Computers & Geosciences 162 (2022) 105076

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

Increasing the interoperability of snow/ice hyperspectral observations

Sabina Di Franco[a,*,1], Roberto Salzano[b,1], Enrico Boldrini[b], Rosamaria Salvatori[a]

[a] CNR-Institute of Polar Sciences, Italy
[b] CNR-Institute of Atmospheric Pollution Research, Italy

ARTICLE INFO

Keywords:
Snow
Field spectroscopy
Metadata
Reflectance
Data model
Interoperability

ABSTRACT

This study aims to set up a metadata profile useful for preparing an interoperable dataset containing snow and ice hyperspectral measurements. The proposed Snow and Ice Spectral Library (SISpec) scheme was prepared for sharing a data collection focused on Antarctica, including 70 observations. Following the perspective to grant "open access" to such a dataset, we found a compromise between the ERC (European Research Council) guidelines, the FAIR (Findability, Accessibility, Interoperability, and Reuse) Data principles defined by the RDA (Research Data Alliance), and the GEO (Group on Earth Observation) Data Sharing Principles. The ISO (International Organization for Standardization) standard 19115 was chosen as the standard framework for describing SISpec. When the available metadata scheme was not sufficient or suitable, metadata extensions or new detailed metadata components were created to be compliant with the ISO 19115 standard. We also considered the INSPIRE (Infrastructure for Spatial Information in Europe) requirements and the result is a metadata model that can be useful to share SISpec metadata both in the European and international contexts. Particularly detailed metadata sections and elements were created for describing spectral signatures and microphysical snow parameters.

1. Introduction

Snow and ice cover a large portion of our planet and they are highly vulnerable to climate changes. To monitor the spatial and temporal extensions of these surfaces it is necessary to use both field and remote sensing data. The snowpack, defined as a layer of ice crystals of diverse sizes and shapes (Rees, 2006), and the ice surface can be recognized in the remotely sensed images considering the peculiarities of their spectral behaviors. While light-absorbing impurities (lithogenic minerals dust, algae, soot) impact more on the visible wavelengths range (Warren, 2019), the snow properties in the short-wave infrared domain (1400–2500 nm) are dominantly affected by the size and shape of the ice crystals (Dominé et al., 2006). Compact ice surfaces have a spectral behavior like the snow surface in the visible range and they absorb almost completely the incident radiation in the short-wave infrared ranges (Rees, 2006). The optical behavior of snow and ice is therefore a key component of the knowledge required for calibrating and validating satellite data. The collection of ground-based observations, provided by portable spectroradiometers, is an ideal data source, which is extremely interesting if those hyperspectral observations were obtained during field campaigns located in remote regions. There is unfortunately still a gap between observations and modeling of the snow cover and the challenge consists in introducing the microphysical properties of the different snow layers, with particular attention to the snow grain habit mixture (Dang et al., 2016; He et al., 2018; Saito et al., 2019).

A set of this kind of data, collected in Antarctica, is partially included in the spectral library (SISpec) aimed at supplying the interpretation of Landsat imagery (Casacchia et al., 2002). This library was adequately coupled to ancillary information and the dataset was distributed in different ways, both as an archive and browsed through a website (Ghergo et al., 2000) but to increase the value of such data, it is now necessary to apply the best practices to optimize interoperability, and reusability of the shared data. Starting from the goal to figure out a better solution to share the SISpec data, we did a state-of-the-art analysis of the metadata of some spectral libraries to define a metadata set that would not only increase the interoperability of SISpec but also define metadata that specifically describes snow.

The proliferation of data characterizes our age, but we risk what is described in the phrase "We are drowning in data but starved for information" (Brown, 2014). This has become a common saying in this

---

* Corresponding author.
E-mail address: sabina.difranco@cnr.it (S. Di Franco).

[1] equally first authors.

https://doi.org/10.1016/j.cageo.2022.105076
Received 8 July 2021; Received in revised form 15 December 2021; Accepted 11 February 2022

time of "big data", the vast amount of scientific data available has increased the need for rules and principles for data management and organization. Many scientific data are organized as a sort of "hidden" archive whose discovery and recovery is a kind of scientific treasure hunt. The effort that is being aimed at in these years is to reach an efficient discovery of data and reuse of research results. Moreover, the scientific community has decided upon principles to grant "open access" in activities as the European Research Council (ERC) guidelines, the principles of findability, accessibility, interoperability, and reusability (FAIR) data, the Group on Earth Observations (GEO) Data Sharing Principles. Some data-sharing initiatives focused on field spectroscopy have lately appeared in the remote sensing community such as NASA's EOSDIS (Earth Science Data and Information System), the LTER (Long Term Ecological Research) network, the Australian Terrestrial Ecosystem Research Network (TERN), SpecNet (Rasaiah, 2015). All these experiences have in common that to make it possible to share data in whatever form it is produced and organized, there is a need for rules to be followed (Wilkinson et al., 2016). The value of shared data increases if datasets are findable or discoverable, enabling the attribute-based search through a detailed description of both data and metadata. Particular attention will be, therefore, necessary to prepare data and metadata that are accessible, and retrievable in a variety of formats suitable for humans and machines alike. A description of metadata that is interoperable must follow the guidelines of the communities that use it and follow a well-defined vocabulary. Finally, it must be taken into account also the reusability when the definition of essential, recommended, and optional metadata should be machine processable and checkable, the use should be straightforward, and data should be referenced to support data sharing and recognition of data value itself. All these requirements refer to the FAIR and GEO principles for the management of scientific data, and it is critical to consider these principles valid for both humans and machines.

The collection of hyperspectral measurements, firstly based on acquisitions performed under laboratory conditions in the 60s, and including later in-situ measurements, is defined as a spectral library (Clark et al., 1990). Spectral libraries (SL) include additional ancillary information (such as chemical compositions, measurement conditions, etc.) to support the identification of many different materials. The first spectral libraries, like those provided by USGS (Clark et al., 1993) and by the NASA JPL (Jet Propulsion Laboratory) (Grove et al., 1992), were dedicated to mineral and rock as an extension of the mineralogical and petrographic analysis and only later were they used to interpret the multi and hyperspectral images. In the last 20 years, spectroscopic techniques have acquired a growing relevance to study the planet's different surfaces, especially because spectroscopic measurements are rapid, precise, and inexpensive (Rasaiah et al., 2014; Viscarra Rossel et al., 2016) and the need for ground-truth observations for satellite data calibration and validation became a major requirement (Bojinski et al., 2003). In the scientific literature, several SL on minerals, soils, rocks, vegetation, and man-made materials are well known. In those SL spectral signatures are collected in different wavelengths that usually cover the wavelength intervals specific to passive optical images (Viscarra et al., 2016, Nidamanuri and Ramiya, 2014, Rasaiah et al., 2012, 2014, Jiménez and Delgado, 2015, Fang et al., 2007, Herold et al., 2004). SLs were in some cases updated as data services where collections can be browsed online (Meerdink et al., 2019; Kokaly et al., 2017) and the spectra directly imported into software packages (QGIS, ENVI) using proprietary or text file formats.

Recently, in addition to the increasing number of datasets, the libraries have evolved to novel paradigms such as Spectral Archives and finally to Spectral Information Systems (SIS). The SIS paradigm, applied for the Swiss Spectral Input/Output (SPECCHIO) (Hueni et al., 2020) and EcoSIS (Wagner et al., 2019) collections, is based on the development of complex systems aimed at managing large datasets and sharing them through application programming interfaces (API) or web portals. While at first spectral signatures were collected in SLs archives used by individual researchers in their workplaces, the need to share information with a larger audience and advances in technology have led to a greater need for sharing.

Information contained in spectral libraries to be shareable and interoperable must be organized. Key components for describing data are metadata since, as the term implies, data about data, are "structured data about everything that can be named" (DCMI, 2021). The metadata profile represents, in this case, the digital content using a set of predefined attributes. This information is useful for the description of the investigated materials and of the database that contains them (Gilliland, 2016). Moreover, metadata is common in information systems and occurs in many forms and is essential to the systems functionality, describes the content, enabling users to find what they are looking for, record important information about data, and helps share them. (Riley, 2017). Well-made metadata schemas should be designed according to international rules and standards such as the ones from the International Organization for Standardization (ISO), the Dublin Core Metadata Initiative (DCMI), or the Infrastructure for spatial information in European Community (INSPIRE). The INSPIRE directive 2002/2/EC imposes strict rules for European countries regarding drafting and sharing of metadata, to enable interoperability across the European Union and beyond. The ISO 19115 standard supports (ISO 2014, 2019) these practices describing structured relations between included elements, and additional components mentioned in other standards of the 19100 series, through the adoption of the Unified Modeling Language (UML) and the eXtensible Markup Language (XML) useful for the implementation of metadata that support the machine-to-machine exchange. The ISO 19115 identifies the minimum metadata set (core Metadata) required for describing datasets and it provides, associated with additional more specific references, guidelines for completing the description of attributes concerning: the base dataset metadata; the acquisition description; and, eventually, the characterization of the surface. While the first base component is already standardized following the ISO and INSPIRE guidelines, the acquisition section is only partially described by the ISO (2019) The information about the observing methodology requires the collection of many details that are domain-specific and, in our case, the experimental setup must consider a scheme provided in literature by different sources (Rasaiah et al., 2014; Jiménez et al., 2014; Hueni et al., 2011). Although the available schemes are an ideal background for describing, with large flexibility, most of the possible experimental setups, all of them are unfortunately user-oriented not aligned to the INSPIRE directives aimed at discovering datasets. Similarly, a snow extension is not already described in the literature for the discovery of snow-related datasets, but a good backbone is represented also in this case by the Canadian Avalanche Association Markup Language (CAAML), described by (Haegeli et al., 2010). This standard was born for the electronic representation of information pertinent to avalanche safety operations, but it is also a tool for sharing data with the community. From this perspective, the alignment of this scheme, and consequently of the International Classification for Seasonal Snow on the Ground Fierz et al. (2009), to the ISO (2014; 2019) and INSPIRE guidelines (INSPIRE, 2007) is one of the gaps that this work aims to fill.

The use of standard attribute naming conventions is an additional point that must be addressed. Looking at the available SLs or SISs, none of them follows a standardized naming convention such as the Climate and Forecast (CF) convention (Eaton et al., 2020) or the Attribute Convention for Data Discovery (ACDD) guideline (ESIP, 2020). These resources contribute to a reliable description of what each variable means, and of the spatial and temporal properties of the data. Several examples of discrepancies between attribute descriptions are possible considering core metadata such as the "institution", "id" and other basic attributes. This mismatch, considering that we are referring to the machine-to-machine interaction, is a major issue if we include the acquisition information ("instrument", "sensor", etc.).

Finally, the selection of the most appropriate file format is a major discussion element since, as in our case (SISpec) a dataset including observations of the snow surface using field spectroscopy represents a complex object difficult to be handled using unstructured file formats (such as an ASCII file). Considering that these measurements are a geospatial data feature and that a self-described data format supports a higher level of interoperability, a reasonable choice seems to be the data model based on the Network Common Data Form (NetCDF) file format (Rew et al., 1997). The NetCDF is a set of machine-independent software libraries and data formats that provide support for the creation, access, and sharing of array-oriented scientific data. This data format is an Open Geospatial Consortium (OGC) standard maintained by the UNIDATA community and is already recognized by several agencies and institutions (Domenico, 2011).

Unfortunately, the approach "one-size-fits-all" is unhelpful and restrictive. Having this in mind, we decided to proceed systematically to approximate the best possible description of a Snow and Ice Spectral (SISpec) scheme starting from the most common metadata standards to describe a specific library of spectral signatures of snow. To reach the goal we created a metadata profile compliant with current ISO standards and INSPIRE directive, starting from snow descriptions collected in the field to support radiometric measurements. The major goal of the proposed scheme is to approach the snow cover as a complex mixture of different crystal shapes and dimensions, using a harmonized classification and a standardized way to describe the microphysical behavior of the observed snow cover.

This study presents a methodology based on existing hyperspectral measurements of different snow surface types and on available interoperability standard resources (data format and guidelines). The proposed approach covers the survey about attributes required by interoperability standards, the definition of encoding specifications, and consequently the description of a specific metadata profile. The novelty is based on the definition of a snow-related extension aimed at obtaining fully-described metadata about the base data information, the instrumental setup and above of all about the surface snow description. The contribution of such a profile is not only investigated in terms of interoperability between spectral libraries but also as a trigger for the ingestion of snow hyperspectral observations into spectral information systems.

2. Methodology

The design of a metadata profile requires a test dataset, in this case, based on field hyperspectral measurements, which supports the definition of the necessary metadata components.

2.1. *Hyperspectral field measurements*

We considered the first version of the SISpec library (Ghergo et al., 2000) where spectroradiometric measurements of snow surface were acquired on the field as well as some snow microphysical properties (snow grains shape and size). This dataset included hyperspectral measurements of snow and ice targets in the spectral range between 350 and 2500 nm. The measurements were collected in Antarctica (Casacchia et al., 2002) with the portable spectroradiometer, Fieldspec FR (ASD Inc. Boulder, CO, USA) as absolute reflectance, i.e., as the ratio between the radiation reflected from the surface and the radiation reflected from a Spectralon reference panel.

The measurements were performed in smooth and open areas, far from the mountainous reliefs, sufficiently wide (about 100 × 100 m) to be easily identified on satellite images with a spatial resolution of 30 m per pixel. Particular attention was paid to selecting areas with homogeneous microphysical characteristics of the snowpack. For each measurement site, the distinctive characteristics of the snowpack were observed and recorded, such as the shape and size of the snow grains. The density, hardness, and temperature of the surface layer of the snowpack were also measured. Local weather conditions were also noted during teach measurement session. The adopted standard for the description of the characteristics of the snow cover at the beginning of the data collection was from Colbeck et al. (1990) and then it was updated to the classification from Fierz et al. (2009).

2.2. *The metadata profile*

The creation of a specific metadata profile for snow and ice was based on two core components: the metadata and the encoding technical specification. The first one, the metadata technical specification, where the metadata model is described using UML diagrams and tables. The metadata sections and elements that are part of the profile are listed (such as title, creator name, place, and date of creation, used to collect information about a resource), including both the ones already existing from international metadata standards and the newly introduced ones. The second one, the encoding technical specification, where technical details are present on how to encode in a specific machine format (e.g., XML, NetCDF, JSON, etc.) metadata documents that are compliant with the metadata profile. The specification may also include schemas, templates, or online tools to help assure compliance with the metadata profile.

The specific metadata set to describe the content of the SISpec library was based firstly on general-purpose standards: the Dublin Core for the general resources (Neiswender and Montgomery, 2010); the ISO standard for the geospatial information (ISO, 2014); and the acquisition standard (ISO, 2019), which is the extensions focused on the acquisition and the processing for the general description of the considered observations.

The appropriate suite of metadata schemas was carefully chosen, to best describe and provide access to databases. To complete the SISpec metadata profile, we chose to adopt the ISO 191xx series of metadata standards (ISO 2014, 2016, 2019). The first two ISO standards define a wide variety of metadata elements for describing geographic information. Part 1 contains fundamental elements (such as information about the metadata itself, citation, spatial-temporal extent, lineage, reference system, data quality), while part 2 focuses on elements for describing acquisitions and processing (such as environmental conditions, instrument, objective, operation, plan, platform, requirements). All the elements described in the two documents have been used as part of the SISpec profile. These elements represent the ISO 19115 comprehensive metadata element set, comprising both the ISO 19115 minimum mandatory components (with mandatory obligations) and all optional elements (in some cases mandating obligations). Additional requirements for the SISpec profile come from the European directive INSPIRE, as the SISpec profile aims to be fully compliant with it. Technical guidelines for the implementation of the INSPIRE required metadata for datasets based on ISO (INSPIRE, 2007) has been reviewed to include in the profile all the needed requirements (e.g., include with mandatory obligation selected optional ISO 19115 elements, or add domain constraints). Finally, ISO (2016) has been considered to realize the XML encoding of documents compliant with the SISpec metadata profile.

Finally, we developed a snow-related extension where both regular standards and existing resources were not enough to fulfil our needs and describe the specific content of the SISpec profile. This component was defined following the ISO rules contained in the ISO 19115 Annexes, especially the Annex F "Metadata extension methodology". The backbone of this extension was the classification for seasonal snow (Fierz et al., 2009), which is already represented for data exchange purposes by the CAAML (Haegeli et al., 2010).

Once we selected the standards for the metadata and defined the structure, we dedicated specific effort to drafting attribute naming conventions. This task is strictly associated with the use of the NetCDF data format since it is a self-described data model that requires standard attribute names for having complete interoperability. We referred our attention to the Climate and Forecast convention and the Attribute Convention for Dataset Discovery.

3. Results

We used the SISpec dataset as a testbed for the creation of snow-specific descriptive metadata. We are working to make available the following products as the result of the standardization process:

- The SISpec metadata profile technical specification, where the SISpec metadata model is described with UML diagrams and tables. The metadata sections and elements that are part of the profile are listed, including both the ones from ISO 19115 and the newly introduced ones (supplementary materials: appendix A-MC1)
- The SISpec XML encoding technical specification, where technical details are present on how to draft XML documents that are compliant with the SISpec metadata profile and can be validated according to the SISpec XML schema (based on ISO 19115 part 3 and community extensions); and the SISpec Schematron (supplementary materials: appendix A-MC2).
- The SISpec NetCDF (Network Common Data Form) encoding technical specification, where technical details are available on how to draft a NetCDF file that is compliant with the SISpec metadata profile, being NetCDF a standard for sharing data. This file is presented in Appendix A-MC3: supplementary materials: appendix A-MC3.

3.1. *Description of the snow metadata profile*

The SISpec metadata profile was created considering different components (Fig. 1), which include: base, acquisition, and snow-information components.

The "mandatory" Base metadata elements are attributes mandatory or highly recommended by ISO (2014) to be included in profiles, such as the dataset title, the reference date, the party responsible for the data and metadata, the abstract about its content and so on. These mandatory base attributes are in total 40 elements and 36 of them are mandatory for INSPIRE requirements. Finally, the "comprehensive" Base metadata from ISO (2014) are all the optional elements, 24 in total, that enable users to describe their datasets with full details. The ACDD convention recommends 19 additional attributes (4 are strongly recommended) useful for the data discovery. The "acquisition" component defined by the ISO (2019) completes the global attribute description of the considered dataset with a set of information about the instrument setup and the campaign geometric design. This component includes 13 elements, and 2 attributes are mandatory. The novel component is the snow information extension, which includes additional information about the experimental conditions for measuring the spectral reflectance, the surface description of the target and finally, the variables associated with the different available measurements and the microphysical conditions of the observed snow surface. While about 42 features are associated with the requirements defined by the different field spectroscopy communities (Hueni et al., 2020; Rasaiah et al., 2014; Jiménez et al., 2014), the observed variables needed for describing the snow optical behavior and microphysics are defined by the snow survey community and by the CF convention (23 variables in this case).

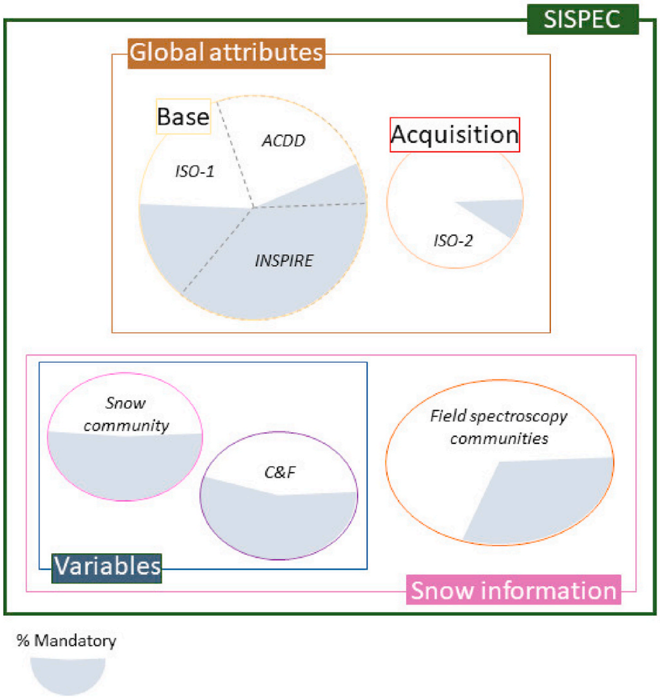


Fig. 1. Representation of the Snow Ice Spectral Library (SISpec) metadata profile and distribution of the metadata components.

The final scheme of the SISpec metadata profile will be composed of three different elements: the metadata base, the acquisition information, and the snow information (Fig. 2).

While the metadata base and the acquisition information are standardized information, mostly about the location, the geographical and time domains, and the instrumental setup, the SISpec information represents the novel component to be described in detail. Referring to the novel extended elements, we created an additional metadata component aimed at including the specific aspects of the studied domain: the snow. We considered 13 variables (5 are mandatory) that can include the specific properties of snow and snow surfaces. The characteristics of the snow were described considering the paper of Fierz et al. (2009). Moreover, we considered in our profile 10 variables (8 are mandatory) derived from the CF convention (Eaton et al., 2020). We considered for the whole profile the use of standardized name alignment that was developed following the Attribute Convention for Data Discovery (ESIP, 2020).

The section dedicated to the snow information (Fig. 3) holds four different elements: i) the surface description, to describe the surface object of the measurements using one string field; ii) the graphic overviews, providing pictorial information (photos) about the target location; iii) the snow conditions during measurements and the specific observations; iv) the observed composition in terms of snow grain size.

The snow grains description was defined by considering the Grain code list (Fig. 4) and identifying the type and relative composition of the snow grains during the measurements. Each observed surface was described, following the International Association of Cryospheric Sciences (IACS) classification (Fierz et al., 2009), considering: the primary classification, which discriminates the different morphological characteristic of the snow crystals (e.g. precipitation particles, machine-made snow, faceted crystals, etc.); the secondary classification, which specifies the subclasses associated with the description of the physical process producing the particles (e.g. precipitation particles-stellar dendrites, faceted crystals-solid faced particle); the percentage, indicating the relative abundance of the grain concerning the total; the size of the grain, quantified as a measure. The metadata profile allows the description of main grain types (more representative) for each observed surface.

The section focused on the snow condition includes different surface descriptors as well as eventually measured parameters. The descriptors are hardness and roughness that are listed in Fig. 5. The quantitative parameters (thickness, density, roughness length and height, temperature, and humidity) require in addition to the measured value, also the description of the used methodology. Moreover, the metadata profile is designed to accept additional conditions, which can be nevertheless described, using the "other Conditions" extension point, present to increase flexibility and usability to the metadata model.

Finally, the acquisition of hyperspectral measurements is included in the components related to the spectral information. The selected parameter is the reflectance value associated with each wavelength

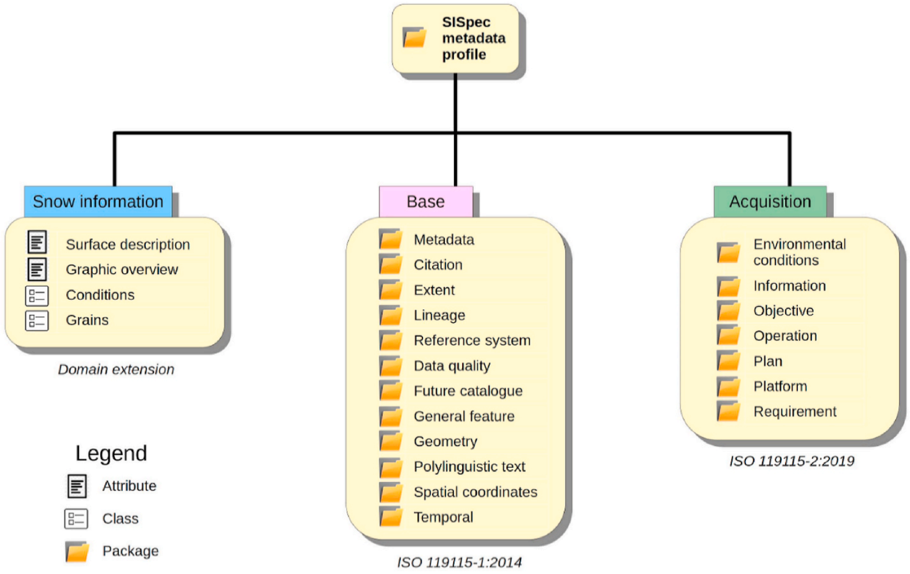


Fig. 2. The SISpec metadata profile and the key components for its description.

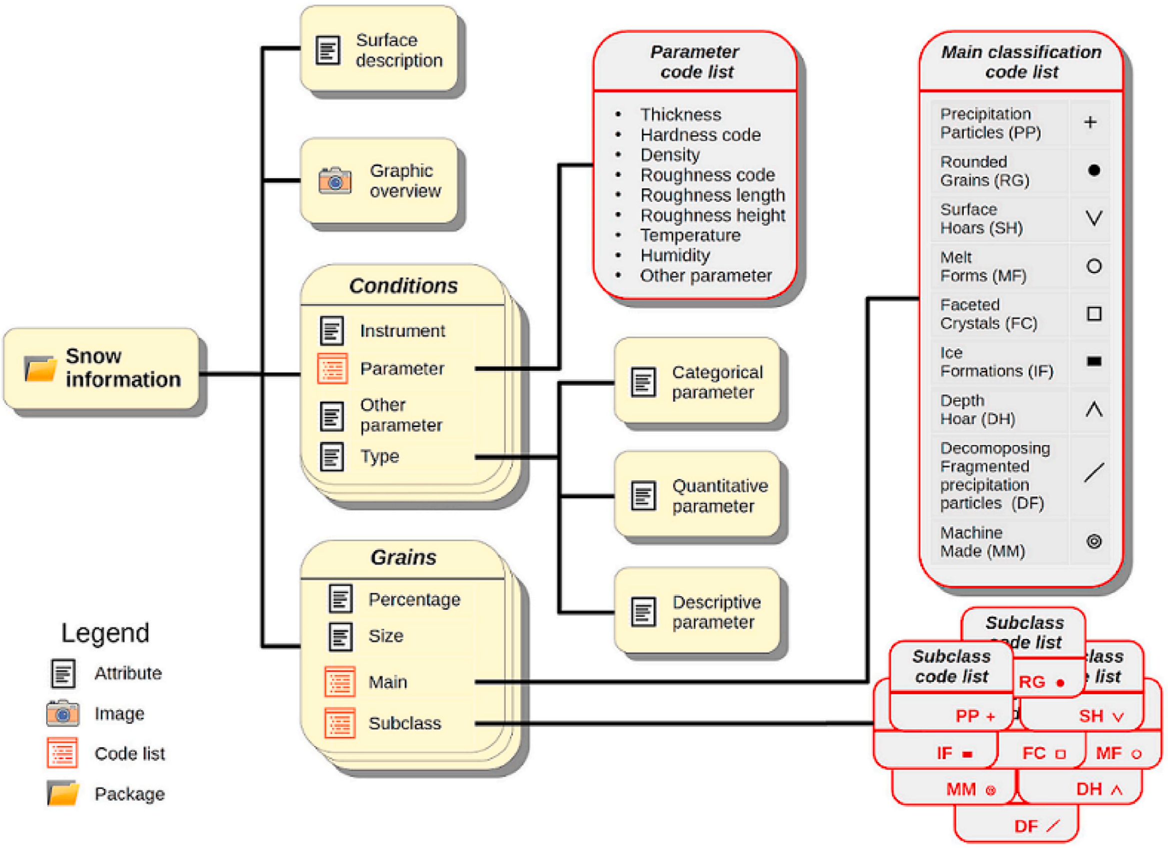
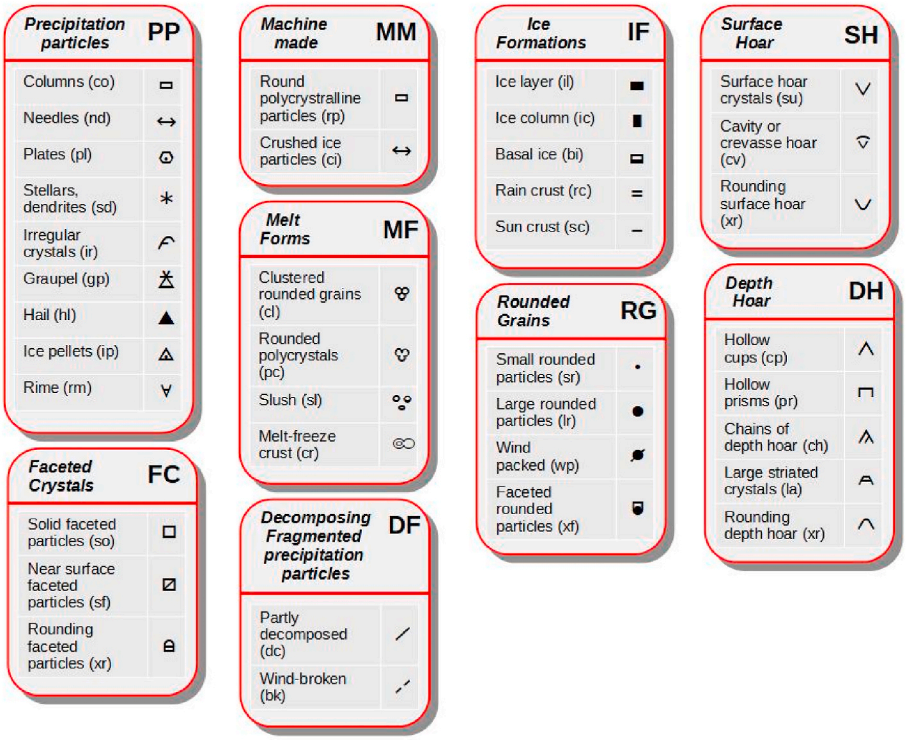


Fig. 3. The detailed structure of the SISpec information.

described in the Instrument description present in the Acquisition Information. Referring to the general metadata component, it is important to state that particular attention was paid to the terminological references for keyword metadatation; specifically, it was chosen to use the EU thesaurus GEMET and the Snowterm thesaurus on snow and ice (Plini et al., 2009).

3.2. *Application of the novel metadata profile to the field data collection*

The impact of the novel metadata description is essentially focused on having hyperspectral measurements, described in terms of experimental setup, associated with the observations about the snow microphysics. The available hyperspectral measurements of the surface snow are generally tagged by labels or comments related to the size distribution neglecting the possibility to consider the crystal shape and the



Precipitation particles	PP
Columns (co)	
Needles (nd)	
Plates (pl)	
Stellars, dendrites (sd)	
Irregular crystals (ir)	
Graupel (gp)	
Hail (hl)	
Ice pellets (ip)	
Rime (rm)	

Faceted Crystals	FC
Solid faceted particles (so)	
Near surface faceted particles (sf)	
Rounding faceted particles (xr)	

Machine made	MM
Round polycrystalline particles (rp)	
Crushed ice particles (ci)	

Melt Forms	MF
Clustered rounded grains (cl)	
Rounded polycrystals (pc)	
Slush (sl)	
Melt-freeze crust (cr)	

Decomposing Fragmented precipitation particles	DF
Partly decomposed (dc)	
Wind-broken (bk)	

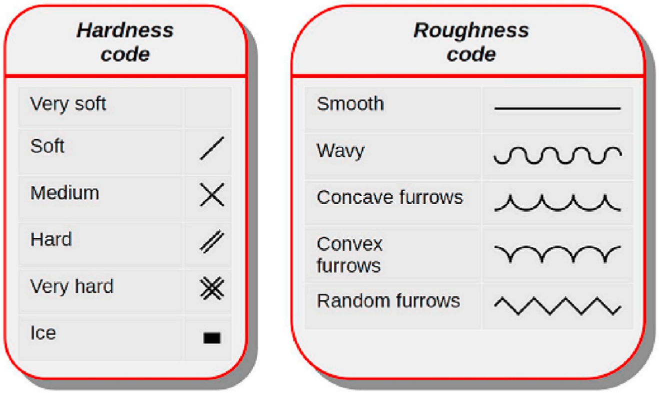
Ice Formations	IF
Ice layer (il)	
Ice column (ic)	
Basal ice (bi)	
Rain crust (rc)	
Sun crust (sc)	

Rounded Grains	RG
Small rounded particles (sr)	
Large rounded particles (lr)	
Wind packed (wp)	
Faceted rounded particles (xf)	

Surface Hoar	SH
Surface hoar crystals (su)	
Cavity or crevasse hoar (cv)	
Rounding surface hoar (xr)	

Depth Hoar	DH
Hollow cups (cp)	
Hollow prisms (pr)	
Chains of depth hoar (ch)	
Large striated crystals (la)	
Rounding depth hoar (xr)	

Fig. 4. Code lists of the snow grain classification based on Fierz et al. (2009).



Hardness code
Very soft
Soft
Medium
Hard
Very hard
Ice

Roughness code
Smooth
Wavy
Concave furrows
Convex furrows
Random furrows

Fig. 5. Code lists of the surface conditions based on Fierz et al. (2009).

complexity of microphysical mixtures. The contribution of the IACS classification (Fierz et al., 2009) highlights the need of combining different crystal shape and size for an appropriate identification of different snow types. Having a rigorous identification of different crystal shapes and size distribution, it is possible to define the processes occurring on the observed snow layer. This information is critical for different disciplines, and it is a major request from communities involved in snow observations and modeling (He et al., 2018; Dang et al., 2016).

Fig. 6 shows some hyperspectral measurements selected in the presented dataset, where the observations highlight large differences between surface types classified in ice, kinetic forms, and equilibrium forms. The difference could be smaller between different equilibrium forms, as in this example, and the IACS classification supports the discrimination between equilibrium growth forms composed both by rounded grains (RG in Fig. 4) with small and large rounded particles, RGsr and RGlr respectively. One of the two showed snow observations is therefore made up of both crystal shapes, where RGsr is the major component and RGlr is the secondary one. The flexibility of the new profile consists also in considering multiple crystal shapes and sizes in the same observations. This feature offers the possibility to describe mixed compositions instead of pure single snow types that rarely occurs in nature, and it enhance to have a detailed description of the observed snow layer.

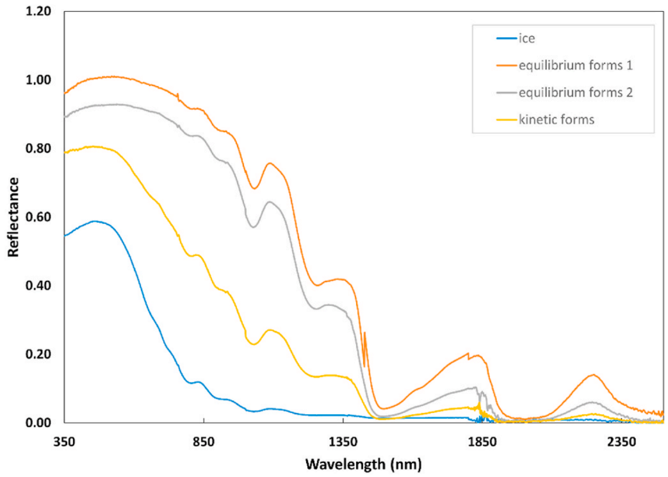


Fig. 6. Examples of hyperspectral measurements of different snow cover types.

4. Discussion

The need for a specific snow-related metadata profile comes out from a review about the availability of such conventions in literature. Considering the availability of spectral libraries (SL), spectral archives (SA) and recently spectral information systems (SIS) focused on different domains, the need for a metadata profile useful for interoperating between different systems is a major requirement. The availability of a snow-related metadata extension strengthens, moreover, the opportunity to connect domain-specific collections (SL or SA) to general-purpose systems (SL, SA or SIS). The evolution of the system paradigm to SIS invokes furthermore the need for standard conventions (ISO and INSPIRE) aimed at increasing the machine-to-machine interaction. The available spectral collections can be grouped in general-purpose and domain specific. The first group includes the SPECLIB SL (Kokaly et al., 2017), the ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) SL (Meerdink et al., 2019), the SPECCHIO SIS (Hueni et al., 2009) and the LUCAS database with its SL (Orgiazzi et al., 2018). The second group includes the INTA SL (Jiménez et al., 2014), and the Global vis-NIR SL (Viscarra et al., 2016). All these datasets are designed for managing observations obtained by laboratory and field instruments. Some of them, like ECOSTRESS and SPECLIB, are characterized by text files, where single measurements (spectra) are coupled to a limited amount (about 20) and not standardized metadata. On the other hand, measurements available in SPECCHIO and INTA are described with a more detailed and standardized metadata profile featuring more than 50 attributes. The most important difference between these groups consists in the use of standardized attributes, which supports the machine-to-machine dialogue, and the availability of aggregated attributes, where information is solely oriented to human interaction. Looking at snow observations, ECOSTRESS, SPECLIB, INTA and SPECCHIO provide few spectra related to the snow matrix, but the description is very coarse and difficult to be considered in specific studies. The presence of such observations in those spectral libraries highlights the importance of having such snow measurements and stresses at the same time the need for more details for describing the snow surface by the International classification defined by Fierz et al. (2009). For example, the ECOSTRESS SL classifies snow observations in the water category, as coarse granular snow, medium granular snow, fine snow, or frost. This size-oriented classification is a limiting factor for snow-related studies since complex mixtures of shape and dimensions are possible. The profile extension supports the description of multiple crystal types, up to three classes are frequent in nature, where the size is coupled to the shape as well as to genesis of such crystals, with more than 30 classes. This description is not only limited to the snow crystallography, but it also opened to harmonise additional microphysical properties, such as hardness and roughness, observed by using international standards. The need for a snow-related extension is, therefore, more emphasized by looking at spectral-domain specific libraries different from snow, where the communities focused on soil and vegetation are more advanced in defining a standardized and detailed metadata scheme. From this perspective, domain-related extensions are important, and the proposed approaches represent a significant guideline for preparing a snow-specific scheme. The Global vis-NIR SL and the LUCAS SL provide, in this case, soil-related extensions composed of attributes (8 and more than 30, respectively) specific for soil chemistry and physics that are not applicable for snow studies. Defining three different components in the metadata scheme (Base, Acquisition and Domain), the considered spectral libraries show significant differences (Table 1).

While the general-purpose and human-oriented libraries (SPECLIB and ECOSTRESS) provides exclusively the Base component, the machine-oriented collections (SPECCHIO, INTA, LUCAS, the Global vis-NIR SL) include both the Base and the Acquisition components and they are already oriented to a machine-to-machine interaction.

The connection between the different spectral libraries represents an interesting perspective since during the last decades smaller databases collapsed into broader services useful for the international communities. While this process occurred in general-purposes databases for collapsing single-institution libraries (Meerdink et al., 2019), domain-specific case studies described how a soil-related metadata scheme is required for harmonizing measurements at a global scale (Viscarra et al., 2016). The example of the collapse process is what happened for the ECOSTRESS SL that represents the union between SLs prepared by the NASA's Jet Propulsion Laboratory (Grove et al., 1992), the USGS (Clark et al., 2003), the Johns Hopkins University previously included in the ASTER database (Baldridge et al., 2009). From this perspective, the potential interaction between the proposed SISPEC metadata profile and the already available spectral libraries is an additional issue that must be analyzed. Two interaction types must be considered: ingestion and harvesting. The first type refers to the ingestion of the SISPEC metadata scheme into the spectral data available in other databases. The first direction considers the possibility to provide SISPEC-described observations to other databases. Although a limited number of snow-related spectra are included in ECOSTRESS, SPECLIB and SPECCHIO, it is critical for SISPEC to interact with those state-of-the-art spectral libraries. Analyzing the different metadata schemes, it is possible to assure a complete overlap between SISPEC and the considered databases. While the conversion is direct for SISPEC-to-SPECCHIO, the SISPEC-to-ECOSTRESS and SISPEC-to-SPECLIB conversions require the aggregation between different attributes included in the SISPEC scheme. The opposite, to-SISPEC, process is unfortunately difficult in terms of conversions not only due to the lack of snow information but also due to the lack of some mandatory SISpec attributes. The required step for the future machine-to-machine ingestion of SISpec data into state-of-the-art spectral information system will be based on preparing specific conversion tools aimed at preparing the data and attributes in the right format. The availability of a larger number of attributes in SISpec will provide the possibility to be aligned with other required profiles as well as the possibility to aggregate attributes in specific metadata requirements.

5. Conclusions

Science needs data, but it is increasingly difficult to share and search them with accuracy and precision now that the increased availability of interconnected sensors and improved storage systems poses the challenges of "Big Data".

Table 1
Metadata structure in different spectral libraries.

Dataset	Snow/total number of spectra	Metadata				Accessibility	Ref
		Base	Acquisition	Spectral	Snow		
ECOSTRESS[a]	3/3400	x				*browse; ascii*	Meerdink et al. (2019)
SPECLIB v7[b]	16/2468	x				*browse; ascii*	Kokaly et al. (2017)
Global vis - NIR spectral library	0/23361	x	x	x		*ascii*	Viscarra Rossel et al., 2016
SPECCHIO[c]	1913/155078	x	x	x		*browse; ascii*	Hueni et al. (2009)
LUCAS spectral library	0/22000	x	x	x		*ascii*	Orgiazzi et al. (2018)
INTA spectral library	0/n.a.	x	x			*private*	Jiménez and Delgado, 2015
SISpec	200/200	x	x		x	*NetCDF*	*this work*

n.a. not available.

[a] https://speclib.jpl.nasa.gov/.

[b] https://crustal.usgs.gov/speclab/QueryAll07a.php.

[c] https://specchio.ch/.

Considering the FAIR rules and metadata standards, the spectral data and ancillary information contained in SISpec have been reconsidered and made compatible with the principles illustrated above. The result is the conversion of ancillary information into metadata. The new setup of the library will provide the polar area monitoring community with an effective tool.

We propose a metadata scheme, which coupled to a NetCDF data model, represents the solution for having a formal and shared standardization aimed at producing well documented and sound metadata for hyperspectral measurements. From this perspective, the availability of metadata for optical, spectrally resolved, field data is an important source of information useful for detecting remotely surface characteristics when field data cannot be collected. Reflectance spectral libraries can help to perform unsupervised hyperspectral image analysis to detect and map surface material. It is well known by now, that there is no "one size fits all" metadata schema and not even a standard for a controlled vocabulary. You must then choose case-by-case finding the most appropriate cataloguing standards to best describe and provide access to your resources. The creation of consistent, standards-based, continuously updated metadata can enable the researchers to publish information about their data and activities in a timely and efficient way and to disseminate this information more widely through specific protocols and data formats (e.g., NetCDF).

Code availability section

This paper does not include the development of specific script/software components. The presented dataset is openly available on a stable public repository (https://zenodo.org/record/4812454#.YOQee-gzahP), where technical encoding specifications and metadata schemas are openly accessible.

Authorship contribution statement

SdF, RSalz, RSalv contributed to the conception, design, analysis, and implementation of the manuscript. RSalv contributed to the acquisition in the field of data measurements; RSalz encoded the NetCDF datasets, and EB encoded the XML metadata scheme. SdF and RSalz shared the joint co-authorship and lead in writing the manuscript. All authors contributed to improving the quality of figures and refining the text in the document during the revision. All authors gave their final approval of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The definition of this metadata profile has been approached to organize datasets obtained during previous and future PNRA field campaigns, in the framework of the CRASI project (PNRA18-00131). The considered snow observations were carried out in collaboration with Mauro Valt (ARPAV). We would also acknowledge Roberto Roncella for reviewing the NetCDF technical encoding specifications.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cageo.2022.105076.

References

Baldridge, A.M., Hook, S.J., Grove, C.I., Rivera, G., 2009. The ASTER spectral library version 2.0. Remote Sens. Environ. 113, 711–715. https://doi.org/10.1016/j.rse.2008.11.007.

Bojinski, S., Schaepman, M., Schlaepfer, D., Itten, K., 2003. SPECCHIO: a spectrum database for remote sensing applications. Comput. Geosci. 29, 27–38. https://doi.org/10.1016/S0098-3004(02)00107-3.

Brown, E.D., 2014. Technology, Strategy, People, Projects. April 2021. http://ericbrown.com/drowning-in-data-starved-for-information.htm.

Clark, R.N., King, T.V.V., Klejwa, M., Swayze, G., Vergo, N., 1990. High spectral resolution reflectance spectroscopy of minerals. J. Geophys. Res. 95, 12653–12680. https://doi.org/10.1029/JB095iB08p12653.

Clark, R.N., Swayze, G.A., Gallagher, A., King, T.V.V., Calvin, W.M., 1993. The U.S. Geological Survey, Digital Spectral Library (Version 1): 0.2 to 3.0 Microns. https://doi.org/10.3133/ofr93592. U.S. Geological Survey, Open-File Report 93–592, 1,340 p.

Clark, R.N., Swayze, G.A., Wise, R.K., Livo, E., Hoefen, T.M., Kokaly, R.F., Sutley, S.J., 2003. USGS Digital Spectral Library Splib05a. U.S. Geological Survey, Open File Report 03-395. https://pubs.usgs.gov/of/2003/ofr-03-395/ofr-03-395.html. (Accessed December 2021).

Casacchia, R., Salvatori, R., Cagnati, A., Valt, M., Ghergo, S., 2002. Field reflectance of snow/ice covers at Terra Nova Bay, Antarctica. Int. J. Rem. Sens. 23 (21), 4653–4667. https://doi.org/10.1080/01431160110113863.

Colbeck, S.C., Akitaya, E., Armstrong, R.L., Gubler, H., Lafeuille, J., Lied, K., McClung, D.M., Morris, E.M., 1990. The international classification for seasonal snow on the ground. In: NTIS (Ed.), International Commission on Snow and Ice of the IAHS, Single volume. NTIS, Springfield, USA, pp. 1–37.

Dang, C., Fu, Q., Warren, S.G., 2016. Effect of snow grain shape on snow albedo. J. Atmos. Sci. 73 (9), 3573–3583. https://doi.org/10.1175/JAS-D-15-0276.1.

DCMI, 2021. Metadata Basics. December 2021. https://www.dublincore.org/resources/metadata-basics/#:~:text=Metadata%2C%20literally%20%22data%20about%20data,data%2C%20concepts%2C%20and%20services.

Domenico, B., 2011. NetCDF Binary Encoding Extension Standard: NetCDF Classic and 64-bit Offset Format. OGC implementation standard. OGC, Arlington, USA, 38pp. OGC-10-092r3.

Domine, F., Salvatori, R., Legagneux, L., Salzano, R., Fily, M., Casacchia, R., 2006. Correlation between the specific surface area and the short wave infrared (SWIR) reflectance of snow. Cold Reg. Sci. Technol. 46, 60–68. https://doi.org/10.1016/j.coldregions.2006.06.002.

Eaton, B., Gregory, J., Drach, B., Taylor, K., Hankin, S., Blower, J., Caron, J., Signell, R., Bentley, P., Rappa, G., Höck, H., Pamment, A., Juckes, M., Raspaud, M., Horne, R., Whiteaker, T., Blodgett, D., Zender, C., Lee, D., Hassell, D., Snow, A.D., Kölling, T., Allured, D., Jelenak, A., Soerensen, A.M., Gaultier, L., Herlédan, S., 2020. NetCDF Climate and Forecast (CF) Metadata Conventions (Ver. 1.10). CF Conventions Committee. Retrieved from. https://cfconventions.org/cf-conventions/cf-conventions.html.

ESIP, 2020. Attribute Convention for Data Discovery. https://wiki.esipfed.org/Attribute_Convention_for_Data_Discovery_1-3.

Fang, L., Chen, S., Zhou, X., Liao, S., Chen, L., 2007. A web-based spectrum library for remote sensing applications of Poyang lake Wetland. Geogr. Inform. Sci. 13 (1–2), 3–9. https://doi.org/10.1080/10824000709480626.

Fierz, C.R., Armstrong, R.L., Durand, Y., Etchevers, P., Greene, E., McClung, D.M., Nishimura, K., Satyawali, P.K., Sokratov, S.A., 2009. The International Classification for Seasonal Snow on the Ground, p. 90. IHP-VII Technical Documents in Hydrology N°83, IACS Contribution N°1, UNESCO-IHP, Paris.

Ghergo, S., Salvatori, R., Casacchia, R., Cagnati, A., Valt, M., 2000. Snow and Ice Spectral Archive (SISpec): un sistema per la gestione di dati spettroradiometrici e nivologici. In: AIT informa - Rivista Italiana di Telerilevamento, 17/18, pp. 3–8.

Gilliland, A.J., 2016. Setting the stage. In: Baca, M. (Ed.), Introduction to Metadata, third ed. Getty Publications, Los Angeles, USA, ISBN 978-1-60606-479-5, pp. 1–20.

Grove, C.I., Hook, S.J., Paylor, E.D., 1992. Laboratory Reflectance Spectra of 160 Minerals, 0.4 to 2.5 Micrometers. NASA, Jet Propulsion Laboratory. Pasadena, CA, USA, Report JPL 92-02. http://hdl.handle.net/2014/40148.

Haegeli, P., Atkins, R., Gerber, M., Hörtnagl, J., Fierz, C., Kelly, J., Morin, S., Nairz, P., Tomm, I., 2010. An international standard for the exchange of snow profile information an example for a domain specific application of CAAML 5.0. In: Proceedings of the International Snow Science Workshop ISSW, Lake Tahoe CA, USA., 17-22 October 2010, pp. 415–416. In: https://arc.lib.montana.edu/snow-science/search.php?workshop=%22International+Snow+Science+Workshop+Proceedings+2010%22.

He, C., Liou, K.N., Takano, Y., Yang, P., Qi, L., Chen, F., 2018. Impact of grain shape and multiple black carbon internal mixing on snow albedo: parameterization and radiative effect analysis. J. Geophys. Res. Atmos. 123, 1253–1268. https://doi.org/10.1002/2017JD027752.

Herold, M., Roberts, D.A., Gardner, M.E., Dennison, P.E., 2004. Spectrometry for urban area remote sensing-Development and analysis of a spectral library from 350 to 2400 nm. Remote Sens. Environ. 91, 304–319. https://doi.org/10.1016/j.rse.2004.02.013.

Hueni, A., Nieke, J., Schopfer, J., Kneubühler, M., Itten, K.I., 2009. The spectral database SPECCHIO for improved long-term usability and data sharing. Comput. Geosci. 35, 557–565. https://doi.org/10.1016/j.cageo.2008.03.015.

Hueni, A., Malthus, T., Kneubuehler, M., Schaepman, M., 2011. Data exchange between distributed spectral databases. Comput. Geosci. 37, 861–873. https://doi.org/10.1016/j.cageo.2010.12.009.

Hueni, A., Chisholm, L.A., Ong, C., Malthaus, T.J., Wyatt, M., Trim, S.A., Schaepman, M. E., Thankappan, M., 2020. The SPECCHIO spectral information system. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 13, 5789–5799. https://doi.org/10.1109/JSTARS.2020.3025117.

INSPIRE, 2007. Technical Guidelines for Implementing Dataset and Service Metadata Based on ISO/TS 19139:2007 (Ver 2.0.1). https://inspire.ec.europa.eu/id/document/tg/metadata-iso19139. (Accessed 15 December 2021).

ISO, 2014. Geographic Information – Metadata – Part 1: Fundamentals (ISO 19115-1: 2014). https://www.iso.org/standard/53798.html. (Accessed 15 December 2021).

ISO, 2016. Geographic Information - Metadata - Part 3: XML Schema Implementation for Fundamental Concepts (ISO 19115-3:2016). https://www.iso.org/standard/32579.html. (Accessed 15 December 2021).

ISO, 2019. Geographic Information — Metadata — Part 2: Extensions for Acquisition and Processing (ISO 19115-2:2019). https://www.iso.org/standard/67039.html. (Accessed 15 December 2021).

Jiménez, M., González, M., Amaro, A., Fernández-Renau, A., 2014. Field spectroscopy metadata system based on ISO and OGC standards. ISPRS Int. J. Geo-Inf. 3, 1003–1022. https://doi.org/10.3390/ijgi3031003.

Jiménez, M., Delgado, R.D., 2015. Towards a standard plant species spectral library protocol for vegetation mapping: a case study in the shrubland of Doñana National Park. ISPRS Int. J. Geo-Inf. 4, 2472–2495. https://doi.org/10.3390/ijgi3031003.

Kokaly, R.F., Clark, R.N., Swayze, G.A., Livo, K.E., Hoefen, T.M., Pearson, N.C., Wise, R. A., Benzel, W.M., Lowers, H.A., Driscoll, R.L., Klein, A.J., 2017. USGS Spectral Library Version 7: U.S. Geological Survey Data Series 1035, 61 pp. https://doi.org/10.3133/ds1035.

Meerdink, S.K., Hook, S.J., Roberts, D.A., Abbott, E.A., 2019. The ECOSTRESS spectral library version 1.0. Remote Sens. Environ. 230, 1–8. https://doi.org/10.1016/j.rse.2019.05.015, 111196.

Neiswender, C., Montgomery, E., 2010. Introduction to metadata interoperability. In: The MMI Guides: Navigating the World of Marine Metadata. https://doi.org/10.23919/OCEANS.2009.5422206.

Nidamanuri, R.R., Ramiya, A.M., 2014. Spectral identification of materials by reflectance spectral library search. Geocarto Int. 29 (6), 609–624. https://doi.org/10.1080/10106049.2013.821175.

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., Fernández-Ugalde, O., 2018. LUCAS Soil, the largest expandable soil dataset for Europe: a review. Eur. J. Soil Sci. 69, 140–153. https://doi.org/10.1111/ejss.12499.

Plini, P., Di Franco, S., De Santis, V., 2009. A state-of-the-art of Italian National Research Council (CNR) activities in the area of terminology and thesauri. In: Hřebíček, J. (Ed.), Proceedings of the European Conference towards eEnvironment. Opportunities of SEIS and SISE: Integrating Environmental Knowledge in Europe, Prague, pp. 614–622.

Rasaiah, B.A., Jones, S.D., Bellman, C., 2012. A novel metadata standard for in situ marine spectroscopy campaigns. In: Arrowsmith, C., Bellman, C., Cartwright, W., Reinke, K., Shortis, M., Soto-Berelov, M., Suarez Barranco, L. (Eds.), Proceedings of the Geospatial Science Research 2 Symposium, Melbourne, ISBN 978-0-9872527-1-5, pp. 1–10.

Rasaiah, B.A., Jones, S.D., Bellman, C., Malthus, T.J., 2014. Critical metadata for spectroscopy field campaigns. Rem. Sens. 6, 3662–3680. https://doi.org/10.3390/rs6053662.

Rasaiah, B.A., Jones, S.D., Bellman, C., Malthus, T.J., Hueni, A., 2015. Assessing field spectroscopy metadata quality. Rem. Sens. 7, 4499–4526. https://doi.org/10.3390/rs70404499.

Rees, W.G., 2006. Remote Sensing of Snow and Ice, first ed. CRC press, Boca Raton, US, ISBN 9780367392307, p. 285.

Rew, R.K., Davis, G.P., Emmerson, S., Davies, H. (1997). NetCDF User's Guide - An Interface for Data Access Version 3, https://wdf.dnw.aero/TOOLS/netCDF/netCDF-cguide.pdf Accessed 28/02/2022.

Riley, J., 2017. Understanding Metadata: what Is Metadata, and what Is it for? NISO Primer Series National Information Standards Organization (NISO), Baltimore, US, ISBN 978-1-937522-72-8, p. 45.

Saito, M., Yang, P., Loeb, N.G., Kato, S., 2019. A novel parameterization of snow albedo based on a two-layer snow model with a mixture of grain habits. J. Atmos. Sci. 76 (5), 1419–1436. https://doi.org/10.1175/JAS-D-18-0308.1.

Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B.G., Bartholomeus, H.M., Bayer, A.D., Bernoux, M., Böttcher, K., Brodský, L., Du, C.W., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C.B., Knadel, M., Morrás, H.J.M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Rufasto Campos, E.M., Sanborn, P., Sellitto, V.M., Sudduth, K.A., Rawlins, B.G., Walter, C., Winowiecki, L.A., Hong, S.Y., Ji, W., 2016. A global spectral library to characterize the world's soil. Earth Sci. Rev. 155, 198–230. https://doi.org/10.1016/j.earscirev.2016.01.012.

Wagner, E.P., Merz, J., Townsend, P.A., 2019. EcoSIS: a spectral library and the tools to use it. In: American Geophysical Union, Fall Meeting 2019 abstract #B11F-2396 Bibcode: 2019AGUFM.B11F2396W.

Warren, S.G., 2019. Optical properties of ice and snow. Philos. Trans. R. Soc. A 377 (2146), 1–17. https://doi.org/10.1098/rsta.2018.0161.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.D., Groth, P., Goble, C., Grethe, J.S., Heringa, J., Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018. https://doi.org/10.1038/sdata.2016.18.