on anomalies in data acquisition	e.g. cloud cover	HYQUAPRO DD 25
Processor ID, SW nameDDs & versions	e.g. dims_ares Version 1.2, DLR PAF	HYQUAPRO DD 27
Synchronization problem	Problems during synchronization of image data with navigation data	HYQUAPRO DD 29
Method of interpolation	Method of interpolation used for geometric correction, e.g. bilinear	HYQUAPRO DD 30
Confidence in atmospheric correction		
model	Confidence in atm. corr. from model itself	HYQUAPRO DD 31
comparison with ground measurements	Confidence in atm. corr. due to comparison with ground measurements	HYQUAPRO DD 32
Information on DEM	Information on DEM (e.g. resolution, accuracy,…) used for processing	HYQUAPRO DD 33
Critical BRDF geometry	Comment on critical BRDF geometry within the scene	HYQUAPRO DD 34
Pixels affected by saturation	Pixels affected by saturation in spatial/spectral neighbourhood	HYQUAPRO DD 35

Table 1. Metadata elements for hyperspectral image data as agreed in EUFAR N6SP

3.2 Data Structure

Currently, providers of hyperspectral image data use varying data formats and/or structures to distribute and archive their data. The intention to harmonize the data structure led to the proposal of using HDF as a common format.

The Hierarchical Data Format (HDF) –developed at the National Center for Supercomputing Applications (NCSA)- is a data model, library and file format for storing large volume and complex data in an efficient and flexible way. A primary advantage is that HDF is a self-describing / self-contained data format. It is supported by multiple platforms (e.g. Windows and Unix) and software APIs including C/C++, JAVA, MATLAB, IDL and Python. HDF is the standard format for NASA's Earth Observing System (EOS), where all data and information of a day's mission (e.g. swaths, grids, in-situ data, instrument metadata) can be stored in a single file. NASA also offers associated tool examples and scripts to view, handle and access HDF-EOS files (http://www.hdfeos.org/index.php).

With HDF, data can be organized in a tree-like -or even more complex- structure, which is the major advantage when it comes to handling the huge variety of data and information (e.g. image data, text files, metadata) associated with airborne hyperspectral image data.

3.3 Implementation of Protocols to DLR's PAF

At DLR, the process of implementing the common protocols on metadata and data structures to its Processing and Archiving Facility (PAF) served as a test case to evaluate the typical workload and allowed development of tools for the conversion of typical data structures (i.e. ENVI bsq/bil) used in hyperspectral remote sensing to the standardized HDF5 format. The report on the implementation and the resulting files can be used as an example and guideline within EUFAR.

The proposed design based on DLR's hyperspectral data tries to fulfil the following requirements:

- Compliance with EUFAR's protocol on metadata including the INSPIRE and ISO 19115 conformity
- Inclusion of all quality layers and data descriptors, which were agreed upon within EUFAR's "HYQUAPRO"
- Design of a user-friendly and self-explanatory layout for the different data using HDF-5
- Mapping of commonly used data structures to simplify the conversion from one system to the other

The use of the guidelines to create INSPIRE metadata ensures that the metadata is not in conflict with ISO 19115. However, full conformance to ISO 19115 implies the provision of additional metadata elements which are not required by INSPIRE. At DLR the complete list of metadata has been implemented to the PAF using XML, which results in an INSPIRE and ISO 19115 conforming metadata file. In addition all harmonized data descriptors of "HYQUAPRO" are included as well. The metadata file is embedded in DLR's multi-mission infrastructure DIMS (Data Information and Management System) via item information file (xml). A section of the metadata information for the Level 2 HyMap data as displayed in DLR's DIMS is shown in Figure 1.

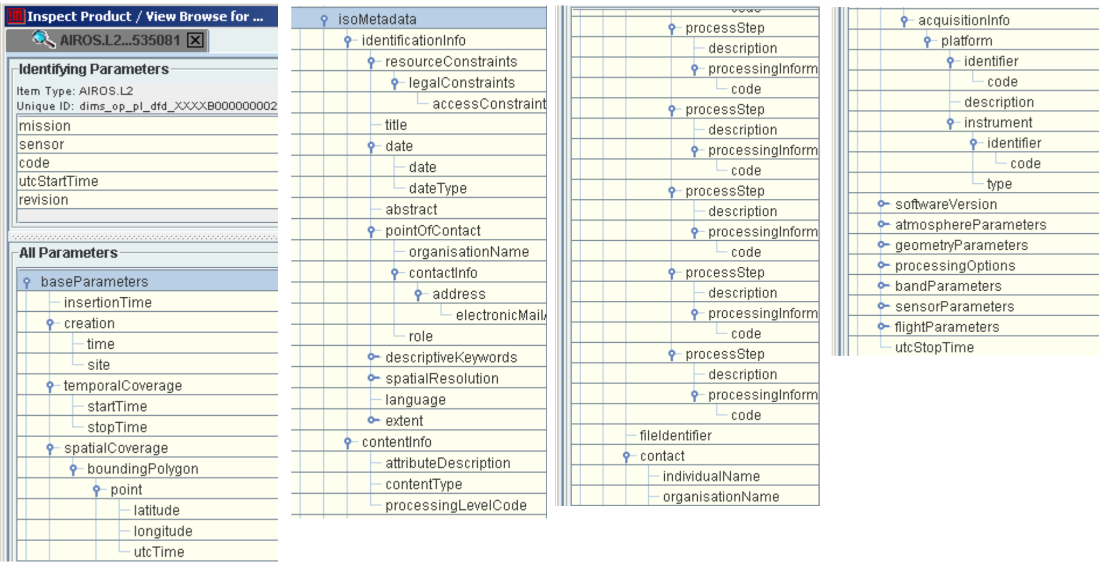

Figure 1. Part of DLR's metadata information as displayed in DIMS

The design for the HDF file contains processed Level2 data with all corresponding information and data. With the possibility of storing data in a tree-like structure, six different groups are defined within the root group to address all different categories, namely:

- Mission information
- Navigation data
- Processing information
- Quality indicators
- Sensor data
- Auxiliary data

"Mission information" includes information about the campaign (e.g. project information), the instrumentation setup during the survey (e.g. sensor characteristics and calibration information), the platform and the survey itself (e.g. acquisition information). The category on "navigation data" contains the raw and processed platform spatial data (position and attitude during data acquisition). Information about the processing parameters, the DEM and the different software versions used for data correction is organised within the subgroup "processing information". "Quality indicators" contains all the data quality descriptors and quality layers, subdivided into atmospheric data (e.g. visibility, water vapour), covariance and histogram information, image geometry (e.g. azimuth and zenith angle), mean spectrum and quality flags (e.g. saturated pixels, cloud coverage). The image data (at-sensor radiance and ground reflectance) is stored bandwise within the "sensor data" including all necessary metadata. In addition all housekeeping data (e.g. dark current) are arranged within "sensor data" as well. "Auxiliary data" is intended for additional airborne data and field data acquired in parallel to the airborne hyperspectral survey. Figure 2 gives an overview of the design of the proposed HDF-file.

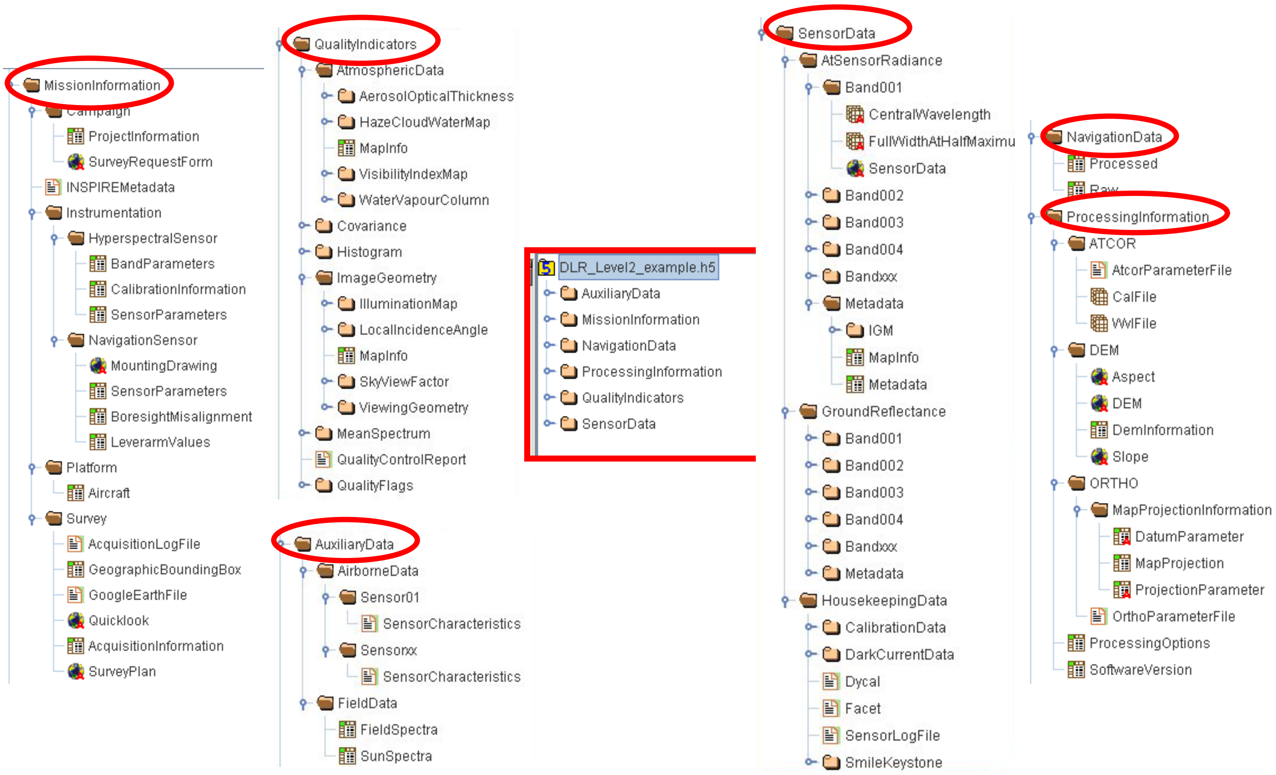

Figure 2. HDF design of DLR's Level 2 data

Currently, DLR is working on a program to include and import all existing data in the HDF file described above. The main function will be the ENVI bsq/bil to HDF transformation, which will be made available through the N6SP toolbox.

4. BEST PRACTICES

Procedures for data processing tend to be instrument/system dependent, therefore binding guidelines, such as common protocols for data processing are not feasible. Thus harmonization using a "best practice" approach was selected. Users and operators are supported with recommendations on system correction, atmospheric correction and geometric rectification of airborne hyperspectral imagery. N6SP summarizes the most important issues when it comes to Level1 and Level 2 processing of the data in a compendium published through the EUFAR web page.

On top of the assembly of best practices, the different software packages for atmospheric and geometric correction have been reviewed and tested, where applicable. The results on software performance, and the information on availability and adaptability summarised within another EUFAR report. It is planned to publish the work of N6SP also within the platform of EUFAR's expert working group on data processing.

5. TOOLBOX

The toolbox component of the framework is currently under development and is called EGADS (EUFAR General Airborne Data processing Software). The Python-based EGADS will include algorithms considered by the expert working groups as best practice for processing and manipulating raw and higher-level probe and image data. Input and output routines will be included for EUFAR recommended file formats (see 3.3). Currently, there are only 10 basic thermodynamics and 6 microphysics algorithms available through EGADS. Shortly it will be complemented by a metadata creator, the HDF5 converter and the implementation of the "HYSOMA" toolbox developed within "HYQUAPRO" (see also "Development of the HYperspectral Soil MApper (HYSOMA) interface: Methods, products, and validation" by Sabine Chabrillat et al.). The call for additional expert-provided algorithms is still open.

6. CONCLUSIONS

The work of EUFAR's N6SP towards standards for airborne hyperspectral image data has been described within this paper. Up to now, a first version of common protocols has been published within the EUFAR community. In addition, N6SP aims for harmonization by providing recommendations on best practices for data pre-processing and by the assembly of different tools for processing and manipulating image data.

The strong involvement of the EUFAR community (e.g. expert working groups) through extensive reviews of the different documents proposed by N6SP and the inclusion of previously existing standards aids the acceptance of the developed standards and protocols. Currently, the protocols are mainly used as best practice, but full implementation is planned within the next EUFAR

program. Additionally, the establishment of the proposed standards and protocols outside the EUFAR community is intended, presenting the results to other international working groups dealing with standards in remote sensing (e.g. CEOS, EnMAP) and atmospheric research (e.g. IWGADTS, ISPRS).
All documents generated within N6SP are openly available at the EUFAR web-page (www.eufar.net).

REFERENCES

Bachmann, M. et al., 2010. FP7-JRA2-DJ2.2.1 – Generic Quality Indicators and Quality Layers. EUFAR Report

Holzwarth, S., Freer, M., 2011. FP7-N6SP-DN6.1.3 - Version 1.0 of common protocols. EUFAR Report

Holzwarth, S., Wang, X., Richter, R., Bachmann, M., Freer, M., 2010. FP7-N6SP-DN6.2.2 - Software test an comparison report & Documentation on best practices. EUFAR Report

Holzwarth, S., Bachmann, M. Freer, M., 2011. FP7-N6SP-DN6.3.1 - Harmonization of data formats. EUFAR Report

Reusen, I. et al., 2008. HYRESSA - HYperspectral REmote Sensing in Europe – specific Support Actions. Final Report

ACKNOWLEDGEMENTS

The research leading to these results has received fundings from the European Community's 7th Framework Programme (FP7/2008-2012) under EUFAR contract n° 227159.