

## Version history

Version	Date	Author(s)	Changes
v0.1 - v0.10	27-Nov-2018 - 12-May-2020	A Hardisty, E Haston	10 early drafts prior to creation of TDWG TG MIDS
v0.11	21-Jul-2020	A Hardisty	Resolve all outstanding comments and suggestions and take into account instructions from review meeting, 17th June 2020.
v0.12	03-Nov-2020	A Hardisty	Conversion to markdown, deposit in TDWG/MIDS repo, cleaning up and separation out in GitHub of this version as the working draft.
v0.13	22-Jan-2021	A Hardisty	Issue #39. Added recommendation in section 3.4 to publish PID alongside MIDS data. ICS proforma updated.
v0.14	29-Mar-2021	A Hardisty	Issues #1, #7, #8, #12, #13 agreements carried into text.
v0.15	29-Jul-2021	A Hardisty	Issues #10, #11, #14, #44, #45 agreements carried into text. Proposed draft text to meet milestone MIDS level 1 proposal.

Title: Minimum Information about a Digital Specimen (MIDS)

Date version issued: 2021-07-29

Date created: 2021-07-29

This version: V0.15

Latest version: V0.15

Previous version: V0.14

Abstract: Minimum Information about a Digital Specimen (MIDS) is a specification defining the information elements expected to be present when publishing digitized information about specimens at various levels of digitization. Digital Specimens are digital representations on the Internet of their physical counterparts in natural science collections. The definition of digitization used here is the process of making physical objects digitally available. Levels of digitization represents a simple categorisation of the type and depth of digitization achieved by heterogeneous approaches to digitization.

Contributors: Alex R Hardisty, Wouter Addink, Mathias Dillen, Quentin Groom, Elspeth Haston, Falko Glöckler, Deborah Paul, Mareike Petersen, Hannu Saarenmaa, Anton Güntsch

Creator: <to be completed>

Bibliographic citation: <to be completed>

Change history: <to be completed. begins with version 1.0>

## Table of contents

- 1 Introduction
  - 1.1 Audience
  - 1.2 RFC 2119 key words
  - 1.3 Content
  - 1.4 Definitions
  - 1.5 Acknowledgements
- 2 Background (informative)
  - 2.1 Processes of digitization (informative)
  - 2.2 Stages of digitization (informative)
  - 2.3 Harmonising the approach (informative)
  - 2.4 Accommodating variability in details of the process (informative)
- 3 Principles of MIDS levels of digitization (normative)
  - 3.1 Authoritative data and variations in digitization
  - 3.2 Other data present or known about a specimen
  - 3.3 Prior to digitization
  - 3.4 Assigning persistent identifiers to digitized specimen data
  - 3.5 MIDS levels of minimum information
  - 3.6 Information elements expected and expansion beyond the minimum
- 4 Information content and element mappings for each MIDS level (normative)
  - 4.1 Information elements expected at MIDS level 1
  - 4.2 Information elements expected at MIDS level 2
  - 4.3 Information elements expected at MIDS level 3
  - 4.4 Information elements expected at MIDS level 0
  - 4.5 Information element definitions
    - 4.6.1 MIDS element - Modified
    - 4.6.2 MIDS element - MIDSLevel
    - 4.6.3 MIDS element - PhysicalSpecimenId
    - 4.6.4 MIDS element - Institution
    - 4.6.5 MIDS element - ObjectType
    - 4.6.6 MIDS element - MaterialType

- 4.6.7 MIDS element - PreparationType
- 4.6.8 MIDS element - Name
- 4.5 Handling of unknown and incomplete data
  - 4.5.1 Unknown datetime in CreatedON/ModifiedON information elements
  - 4.5.2 Geographical data information elements
  - 4.5.3 Unknown values for other information elements
- 5 Images and other media types (normative)
- 6 Guidelines for machine-actionability (normative)
- 7 Recommendations for collection-holding institutions
- 8 Guarding against publication of junk data
- 9 Conformance (normative)
  - 9.1 Principle means of conformance
  - 9.2 Implementation Conformance Statement (ICS)
- 10 Illustrative use cases and guidelines for use (informative)
- 11 References

## 1 Introduction

---

Minimum Information about a Digital Specimen (MIDS) specifies the information elements expected to be present when providing access to specimens within a digital framework. Digital Specimens are online digital representations of their physical counterparts in natural science collections. The definition of digitization used here is the process of making physical objects digitally available in terms of data and/or images. Several 'levels of digitization' (0-3) represent a simple categorisation of the type and depth of digitization achieved by heterogeneous approaches to digitization.

MIDS has multiple aims:

1. To improve the quality of published data by offering clarity to stewards of collections about the minimum quantity and quality of information they should be publishing to make digital specimen information useful for multiple purposes of teaching and learning, research, etc.;
2. To assist the global effort to digitize natural science collections, estimated to be 3 billion specimens worldwide by providing a structured framework that clarifies the outcomes of digitization and the level of digitization achieved, and assists prioritization of the remaining work;
3. To support and contribute towards assessments of fitness for purpose of data (suitability) for feeding specific types of data processing pipelines; and,
4. To assist researchers to know what information to include in their journal articles about specimens they have used in their research .

MIDS defines what information about a specimen should be available but it does not attach specific value to that information because attaching value depends both upon the information itself and the purpose to which that is put. MIDS does not codify these aspects.

The levels of digitization and their expected information content defined herein, and named as Minimum Information about a Digital Specimen (MIDS) are based on major research and curation requirements put forward by national museums and herbaria.

### 1.1 Audience

---

This document is intended primarily for those who are responsible for digitizing and sharing data publicly (publishing data) about natural science specimens. It can also be useful to those who are developing applications, tools and workflows related to digitization, and for those reporting to collection management personnel and funding agencies.

### 1.2 RFC 2119 key words

---

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119. [RFC-2119]

### 1.3 Content

---

This is the only document concerned with the specification of Minimum Information about a Digital Specimen (MIDS). Sections are designated as either 'informative' or 'normative'. This document first explains the rationale, principles of MIDS and the levels of digitization, then details the information elements to be expected at each MIDS level. A normative definition of each MIDS information element is provided, including constraints and mappings to Darwin Core (DwC) terms and Access to Biological Collection Data (ABCD) terms. An implementation conformance statement (ICS) proforma is provided to allow declarations of conformance to be made. Illustrative use cases and guidelines for use are also provided.

### 1.4 Definitions

---

**bare record** – the association between an identifier of a physical specimen and its initial digital representation, allowing for subsequent unambiguous attachment of all other information (cf. basic record, regular record and extended record).

**basic record** – a limited set of information about the specimen; for example, physical specimen identifier and barcode details, and details of the collection it belongs to (cf. bare record, regular record and extended record).

**data (also known as content data)** – data relating directly to describing physical specimens, such as images of those specimens, information from specimen labels (scientific name, location where collected, date collected, collector name, etc.), links to third-party semantic resources, or measurements and other analyses of specimens.

**digital specimen** – a digital representation on the Internet (digital twin) corresponding to a physical specimen in a natural science collection.

**digitization** – the process of converting analog information about physical specimens (e.g. in natural science collections) to digital form, which therefore includes electronic text, images and other representations; and making that publicly available.

**extended record** – a comprehensive set of elements describing the specimen; for example: all label data, annotations and determination history (cf. bare record, basic record and regular record).

**levels of digitization** – a simple categorisation of the type and depth of digitization achieved by heterogeneous approaches to digitization.

**natural science collection** – a collection of natural science objects, such as preserved plants and animals, fossils, rocks and minerals, gems, etc. curated and maintained by an institution such as a museum or a university.

**persistent identifier (PID)** – a persistent identifier is a string (functioning as a symbol) that uniquely identifies a thing. The identifier can be persistently resolved to digitally actionable meaningful information about the identified thing.

**referent** – in the context of two-part information elements that name things, the referent is an identifier of an actual person, organisation, place or other thing signified by a name string, such as a collector name, museum name, place name, etc.

Note 1: Examples of referents include: i) a person identification to disambiguate between persons sharing the same name, where an ORCID iD (<https://orcid.org/>) can be used to disambiguate living persons while a Wikidata item (<https://www.wikidata.org/wiki/Help:Items>) can be used to disambiguate persons unable to obtain an ORCID iD (typically, deceased collectors' names); ii) a research organisation or collection-holding institution can be referred to by its GRID (<https://www.grid.ac/>) or ROR (<https://ror.org/>) identifier; iii) place names can be uniquely identified using GeoNames entities (<https://www.geonames.org/>); and iv) in the case of a scientific name a scientific name identifier can be used to unambiguously establish a reference to the correct nomenclatural details.

**regular record** – a set of information describing the specimen; for example, physical specimen identifier, barcode details, taxon name, collection date and place, collector, etc. (cf. bare record, basic record and extended record).

## 1.5 Acknowledgements

This minimum information specification is the result of work carried out jointly with contributions from multiple funded projects that include:

- CETAF Digitization Working Group;
- European Union Horizon 2020-funded ICEDIG project (grant agreement no. 777483);
- European Union Horizon 2020-funded SYNTHESYS+ project (grant agreement no. 823827);
- European Union Horizon 2020-funded DiSSCo Prepare project (grant agreement no. 871043);
- European Union COST Action CA17106 MOBILISE; and,
- US National Science Foundation (NSF) Advancing Digitization of Biodiversity Collections Program DBI-1547229 (2016-2021), iDigBio.

## 2 Background (informative)

### 2.1 Processes of digitization (informative)

More than 3 billion physical objects cared for, organised and catalogued in thousands of natural science collections around the world represent an important data resource in botany, zoology, geology and other disciplines for research addressing scientific and societal questions. In common with most other contemporary scientific endeavours, once scientists working with such objects find themselves able to access data resources electronically without leaving their desks, they quickly find that they want to access further resources that are sometimes not so advanced in terms of 'digital transformation' [Meyer 2015]. An important component of digital transformation in natural sciences is the process of making data about collections and the physical objects in those collections digitally available, findable, and accessible by acquiring data from the physical objects themselves (i.e., by digitization), through curation and further processing of that data, to open data publication and use.

Digitization can lead to digital data for collections as a whole (i.e., an overview), for sub-parts of collections (inventories of trays of insects, boxes of herbarium sheets, for example) and for individual specimens. The first two categories contribute towards providing coverage and access information about the holdings of a collection-holding institution, in terms of scope and extent and are the subject of a companion Collection Descriptions standard. Digitization of specimens, which is the focus of the present specification provides explicit and precise data about each object curated in a collection. Capturing and presenting such data in standard formats is essential so that it can be easily understood, compared, analysed and communicated. Similarly, ensuring that enough data are captured, curated and published is essential so that it is useful for the widest possible range of purposes. Finally, ensuring that such data are consistently indexed and identified, such that they can be found, accessed and used is critical. This is the rationale for minimum information specifications such as the present one.

### 2.2 Stages of digitization (informative)

At the level of specimens, digitization can be characterised generally as consisting of several activities, with data created as a result of each activity:

- Attaching an identifier to a physical specimen and creating a digital catalogue record, perhaps with its own digital identifier. This can include attaching collection-level data (from cabinets, boxes, folders, trays, etc.) to a group of specimens;
- Specimen data capture, which can include:
  - Making one or more images of the specimen and/or its labels; and,
  - Extracting, processing and encoding information from those labels and images into a digital record;
- Enriching the digital record with supplementary information that can include determinations, annotations, and/or information from a wide variety of third-party sources (references to published literature, genetic sequences, etc.).

These stages are often implemented as workflows, which can vary in their details from one institution or project to another. Nevertheless, the heterogeneity of approaches and the variations are understood [Nelson 2015]. These and other differences of approach are accommodated by MIDS.

### 2.3 Harmonising the approach (informative)

Multiple digitization initiatives around the world capture digital data about specimens. An important concern in digitization is how much detail to digitize from each physical specimen. While a photographic image can often be made quickly, transcribing and interpreting all the details from labels, enriching the data with external information, and making specific measurements of the specimen take more time and resources. Often, digitization projects have been conceived to support specific scientific needs leading to variable outcomes in terms of what data is captured and how it is presented. The idea of 'Minimum Information about a Digital Specimen' (MIDS) has been conceived to structure this complexity.

Mobilising, unifying and delivering natural science (bio- and geo-diversity) information at the scale, form and precision required by the scientific communities can be accomplished in part by harmonising policy into guidelines about practical levels of digitization to apply and also in part by harmonising the information to be expected from each level of digitization. The present specification addresses the latter.

Harmonisation of information provides clarity about the minimum information that collection-holding institutions should be publishing in the future to make collections and digital specimens useful for multiple purposes of research, teaching and learning, etc. Similarly, by harmonising a framework that clarifies what is meant by different levels of digitization it becomes easier to consistently measure the extent of digitization achieved over time (e.g., via a collection digitization dashboard) and to set priorities for the remaining work.

The notions of extended specimens [Webster 2017] and Next Generation Collections [Schindel 2018], together with work of initiatives such as iDigBio, DiSSCo, NRCA Digital, NSII, the Biodiversity Collections Network (BCoN) and the One World Collection project further drive the need for data harmonisation. Collections-based science is global in nature and increasingly digital. In this context, barriers to interoperability and re-use must be lowered and eventually removed and global solutions are highly desirable.

The Minimum Information about a Digital Specimen (MIDS) specification, and its companion specification for information about collections – the Collection Description (CD) standard – aims to address these problems. Adoption and use of MIDS (and CD) contributes to making natural sciences specimen data more compliant with the guiding principles of FAIR ('Findable, Accessible, Interoperable, Reusable') [Wilkinson 2016, Mons 2017].

## 2.4 Accommodating variability in details of the process (informative)

---

Heterogeneous approaches to digitization are accommodated in MIDS via a simple categorisation based on the form and depth of digitization achieved.

MIDS is divided into several levels that represent a progression through the stages of digitization. Not all digitization programmes can publish all the information right from the beginning, and it is useful when collections owners publish lesser quantities at the earliest stages. Indeed, it is likely that many digitization programmes cannot or do not publish all the information at all. Only through subsequent processes of research, annotation and interpretation, and extension can specimen information become increasingly enriched and more (but never) complete. Such processes ebb and flow with the tide of scientific enquiry, contributing to and building up an ever-richer corpus of information associated with digitised artefacts. Since we can never know all there is to know about a specimen, the processes never finish and thus there can be no notion of maximum or complete information. The highest 'gold standard' MIDS level 3 is open-ended to account for this.

In support of the generalised digitization workflow, MIDS defines the information elements expected at each level of digitization and strongly encourages that data resulting from digitization should be made openly available online on the Internet. Thus, in the context of MIDS, "digitized" means not only existing in digital form but also existing as "online representations that are publicly available i.e., openly findable and accessible". Note here that 'accessible' means to the extent that access is not restricted based on objective criteria defined by national security, legislation or other regulatory compliance, sensitivity of collection information, and third-party rights (such as personal privacy).

MIDS is a 'minimum specification'. This means that at any specific MIDS level a defined set of expected information elements should be the minimum amount of information made digitally and openly available online following each of the major stages of digitization identified in the generalised workflow (2.2 above). Because MIDS is a minimum specification, publishing more extensive sets of information elements is not precluded and is, in fact, strongly recommended.

Specific practices of digitization and data publication vary widely among institutions, meaning that different elements of information can become available as the result of specific detailed workflows steps, the order of which can vary from one institution to another. Similarly, the way in which different elements of information are mapped to standard terms also varies. MIDS accommodates this variability and always allows the presence of information elements additional to the expected elements at any MIDS level. These can include elements listed at higher MIDS levels, as well as other elements beyond those listed; for example, institution or collection specific elements. As a general principle, institutions should publish the fullest available data about their collections and individual specimens at the earliest opportunity, expecting that such data is likely to become enriched and annotated over time.

## 3 Principles of MIDS levels of digitization (normative)

---

### 3.1 Authoritative data and variations in digitization

---

MIDS must focus on the primary, curated scientific information denoted as being authoritative in relation to the specimen. This generally includes the 'what, when, where, and who' of the specimen i.e., What is it? When was it collected? Where was it collected? And who by? Apart from management information such as accession numbers, barcodes and catalogue numbers, 'what, when, where, who' is the minimum information usually curated in natural science collections. It is also normally a principal product of digitization processes, alongside capturing photographic images of specimens. Thus, establishing this quartet of 'what, when, where, and who' as the standard to be achieved as the regular level of information to be expected, is completely rational. Nevertheless, digitization processes vary from one specimen category to another and between institutions. Different levels of information – some less comprehensive, some more so – become available at different points in digitization processes. Thus, MIDS offers several levels of minimum information to be expected or achieved as the output of various stages of what is quite a variable process in the way it is conducted from one institution or collection to another.

### 3.2 Other data present or known about a specimen

---

In contrast to the authoritative data about a specimen, a wide variety of secondary or supplementary data may also be associated with a specimen. This can include references in literature, the results of various kinds of analysis such as (bio)chemical analysis, DNA sequencing or radiocarbon dating, links to audio and video recordings, habitat information, etc. Such data can be held with the specimen or its catalogue record, or it can be held in other institutes and databases beyond the institution holding the physical specimen. It is data that can become associated with a specimen.

Supplementary data can either be directly related to the physical specimen, as in the case of a database record of a DNA sequence taken from a tissue sample cut from the specimen; or such data can be related only indirectly. An audio recording of the birdsong of a species that the specimen represents is an example of the latter i.e., the recording is not of that specific individual but of another in the same taxon. Indirect relations are characterised as conceptual relations rather than direct or actual relations.

MIDS optionally allows for the minimum information about a specimen to be extended by making provision to capture and publish related data.

### 3.3 Prior to digitization

In different cases, and to support processes in collection-holding institutions including accession processes and 'digital by default' processes, the need to create a digital record of a specimen can arise at a very early stage in the life cycle of an acquired specimen, even before digitization processes have been initiated.

In some instances, a photographic image can also be made or be already available at such a moment. Or data related to other actions, such as DNA or chemical analysis and publication can become available. Actions like taking a photograph or carrying out an analysis can take place before accession to a collection and digitization. They can occur even without a physical specimen ending up as preserved material in a collection. Such actions represent the very beginning of a digitization process and can be precursive to further digitization steps later.

When needed at such moments, a bare catalogue record may be created to associate a specimen with a catalogue entry or database record. This should normally be via the specimen's identifier, associating a bare minimum of digital data (a catalogue number and, ideally a persistent identifier – see 3.4 below) with a physical specimen.

Note 2: MIDS supports early stage digital bare catalogue record creation by providing a pre-level of minimum information about a digital specimen, level zero (0) that is precursive to later steps.

### 3.4 Assigning persistent identifiers to digitized specimen data

As part of the minimum information required to make digital specimen data publicly available, a persistent identifier (PID) should be assigned to a specimen and its digital data as early as possible in the digitization process; even prior to digitization commencing (see 3.3 above). If a PID is available for the digital specimen, this should be published alongside the MIDS information.

Note 3: Persistent and global unique identification is an important step towards making digitized specimen data publicly available online so that they can be referred to unambiguously, for example in a journal article, or so that its data can be easily found, used and updated whenever necessary. With the advent of innovative services making greater use of PIDs in the future, persistently identified specimen data is expected to become more valuable over time.

Note 4: By combining a generated suffix with a prefix from the Handle system i.e., <prefix>/<suffix> an unambiguous, persistent (long-lasting) resolvable reference to the digital specimen's information can be created. A Digital Object Identifier (DOI™) is an example of such a persistent identifier.

Note 5: Beyond the recommendation above, the topic of persistent identification of specimen data is outside the scope of the present document.

### 3.5 MIDS levels of minimum information

As outlined in Table 1, MIDS specifies three levels of 'minimum information', together with a pre-level, level zero (0). Excepting level 0, each MIDS level of minimum information is a superset of the preceding level.

Table 1: MIDS levels of minimum information

MIDS level	Record extent	Purpose
1	Basic	A basic record of specimen information.
2	Regular	Key information fields that have been agreed over time as essential for most scientific purposes.
3	Extended	Other data present or information known about the specimen, including links to third-party sources.
0 (Note)	Bare	<i>A bare or skeletal record making the association between an identifier of a physical specimen and its digital representation, allowing for unambiguous attachment of all other information.</i>

**Note:** Level 0 is equivalent to creating a simple catalogue record containing a physical specimen identifier, such as a barcode number. Level 0 often precedes fuller digitization steps that yield more detailed information. Hence, level 0 is termed a pre-level. Nevertheless, level 0 data is useful minimum information for advertising or knowing about the existence of specimens.

**MIDS level 1** is a basic level of information about a specimen. The creation of a basic digital record with a globally unique identifier enables all other information (including images and/or other media types) to be associated unambiguously with a single physical specimen. The elements of name and geographical region are often included to create a virtual representation of the physical collections, enabling similar search and browsing capabilities online as researchers and curators would have at the cabinets of physical specimens.

**MIDS level 2** is the regular level of information expected to be known about a specimen. Regular records include the latitude/longitude data which, along with taxon name, are critical for many research applications. For botanical specimens the inclusion of collector name, number and date are a priority, given their historical use in acting as the identifier for the specimen in literature. These terms are routinely used to identify the specimen in requests to curation staff.

**MIDS level 3** is an extended or rich level of information about a specimen. Extended or rich records include other data present or known about the specimen. Such records will need to be updated as additional labels and other information are added to the specimen over time. At level 3 there is no notion of completeness or of a full or complete record because new information is always valuable and can be added to an existing record at any time.

**MIDS level 0** (usually being established prior to any formal digitization and hence considered a pre-level) is a bare level of information that exists to make an association between a physical specimen with its identifier (barcode, for example) and an entry in a catalogue or database. One early action in a digitization process is the creation of a database record that acknowledges or coincides with the existence or accession of a physical specimen into a collection, but not necessarily with any other information digitized at that stage. Note, however that image(s) can be generated and referenced at this early stage.

### 3.6 Information elements expected and expansion beyond the minimum

At each level, MIDS defines the information elements that must be present. This is specified in section 4 below

In addition to the information elements expected to be present as a minimum, MIDS must always allow the presence of other information elements beyond the minimum, including institutional and collection specific elements.

Note 6: It is not precluded and is encouraged that institutions publish information beyond the minimum. However, information does not qualify as meeting a specific MIDS level unless all the information elements expected at that level are available with actual values.

## 4 Information content and element mappings for each MIDS level (normative)

Sections 4.1 - 4.4 summarise in tables the information elements expected to be present at MIDS levels 1, 2, 3 and 0 respectively. Full definitions of each information element, their mappings to other standards, constraints and examples are specified in tables in section 4.5.

Section 4.6 specifies additional requirements for handling of unknown or incomplete data that must be applied in conjunction with the relevant information elements.

### 4.1 Information elements expected at MIDS level 1

The information elements expected to be present in digital specimen data published at MIDS level 1 are listed in Table 2.

**Table 2: MIDS Level 1: Expected elements, mappings and recommendations for a basic record**

	MIDS information element	Definition	Requirement
1	Modified	UTC date/time of first creation or subsequent modification, if any	Use [ISO 8601]. UTC date/time, extended format (for human readability).
2	MIDSLLevel	A categorisation of the type and depth of digitization achieved: Basic	Integer, value = 1.
3	PhysicalSpecimenId	A unique identity for the specimen within the curating institution. Whatever the institution uses to uniquely identify the item, for example: specimen identifier, barcode, catalogue number, etc.	Cardinality 1..n.
4	Institution [Code, Referent]	Identifier(s) of the institution. Has two parts: i) Code: The name (or acronym) in use by the institution having custody of the object(s) or information referred to; and ii) Referent: An identifier of the organisation signified by Code.	Code from GBIF Registry. Referent from ROR or Wikidata; else 'unknown'.
5	ObjectType	High-level term for the classification of curated objects. The type or general nature of a material object/sample.	objectType should be set in conjunction with preparationType, as each objectType has specific preparationTypes.
6	MaterialType	The material the object is composed of.	
7	PreparationType	General terms for the classification of curated objects according to the way they have been prepared. In combination with ObjectType - hierarchical; a more specific classification than described by ObjectType.	Preparation Type should be set in conjunction with ObjectType, as each ObjectType has specific Preparation Types.
8	Name	A name given to the object. Equivalent to the Dublin Core term dc:title. Any string of characters and/or numbers by which the object is referenced within a collection. For example, the name the specimen is stored under, it's scientific or taxonomic name if known, how it is labelled, etc. This name is not necessarily its name according to an accepted scientific classification, identification, or taxonomic determination (i.e., scientific name) but it often can be the same as that.	Examples: <i>Colias croceus</i> (Geoffroy, 1785), Ophthalmosaurus, Australian Baobab, Scottish rock, etc.

An image and/or other multimedia may be present (see 5 below).

### 4.2 Information elements expected at MIDS level 2

In addition to the information elements expected at MIDS level 1, the information elements expected to be present in digital specimen data published at MIDS level 2 are listed in Table 3.

29Jul2021: MIDS level 2 is still open for discussion. Progress of the discussion can be following in GitHub via the issues collected under the MIDS level 2 proposal milestone.

**Table 3: MIDS Level 2: Expected elements, mappings and recommendations for a regular record**

	MIDS information element	Definition	Requirement
1	Modified	UTC date/time of first creation or subsequent modification, if any	Use [ISO 8601]. UTC date/time, extended format (for human readability).
2	MIDSLLevel	A categorisation of the type and depth of digitization achieved: Regular	Integer, value = 2.
3 - 8	<i>Elements 3 - 8 as for MIDS level 1</i>	<i>per Table 2</i>	
9	Continent		Note
10	Country		Use [ISO 3166]
11	State/province		Use [ISO 3166]
12	County [Name, Referent]		Use [ISO 3166]. GeoNames identifier if known; else 'unknown'
13	Locality [Name, Referent]	locality	Locality name. GeoNames identifier if known; else 'unknown'.
14	Latitude/longitude	decimal latitude, decimal longitude	If known; else "unknown". Include georeference datum, source and uncertainty if known; else "unknown".
15	Altitude/depth	verbatim elevation or verbatim depth	If known; else "unknown".
16	Collector name [Name, Referent]	recordedBy	Person's name. Wikidata identifier if known; else 'unknown'.
17	Collector number		"s.n." or equivalent if not present
18	Collection date		"s.d." or equivalent if not present
19	Collection code/name		From GBIF Registry
20	Type status		Controlled vocabulary
21	Geographical region		Note
22	Deposited/accession date		Use [ISO 8601]
23	Name identifier		Uniquely identifying name in 8. ScientificName
<p><b>Note:</b> Geographical region should be identified in one of three major categories: terrestrial (WGSRPD level 1 areas, [Brummitt 2001]), marine (IHO World Seas, [FMI 2018]) or extra-terrestrial. The geographical region value must be specified in two parts i.e., Terrestrial: &lt;regionname below&gt; or Marine:&lt;seaname below&gt; or Extra-terrestrial:Extra-terrestrial. World/NA should be used when a more specific regional classification is not known.</p>			
<p><b>Terrestrial:</b> Europe, Africa, Asia Temperate, Asia Tropical, Australasia, Pacific, North America, South America, Antarctic, World/NA</p>		<p><b>Marine:</b> North Pacific, South Pacific, North Atlantic, South Atlantic, Indian, Southern, Arctic Marine, World/NA</p>	

An image and/or other multimedia may be present (see 5 below).

### 4.3 Information elements expected at MIDS level 3

In addition to the information elements expected at MIDS levels 1 and 2, the information elements expected to be present in digital specimen data published at MIDS level 3 are listed in Table 4.

*29Jul2021: MIDS level 3 is still open for discussion and for further study. At MIDS level 3, Extended, we expect interpreted, added data that has not been directly extracted from the physical object but rather from already available digital data. Examples are links to publications, sequences, georeferenced data etc. This level can have many/any enrichments and is open-ended.*

**Table 4: MIDS Level 3: Expected elements, mappings and recommendations for an extended record**

	MIDS information element	Definition	Requirement
1	Modified	UTC date/time of first creation or subsequent modification, if any	Use [ISO 8601]. UTC date/time, extended format (for human readability).
2	MIDSLLevel	A categorisation of the type and depth of digitization achieved: Extended	Integer, value = 3.
3 - 8	<i>Elements 3 - 8 as for MIDS level 1</i>	<i>per Table 2</i>	

9 - 23	<i>Elements 9 - 23 as for MIDS level 2</i>	<i>per Table 3</i>	
24	Determination(s)	All determinations (with taxon name identifier) including the determiner (with the person name identifier)	
25	Remaining transcription from original collection label	All information from the original collection label not already transcribed	
26	Quality assertions	<i>Element for further study.</i>	
27	PID links to image(s)	<i>Element for further study.</i>	
28	PID links to other data present or information known about the specimen	<i>Element for further study.</i>	
29	PID links to annotations	<i>Element for further study.</i>	
30	PID links to interpretations	<i>Element for further study.</i>	
31	PID links to provenance	<i>Element for further study.</i>	
32	Institutional persistent identifier (CETAF Stable Identifier/URI)	<i>Element for further study.</i>	

An image and/or other multimedia may be present (see 5 below).

#### 4.4 Information elements expected at MIDS level 0

The information elements expected to be present in digital specimen data published at MIDS level 0 are listed in Table 5.

**Table 5: MIDS Level 0: Expected elements, mappings and recommendations for a bare record**

	MIDS information element	Definition	Requirement
1	Modified	UTC date/time of first creation or subsequent modification, if any	Use [ISO 8601]. UTC date/time, extended format (for human readability).
2	MIDSLevel	A categorisation of the type and depth of digitization achieved: Bare	Integer, value = 0.
3	PhysicalSpecimenId	A unique identity for the specimen within the curating institution. Whatever the institution uses to uniquely identify the item, for example: specimen identifier, barcode, catalogue number, etc.	Cardinality 1..n.
4	Institution [Code, Referent]	Identifier(s) of the institution. Has two parts: i) Code: The name (or acronym) in use by the institution having custody of the object(s) or information referred to; and ii) Referent: An identifier of the organisation signified by Code.	Code from GBIF Registry. Referent from ROR or Wikidata; else 'unknown'.

An image and/or other multimedia may be present (see 5 below).

#### 4.5 Information element definitions

The following subsections specify the details of the individual MIDS information elements.

*29Jul2021: MIDS elements beyond MIDS level 0 and 1 are not yet discussed and agreed. Subsections will be added for each as they are agreed.*

##### 4.6.1 MIDS element - Modified

MIDS information element	Modified
Definition	UTC date/time of date/time of first creation or subsequent modification, if any
DwC term (latest, 2014-11-08)	modified ( <i>note, is an imported dcterms</i> )
ABCD term name(3.0)	hasDateModified
Applicable standard(s)/recommendation(s)	Use ISO 8601. UTC date/time, extended format (for human readability)
Element identifier	
Required	Yes
Repeatable	No
Constraints	
Examples	2021-01-20T10:54:18Z
Element specification status	<i>agreed; accepted in specification</i>
Notes	...

##### 4.6.2 MIDS element - MIDSLevel

##### 4.6.3 MIDS element - PhysicalSpecimenId



MIDS information element	PhysicalSpecimenId
Definition	A unique identity for the specimen within the curating institution. Whatever the institution uses to uniquely identify the item, for example: specimen identifier, barcode, catalogue number, etc.
DwC term (latest, 2014-11-08)	catalogNumber ( <i>note, term used by GBIF is catalogNumber</i> )
ABCD term name(3.0)	physicalObjectID
Applicable standard(s)/recommendation(s)	Cardinality 1..n
Element identifier	
Required	Yes
Repeatable	Yes. An object can have several identifiers.
Constraints	The physical specimen identifier, when given back to the organization indicated by the Institution information element must be sufficient to allow the organization to locate the object(s) the user is interested in. In cases where a single identifier identifies multiple objects (as is possible in some collections) information such as objectType, materialType and preparationType may also need to be given.
Examples	<i>To be added</i>
Element specification status	<i>agreed</i>
Notes	...

#### 4.6.4 MIDS element - Institution

MIDS information element	Institution [Code, Referent]
Definition	Identifier(s) of the institution. Has two parts: i) Code: The name (or acronym) in use by the institution having custody of the object(s) or information referred to; and ii) Referent: An identifier of the organisation signified by Code.
DwC term (latest, 2014-11-08)	i) institutionCode, ii) No equivalent
ABCD term name(3.0)	i) sourceInstitutionID, ii) No equivalent
Applicable standard(s)/recommendation(s)	Acronyms (namestrings) for Code will typically come from the GBIF Registry of Scientific Collections (nee GrSciColl); else 'unknown'. Unique identifiers for Referent should come from either ROR or Wikidata; else 'unknown'.
Element identifier	
Required	Yes
Repeatable	Yes. Multiple organisation identifiers are likely for the same organisation. Multiple name acronyms are also possible. See constraints
Constraints	When there are multiple identifiers for the organisation represented by Code, the Referent part can be repeated as necessary i.e., <i>Institution [Code, Referent1, Referent2, ... ReferentN]</i> . When an organisation is known by multiple Codes (e.g., due to historical reasons) the information element must be repeated for each Code with all of the Referents i.e., <i>Institution [Code1, Referent1, Referent2, ... ReferentN], Institution [Code2, Referent1, Referent2, ... ReferentN]</i> .
Examples	<i>To be added</i>
Element specification status	<i>agreed; accepted in specification</i>
Notes	...

#### 4.6.5 MIDS element - ObjectType

MIDS information element	ObjectType
Definition	High-level term for the classification of curated objects. The type or general nature of a material object/sample.
DwC term (latest, 2014-11-08)	proposed: materialSampleType
ABCD term name (3.0)	KindOfUnit
Applicable standard(s)/recommendation(s)	
Element identifier	
Required	No
Repeatable	No
Constraints	Controlled vocabulary. objectType should be set in conjunction with preparationType, as each objectType has specific preparationTypes.
Examples	<i>To be added</i>
Element specification status	<i>agreed; accepted in specification. Created following TG mtg no. 7.</i>

<b>Notes</b>	<i>Broadly, same interpretation as cd:objectType. Must not be used for classifying objects by taxon. Definition of the controlled vocabulary is needed</i>
--------------	------------------------------------------------------------------------------------------------------------------------------------------------------------

#### 4.6.6 MIDS element - MaterialType

MIDS information element	MaterialType
Definition	The material the object is composed of.
DwC term (latest, 2014-11-08)	
ABCD term name (3.0)	
Applicable standard(s)/recommendation(s)	
Element identifier	
Required	Yes
Repeatable	No
Constraints	Controlled vocabulary
Examples	<i>To be added</i>
Element specification status	<i>agreed; accepted in specification</i>
Notes	<i>Definition of controlled vocabulary is needed.</i>

#### 4.6.7 MIDS element - PreparationType

MIDS information element	PreparationType
Definition	General terms for the classification of curated objects according to the way they have been prepared. In combination with ObjectType - hierarchical; a more specific classification than described by ObjectType.
DwC term (latest, 2014-11-08)	preparations
ABCD term name (3.0)	Unit/SpecimenUnit/Preparations/Preparation/PreparationType
Applicable standard(s)/recommendation(s)	
Element identifier	
Required	Yes
Repeatable	No
Constraints	Controlled vocabulary. Preparation Type should be set in conjunction with ObjectType, as each ObjectType has specific Preparation Types.
Examples	<i>To be added</i>
Element specification status	<i>agreed. Created following TG mtg. no. 7.</i>
Notes	<i>Same interpretation as cd:preparationType. Must not be used for classifying objects by taxon. Definition of the controlled vocabulary is still needed.</i>

#### 4.6.8 MIDS element - Name

MIDS information element	Name
Definition	A name given to the object. Equivalent to the Dublin Core term dc:title. Any string of characters and/or numbers by which the object is referenced within a collection. For example, the name the specimen is stored under, it's scientific or taxonomic name if known, how it is labelled, etc. This name is not necessarily its name according to an accepted scientific classification, identification, or taxonomic determination (i.e., scientific name) but it often can be the same as that.
DwC equivalent term (latest, 2014-11-08)	scientificName, organismName, vernacularName, etc.
ABCD equivalent term (3.0)	name, fullScientificName, informalName, etc.
Applicable standard(s)/recommendation(s)	
Element identifier	name
Required	Yes
Repeatable	No
Constraints	None
Examples	<i>Colias croceus (Geoffroy, 1785), Ophthalmosaurus, Australian Baobab, Scottish rock, etc.</i>
Element specification status	<i>agreed; accepted in specification</i>
Notes	...

## 4.5 Handling of unknown and incomplete data

Best practice dictates that wherever possible data should not be published with empty field values as this is misleading for both human users and machines. There are many reasons why data can be missing, unknown, incomplete or explicitly withheld [Groom 2019] and various tactics have been used in the past to deal with such situations. However, with the increasing use of machines to interpret and act upon data, variable past practices should be deprecated, and new, more consistent practices promoted. In the interest of presenting more meaningful corpora of minimum specimen information in the future, the present specification (4.5.1 - 4.5.3 below) recommends harmonising an approach for presenting minimum information fields where the content of those fields can be unknown or incomplete for a variety of reasons.

### 4.5.1 Unknown datetime in CreatedON/ModifiedON information elements

If the creation and/or modification date/time of a specimen data record is not known, the standard form for unspecified date/time, as specified in ISO 8601-2:2019 should be used for the createdON and/or modifiedON information element(s) i.e., "X\*" or "XXXX-XX-XX".

### 4.5.2 Geographical data information elements

To achieve MIDS level 2, each geographical information element listed in Table 3 (i.e., Continent, Country, State/province, County, Locality, Latitude/longitude, Altitude/depth, Geographical Region) must be provided.

If such information is not present on the specimen label, but can be inferred/interpreted from other information present or from other sources, it should be provided. If such information is not present and cannot be inferred it should be recorded as "unknown:missing" (4.5.3 below).

Country level information has particular importance and this (or at least one other more localised named area) should always be provided.

### 4.5.3 Unknown values for other information elements

Table 6 lists the terms that should be used in lieu of any of the information elements listed in Table 2 – Table 4 above being left empty.

**Table 6: Terms for missing data values in lieu of empty information elements**

Based on [Groom 2019], Table 1. CC BY.

Missing data term	Definition	Example
unknown (Note 1)	The information is not digitally available.	Empty value in a digital record of unknown provenance.
unknown:undigitized (Note 1, Note 2)	The information is not digitally available. No attempt has been made to digitize it.	Empty value in a skeletal record to which data still need to be added from the label.
unknown:missing (Note 1, Note 2)	The information is not digitally available. It appeared to be absent during digitization.	A value of S.D. ( <i>sine dato</i> ) used by transcription platforms to indicate the absence of a date value.
unknown:indecipherable (Note 1, Note 2)	The information is not digitally available. It appeared to be present during digitization but failed to be captured.	An indication made by a transcriber that they failed to transcribe the information.
known:withheld (Note 3)	The information is digitally available, but it has been withheld by the provider.	A georeferenced record for which coordinate data are available but withheld for conservation reasons.
known:undigitized	The information is known but has not been digitized (i.e., it is not digitally available).	Label information that has not been transcribed because the digitization request/purpose does not require it.
Not applicable	The information element is not applicable	One or more of the information elements 23 onwards in Table 4 may not be applicable. Some information elements are not applicable for some kinds of collection.
<p><b>Note 1: The generic unknown indicates that the information is indeed not available.</b></p> <p><b>Note 2: The additives undigitized, missing and indecipherable allow elaboration as to why the data are unavailable, if this reason is known.</b></p> <p><b>Note 3: known:withheld indicates that the data are digitally available (e.g., in a primary source) and could potentially be retrieved after contacting the data provider.</b></p>		

## 5 Images and other media types (normative)

Many but not all digitization projects and programmes produce images of specimens and their accompanying label(s). It is helpful to users and machines, when searching for available digital specimens to know if images are available. Thus, a MIDS level information element (2) may be supplemented with one or both the designators specified in Table 7.

As part of the minimum information about a specimen, digital specimens should (if possible) include at least one image of the complete specimen and its accompanying label(s). If images are present, the MIDS level of the digital specimen must be supplemented with the indicator 'i' indicating the availability of image(s).

Digital specimens may include multiple images of different types and/or other types of media. If other media type(s) are present, the MIDS level of the digital specimen must be supplemented with the indicator 'm' to indicate the availability of other media type(s).

**Table 7: Media elements of MIDS record types**

MIDS level	Object image type	Recommendation
i	Object image with label	Image should include the whole specimen where possible. Original label(s) should be included either in the same image or as separate images.
m	Other multimedia present	No specific recommendation.

Specimens can require more than one image to properly display all features of interest. Thus, multiple images can be present for a specimen. However, characterising the number and nature of image and other multimedia types available as outputs of digitization processes is outside the scope of the present specification.

## 6 Guidelines for machine-actionability (normative)

To do: <insert text. This is the publishing/sharing aspect that is often treated separate from and after digitization. Publicly available means not only to humans but also for machines to find/access as well. FAIR statement of how MIDS supports each of the 15 principles.>

## 7 Recommendations for collection-holding institutions

MIDS is a minimum standard. As a general principle, institutions should aim to publish the fullest available data about their collections and individual specimens at the earliest opportunity. This means as soon as practically possible after digitization, verification and curation activities have been carried out, expecting that such data can and will be enriched and annotated over time by exposure to a broad expert audience.

The minimum amount of information to be made available (published) must be the expected information elements at any chosen level of digitization.

As best practice, level 2 Regular should be the minimum standard to aim for. However, publishing information in accordance with levels 1 Basic and/or 0 Catalogue and enriching it over time is acceptable.

The information elements expected by MIDS at any level of digitization are the minimum amount of information to be published. As the gold standard, collection-holding institutions are encouraged to publish beyond the minimum where possible, and to aspire to Level 3 Extended.

To do>> We probably want to say something here about how to apply MIDS retrospectively to records already digitized.

## 8 Guarding against publication of junk data

Concerns about publishing junk data are not solved by MIDS. Collection-holding institutions should operate adequate quality controls during digitization processes to ensure that published data are valid data, to the best of knowledge. Attention to prevention of digitization errors, simple (automated) checks and basic rules, such as avoiding unfilled information fields (i.e., use specific 'unknown' or 'not specified' values instead of leaving fields blank) each contribute to ensuring valid data.

## 9 Conformance (normative)

### 9.1 Principle means of conformance

29Jul2021: ICS proforma tables to be revised - not presently uptodate. See current proposal under issue #50.

The principle means of conformance with the present specification must be through implementation of its requirements in workflows and processes of specimen digitization and data management.

Note 7: Tools for automated verification of MIDS levels may become available in the future.

### 9.2 Implementation Conformance Statement (ICS)

The Implementation Conformance Statement (ICS) proforma in Table 8 – Table 11 is a checklist collection-holding institutions and others can use to check and self-declare their conformance with the requirements of the present specification.

In the support column for each item, an assessor must answer the status questions with "yes" or "no" by ticking/unticking the checkboxes. Status values have the following meanings:

m mandatory to support the item/provide the information. o optional to support the item/provide the information. m.N mandatory to support at least one of the items/provide the information in the group N. o.N optional to support one or more of the items/provide the information in the group N. c.N conditional on item N. Answering the items in the table is conditional on the answer for item N.

\_Editor's note: On completion of proforma rows, numbering and cross-referencing to be checked/completed.\_

Table 8: ICS proforma for MIDS conformance (general)

Item	Requirement	Reference	Status	Support
1	Focus on the primary, curated scientific information denoted as being authoritative for the specimen	§3.1	m	<input type="checkbox"/>
2	Publish specimen data at MIDS level 0	§3.5	m.1	<input type="checkbox"/>
3	Publish specimen data at MIDS level 1	§3.5	m.1	<input type="checkbox"/>
4	Publish specimen data at MIDS level 2	§3.5	m.1	<input type="checkbox"/>
5	Publish specimen data at MIDS level 3	§3.5	m.1	<input type="checkbox"/>
6	PID assigned to digital specimen data	§3.4	o	<input type="checkbox"/>
7	PID published alongside MIDS information	§3.4	c6:o	<input type="checkbox"/>
8	Unknown / unspecified datetime form supported i.e., "X*" or "XXXX-XX-XX"	§4.5.1	o	<input type="checkbox"/>
9	?????Issue: remainder of ICS items need to be re-number and cross-references need to be checked and corrected. Need to look into autonumbering to solve this.?????	§4.5.2	o	<input type="checkbox"/>

m.1: It is mandatory to support at least one of these items 2 - 5.

**Table 9: ICS proforma for MIDS level 0 conformance**

c.2: yes, no, n/a* (select according to answer in Table 8). If yes, complete this table.			Choose an item.	
Item	Requirement	Reference	Status	Support
6	CreatedOn	Table 5 1.	m	<input type="checkbox"/>
7	ModifiedOn	Table 5 2.	m	<input type="checkbox"/>
8	RecordCreator i) Name ii) Referent	Table 5 3.	m o	<input type="checkbox"/> <input type="checkbox"/>
9	MidsLevel	Table 5 4.	m	<input type="checkbox"/>
10	PhysicalSpecimenId	Table 5 5.	m	<input type="checkbox"/>
11	Institution i) Code ii) Referent	Table 5 6.	m o	<input type="checkbox"/> <input type="checkbox"/>
12	MaterialType	Table 5 7.	m	<input type="checkbox"/>
13	ScientificName i) Name appropriate to the specimen ii) Referent	Table 5 8.	m o	<input type="checkbox"/> <input type="checkbox"/>
14	Has image(s)	§5	o	<input type="checkbox"/>
15	Has other media type(s)	§5	o	<input type="checkbox"/>

**Table 10: ICS proforma for MIDS level 1 conformance**

c.3: yes, no, n/a* (select according to answer in Table 8). If yes, complete this table.			Choose an item.	
Item	Requirement	Reference	Status	Support
16	CreatedOn	Table 2 1.	m	<input type="checkbox"/>
17	ModifiedOn	Table 2 2.	m	<input type="checkbox"/>
18	RecordCreator i) Name ii) Referent	Table 2 3.	m o	<input type="checkbox"/> <input type="checkbox"/>
19	MIDSLevel	Table 2 4.	m	<input type="checkbox"/>
20	PhysicalSpecimenId	Table 2 5.	m	<input type="checkbox"/>
21	Institution i) Code ii) Referent	Table 2 6.	m o	<input type="checkbox"/> <input type="checkbox"/>
22	MaterialType	Table 2 7.	m	<input type="checkbox"/>
23	ScientificName i) Name appropriate to the specimen ii) Referent	Table 2 8.	m o	<input type="checkbox"/> <input type="checkbox"/>
24	Has image(s)	§5	o	<input type="checkbox"/>
25	Has other media type(s)	§5	o	<input type="checkbox"/>

**Table 11: ICS proforma for MIDS level 2 conformance**

\_Editor's note: Proforma to be completed.\_

c.4: yes, no, n/a* (select according to answer in Table 8). If yes, complete this table.			Choose an item.	
Item	Requirement	Reference	Status	Support
26	CreatedOn	Table 3 1.	m	<input type="checkbox"/>
27	ModifiedOn	Table 3 2.	m	<input type="checkbox"/>
28	RecordCreator i) Name ii) Referent	Table 3 3.	m o	<input type="checkbox"/> <input type="checkbox"/>
29	MIDSLevel	Table 3 4.	m	<input type="checkbox"/>
30	PhysicalSpecimenId	Table 3 5.	m	<input type="checkbox"/>
31	Institution i) Code ii) Referent	Table 3 6.	m o	<input type="checkbox"/> <input type="checkbox"/>
32	MaterialType	Table 3 7.	m	<input type="checkbox"/>
33	ScientificName i) Name appropriate to the specimen	Table 3 8.	m	<input type="checkbox"/>

	ii) Referent		<input type="radio"/>	<input type="checkbox"/>
	<i>Editor's note: Proforma to be completed.</i>			
nn	Has image(s)	\$5	<input type="radio"/>	<input type="checkbox"/>
nn	Has other media type(s)	\$5	<input type="radio"/>	<input type="checkbox"/>

**Table 12: ICS proforma for MIDS level 3 conformance**

*Editor's note: Proforma to be completed.*

c.5: yes, no, n/a* (select according to answer in Table 8). If yes, complete this table.			Choose an item.	
Item	Requirement	Reference	Status	Support
	CreatedOn	Table 4 1.	m	<input type="checkbox"/>
	ModifiedOn	Table 4 2.	m	<input type="checkbox"/>
	Unknown / unspecified datetime form i.e., "X*" or "XXXX-XX-XX"	Table 4 Note 1	<input type="radio"/>	<input type="checkbox"/>
	RecordCreator i) Name ii) Referent	Table 4 3.	m o	<input type="checkbox"/> <input type="checkbox"/>
	MIDSLevel	Table 4 4.	m	<input type="checkbox"/>
	PhysicalSpecimenId	Table 4 5.	m	<input type="checkbox"/>
	Institution i) Code ii) Referent	Table 4 6.	m o	<input type="checkbox"/> <input type="checkbox"/>
	MaterialType	Table 4 7.	m	<input type="checkbox"/>
	ScientificName i) Name appropriate to the specimen ii) Referent	Table 4 8.	m o	<input type="checkbox"/> <input type="checkbox"/>
	<i>Editor's note: Proforma to be completed.</i>			
	Has image(s)	\$5	<input type="radio"/>	<input type="checkbox"/>
	Has other media type(s)	\$5	<input type="radio"/>	<input type="checkbox"/>

**Table 13: ICS proforma for handling of unknown or incomplete data**

Item	Requirement	Reference	Status	Support
13.1	Unknown / unspecified datetime "X*" form (ISO 8601-2)	\$4.5.1	<input type="radio"/> .2	<input type="checkbox"/>
13.2	Unknown / unspecified datetime "XXXX-XX-XX" form (ISO 8601-2)	\$4.5.1	<input type="radio"/> .2	<input type="checkbox"/>
13.3	... ..	\$4.5.2	<input type="radio"/>	<input type="checkbox"/>
13.4	Unknown data term "unknown"	\$4.5.3	<input type="radio"/>	<input type="checkbox"/>
13.5	Unknown data term "unknown:undigitized"	\$4.5.3	c13.3:o.3	<input type="checkbox"/>
13.6	Unknown data term "unknown:missing"	\$4.5.3	c13.3:o.3	<input type="checkbox"/>
13.7	Unknown data term "unknown:indcipherable"	\$4.5.3	c13.3:o.3	<input type="checkbox"/>
13.8	Withheld data term "known:withheld"	\$4.5.3	<input type="radio"/>	<input type="checkbox"/>
13.9	Information element not applicable	\$4.5.3	<input type="radio"/>	<input type="checkbox"/>

*o.2: It is optional to support one of these items 13.1 and 13.2. \*

*o.3: IF c13.4, optional to support items 13.5 - 13.7 but all items must be supported.*

## 10 Illustrative use cases and guidelines for use (informative)

*\_To do>> include some illustrative use cases here, drawn from real practice to show what information would typically be published for that use case when complying with, for example MIDS level 2.\_*

*Still to do>>*

- Levels need to link to user stories: A certain level should be able to serve some specific stories. Analyse MIDS against DiSSCo user stories – see which levels are needed by which user stories to see what this tells us. Cross-refer to Niels article on user stories.*
- Review levels against ICEDIG Hannu's MS37/D6.5 proposals to ensure we align.*

## 11 References

[Brummitt 2001] Brummitt, R.K. (2001). World Geographical Scheme for Recording Plant Distributions, Edition 2. Biodiversity Information Standards (TDWG). <http://www.tdwg.org/standards/109>.

[FMI 2018] Flanders Marine Institute (2018). IHO Sea Areas, version 3. Available online at <http://www.marineregions.org/>. doi: 10.14284/323.

[Groom 2019] Groom G, Dillen M, Hardy H, Phillips S, Willemsse L, Wu Z. (2019). Improved standardization of transcribed digital specimen data. Database, Volume 2019, baz129. doi: 10.1093/database/baz129.

[ISO 3166] International Standards Organisation (ISO) (2013). ISO 3166. Codes for the Representation of Names of Countries and Their Subdivisions. Part 1: Country codes; Part 2: Country subdivision code; and Part 3: Code for formerly used names of countries. [www.iso.org](http://www.iso.org).

[ISO 8601] International Standards Organisation (ISO) (2019). ISO 8601. Date and Time — Representations for Information Interchange. Part 1: Basic Rules; and Part 2: Extensions. [www.iso.org](http://www.iso.org).

[Meyer 2015] Meyer, Eric T., and Ralph Schroeder. Knowledge machines: digital transformations of the sciences and humanities. MIT Press, 2015.

[Mons 2017] Mons B, Neylon CD, Velterop J, Dumontier M, da Silva Santos LOB, Wilkinson M (2017) Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information Services and Use, vol. 37, no. 1, pp. 49-56, 2017. doi: 10.3233/ISU-170824.

[Nelson 2015] Nelson, G., Sweeney, P., Wallace, L. E., Rabaler, R. K., Allard, D., Brown, H., ... & Gilbert, E. (2015). Digitization workflows for flat sheets and packets of plants, algae, and fungi. Applications in Plant Sciences, 3(9), 1500065. doi: 10.3732/apps.1500065.

[RFC 2119] IETF RFC 2119. (1997). Key words for use in RFCs to Indicate Requirement Levels. url: <https://www.ietf.org/rfc/rfc2119.txt>.

[Schindel 2018] Schindel DE, Cook JA. The next generation of natural history collections. PLoS Biology. 2018 Jul 16;16(7):e2006125. doi: 10.1371/journal.pbio.2006125.

[Webster 2017] Webster, MS (ed). The Extended Specimen: Emerging Frontiers in Collections-based Ornithological Research. CRC Press, 2017. ISBN 978-1-315-12045-4.

[Wilkinson 2016] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3. doi: 10.1038/sdata.2016.18.

END.

---