

DiSSCo related output

This template collects the required metadata to reference the official Deliverables and Milestones of DiSSCo-related projects. More information on the mandatory and conditionally mandatory fields can be found in the supporting document 'Metadata for DiSSCo Knowledge base' that is shared among work package leads, and in Teamwork > Files. A short explanatory text is given for all metadata fields, thus allowing easy entry of the required information. If there are any questions, please contact us at info@dissco.eu.

Title

DiSSCo Prepare Milestone Report MS5.5 "Compilation of relevant data standards"

Author(s)

Petersen, Mareike
von Mering, Sabine
Glöckler, Falko

Identifier of the author(s)

<https://orcid.org/0000-0001-8666-1931> (MP)
<https://orcid.org/0000-0003-2982-7792> (SvM)
<https://orcid.org/0000-0002-7127-2738> (FG)

Affiliation

Museum für Naturkunde – Leibniz Institute for
Evolution and Biodiversity Science
(<https://ror.org/052d1a351>)

Contributors

Addink, Wouter
Hardisty, Alex
Abarenkov, Kessy
Fichtmüller, David

Publisher

Identifier of the publisher

Resource ID

<https://doi.org/10.34960/3mg1-7n14>

Publication year

2021

Related identifiers

Is it the first time you submit this outcome?

Yes

Creation date

15/01/2021

Version

1

Citation

Petersen, M., von Mering, S. & Glöckler, F. 2021: DiSSCo Prepare Milestone Report MS5.5 "Compilation of relevant data standards". <https://doi.org/10.34960/3mg1-7n14>

Abstract

This Milestone Report covers data standards and ontologies relevant for the DiSSCo research infrastructure. Together with other project outputs of DiSSCo Prepare this report is identifying and documenting important building blocks for the technical Infrastructure of DiSSCo. The compilation will be available via the currently developed DiSSCo Knowledgebase.

Content keywords

scientific

Project reference
DiSSCo Prepare (GA-871043)

WP number
WP5

Project output
Milestone report

Deliverable/milestone number
MS5.5

Dissemination level
Public

Rights

License
CC0 1.0 Universal (CC0 1.0)

Resource type
Text

Format
PDF

Funding Programme
H2020-INFRADEV-2019-2

Contact email
sabine.vonmering@mfn.berlin



DiSSCo Prepare

WP 5 Common Resources and Standards –

MS 5.5 Compilation of relevant data standards

WP Lead: Mareike Petersen (MfN)

Task Lead: Mareike Petersen & Falko Glöckler (both MfN)

Authors: Mareike Petersen, Sabine von Mering & Falko Glöckler (all MfN)

Contributors: Kessy Abarenkov (Tartu), Wouter Addink (Naturalis), David Fichtmüller (BGBM), Alex Hardisty (Cardiff)

Abstract

This Milestone Report covers data standards and ontologies relevant for the DiSSCo research infrastructure. Together with other project outputs of DiSSCo Prepare this report is identifying and documenting important building blocks for the technical Infrastructure of DiSSCo. The compilation will be available via the currently developed DiSSCo Knowledgebase.

Key words

Knowledgebase, ontology, data standard



H2020-INFRADEV-2019-2
Grant Agreement No 871043

Grant Agreement number: 871043 — DiSSCo Prepare — H2020-INFRADEV-2018-2020 / H2020-INFRADEV-2019-2

INDEX

01	INTRODUCTION	5
02	APPROACH	5
03	RESULTS	6
	Data Standards	6
	a. Standards relevant for the description of specimen and Bio- and Geodiversity	6
	Access to Biological Collection Data	6
	Access to Biological Collection Data Extension for Geosciences	7
	CETAF Specimen Preview Profile	Fehler! Textmarke nicht definiert.
	Darwin Core	8
	Ecological Metadata Language	8
	Extended Specimen Concept	9
	Humboldt Core	9
	Open Digital Specimen	10
	Taxonomic Concept Transfer Schema	10
	b. Standards relevant for the description of collections	11
	Collection Description Standard	11
	Minimum Information about a Digital Specimen	11
	c. Standards relevant for the description of tissues and sequences	12
	Global Genome Biodiversity Network Data Standard	12
	Minimum Information about any (x) Sequence	12
	d. Standards relevant for the description of digital multimedia assets	13
	Audubon Core	13
	e. Standards related to localities	13
	Geography Markup Language	13
	ISO 3166 – 1	14

ISO 3166 – 2	14
f. Other relevant standards	15
Ecological Trait-data standard	15
Bioschemas.org	15
CloudEvents	16
Data Quality Vocabulary	16
Datacite Metadata Schema	17
Digital Object Interface Protocol	17
FAIR Digital Object Framework	18
FAIR Implementation Profile	18
Schema.org	19
Structured Descriptive Data	19
W3C PROV	20
Web Annotation Data Model	20
Ontologies	21
PROV-O	21
Biological Collections Ontology	21
Ontology for Biomedical Investigations	21
Information Artifact Ontology	22
Open Digital Specimen Ontology	22
MarineTLO ontology	22
04 OUTLOOK	23

01 INTRODUCTION

DiSSCo Prepare will build on profound technical knowledge from various sources and initiatives. In order to allow for efficient knowledge and technology transfer for partners building the DiSSCo technical backbone, a central and freely-accessible DiSSCo Knowledgebase will be designed and implemented.

The knowledgebase will include structured technical documentation of identified DiSSCo technical building blocks, an assessment of their technical readiness for DiSSCo, as well as specifications on their relevance for the overall DiSSCo technical infrastructure and the DiSSCo data model.

In this first Milestone Report linked to the Knowledgebase development we compile a list of relevant data standards and ontologies for bio- and geo-collection objects, collection descriptions, and digital assets. Relevant Persistent Identifier (PID) Systems are in focus of MS 5.3 “Documentation of PIDs relevant for DiSSCo technical infrastructure” due in Month 18 and are not part of this Report.

Data standards are the rules by which data are described and recorded. In order to share, exchange, and understand data, we must standardize the format as well as the meaning¹. Ontologies are usually domain specific and describe categories properties and relations between concepts, data, or entities.

The compilation will be available via the DiSSCo Knowledgebase (currently under development). This list of standards and further building blocks to be prepared will be an important reference for DiSSCo Preparatory Phase (DPP) project partners, tasks within the project as well as tasks and developments in other DiSSCo related projects (e.g. Helpdesk developed under Synthesys+).

02 APPROACH

Relevant building blocks were collected from known organizations coping with the standardization, publication and exchange of data related bio- and geodiversity and beyond. Especially important for the development of standards in our domain is Biodiversity Information Standards (TDWG). This not-for-profit, scientific and educational association was formed to establish international collaboration among the creators, managers and users of biodiversity information and to promote the wider and more effective dissemination and sharing of knowledge about biological organisms. The TDWG community develops, ratifies and promotes standards and guidelines for the recording and exchange of data about organisms and brings together experts/specialists in different TDWG Interest and Task Groups working on specific topics and standards.

Other important organizations with respect to relevant outputs are the Research Data Alliance and the World Wide Web Consortium. The Research Data Alliance (RDA) is a community-driven international initiative with the goal of sharing and re-use of data by building the technical but also societal infrastructure. It can be defined by its vision: “Researchers and innovators openly share and re-use data across technologies, disciplines, and countries to address the grand challenges of society.” and mission: “RDA builds the social and technical bridges that

¹ U.S. Geological Survey: Data Standards:

<https://www.usgs.gov/products/data-and-tools/data-management/data-standards#:~:text=Data%20standards%20are%20the%20rules,as%20well%20as%20the%20meaning>

enable open sharing and re-use of data.” Many of the RDA recommendations and developments will be part of the data lifecycle in DiSSCo². The World Wide Web Consortium (W3C) is an international community where Member organizations, a full-time staff, and the public work together to develop Web standards. The W3C mission is to lead the World Wide Web to its full potential by developing protocols and guidelines that ensure the long-term growth of the Web³.

03 RESULTS

Data Standards

Data standards were collected from relevant sources in a tabular format including information on Name, Focus, Status, Maintenance, Documentation, Citation, and Further reading. The standards are grouped according to their field of application. In each subsection the standards are in alphabetical order. Although the list is already very comprehensive, further standards might be of interest or relevance during ongoing development of the DiSSCo Research Infrastructure.

a. Standards relevant for the description of specimen and Bio- and Geodiversity

Name	Access to Biological Collection Data
Focus	The Access to Biological Collection Data (ABCD) Schema is an evolving comprehensive standard for the access to and exchange of data about specimens and observations (a.k.a. primary biodiversity data). Use cases are e.g. GBIF, GFBio, Europeana.
Maintenance	TDWG current (2005) standard (Version 2.06), a new version (3.0) is already available but not yet ratified by TDWG; maintained by the ABCD Maintenance Group
Documentation	https://github.com/tdwg/abcd/ https://abcd.tdwg.org/terms/
Citation	Access to Biological Collection Data task group. 2007. Access to Biological Collection Data (ABCD), Version 2.06. Biodiversity Information Standards (TDWG) http://www.tdwg.org/standards/115
Further reading	ABCD 3.0 is the new version of the standard Fichtmüller D, Berendsohn W, Droege G, Glöckler F, Güntsch A, Hoffmann J, Holetschek J, Petersen M, Reimeier F (2019) ABCD 3.0 Ready to Use. Biodiversity Information Science and Standards 3: e37214. https://doi.org/10.3897/biss.3.37214

² Islam S, Hardisty A, Addink W, Weiland C and Glöckler F (2020) Incorporating RDA Outputs in the Design of a European Research Infrastructure for Natural Science Collections. Data Science Journal, 19(1), p.50. DOI: <http://doi.org/10.5334/dsj-2020-050>

³ World Wide Web Consortium: <https://www.w3.org/Consortium/>

Name	Access to Biological Collection Data Extension for Geosciences
Focus	Extension for Geosciences (EFG) extends the ABCD schema to integrate geological data, including information about paleontological and geological collection items. Use cases are e.g. GeoCAsE Portal, Europeana.
Maintenance	Since March 2017 under ratification by TDWG, maintained by the TDWG EFG Task Group
Documentation	https://github.com/tdwg/efg
Citation	<i>Original Citation:</i> Kiessling W, Rissoné A, Copp C, Döring M, Mewis H (2006): The EFG extension to the ABCD schema
Further reading	Petersen M, Glöckler F, Kiessling W, Döring M, Fichtmüller D, Laphakorn L, Baltruschat B, Hoffmann J (2018). History and development of ABCDEFG: a data standard for geosciences. Fossil Record, 21(1), 47-53. https://doi.org/10.5194/fr-21-47-2018

Name	CETAF Specimen Preview Profile
Focus	The CETAF Specimen Preview Profile (CSPP) is a minimal set of agreed (RDF) collection metadata elements implemented consistently across CETAF (Consortium of European Taxonomic Facilities) organisations. The purpose of CSPP is to provide a stable resource enabling preview functions in specimen portals. The CSPP is not meant to be comprehensive, which means that Linked Open (collection) Data implementations of CETAF institutions will usually provide much richer metadata with additional RDF-elements. Use case e.g. CETAF/JACQ Botany Pilot (https://services.bgbm.org/botany/pilot/person/q/Q1349394).
Maintenance	CETAF ISTC (Information Science & Technology Commission); actively supported by many CETAF institutions, more to come
Documentation	https://cetafidentifiers.biowikifarm.net/wiki/CETAF_Specimen_Preview_Profile_(CSPP)
Citation	-
Further reading	https://cetafidentifiers.biowikifarm.net/wiki/Questions,_problem_solutions_and_further_discussions_(Guide_of_best_practices)#CETAF_Specimen_Preview_Profile_.28CSPP.29

Name	Darwin Core
Focus	Sharing of information about biological diversity. Darwin Core is primarily based on taxa, their occurrence in nature as documented by observations, specimens, samples, and related information. Multiple extensions exist. Use e.g. GBIF
Maintenance	TDWG Current Standard, maintained by the TDWG Darwin Core Maintenance Interest Group
Documentation	https://github.com/tdwg/dwc https://dwc.tdwg.org/terms/
Citation	Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, et al. (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7(1): e29715. https://doi.org/10.1371/journal.pone.0029715
Further reading	-

Name	Ecological Metadata Language
Focus	The Ecological Metadata Language (EML) is a metadata standard developed by the ecology discipline and for the ecology discipline. It is based on prior work done by the Ecological Society of America and associated efforts (Michener et al., 1997, Ecological Applications). EML is implemented as a series of XML document types that can be used in a modular and extensible manner to document ecological data. Each EML module is designed to describe one logical part of the total metadata that should be included with any ecological dataset.
Maintenance	EML project
Documentation	https://eml.ecoinformatics.org/ https://github.com/NCEAS/eml
Citation	Jones MB, O'Brien M, Mecum B, Boettiger C, Schildhauer M, Maier M, Whiteaker T, Earl S, Chong S (2019) Ecological Metadata Language version 2.2.0. KNB Data Repository. doi:10.5063/F11834T2
Further reading	Michener WK, Brunt JW, Helly JJ, Kirchner TB, & Stafford, SG (1997) Nongeospatial metadata for the ecological sciences. Ecological Applications, 7(1), 330-342.

Name	Extended Specimen Concept
Focus	“The concept of an extended specimen (Webster 2017) conveys the current perspective of the biodiversity specimen as extending beyond the singular physical object to potentially limitless additional physical preparations and digital resources. [...] These extensions will scale from molecules to the ecosphere, and would include genetic, phenotypic, behavioral, and environmental data, as well as biotic interaction networks and new multimedia components (e.g., 2D and 3D specimen images, in situ field images, videos of field conditions).” (Lendemer et al., 2019)
Maintenance	Extended Specimen Network (ESN), in active use
Documentation	-
Citation	Webster M (Ed.) (2017) The Extended Specimen- Emerging Frontiers in Collections-Based Ornithological Research. CRC Press, 240 pp. [ISBN 9781498729154]
Further reading	Lendemer J, Thiers B, Monfils AK, Zaspel J, Ellwood ER, Bentley A, LeVan K, Bates J, Jennings D, Contreras D, Lagomarsino L, Mabee P, Ford LS, Guralnick R, Gropp RE, Revelez M, Cobb N, Seltmann K, Aime MC (2019) The Extended Specimen Network: A Strategy to Enhance US Biodiversity Collections, Promote Research and Education. <i>BioScience</i> 70 (1): 23-30. https://doi.org/10.1093/biosci/biz140

Name	Humboldt Core
Focus	Humboldt Core is a community-developed standard for representing critical information about scope, method and completeness of biological inventories. [...] [I]t provides a means for standardized capture of information that are typically reported in any inventory. The standard has been developed to be usable across the wide range of inventories that are performed, and has been rigorously tested to assure its usability. Terms in the Humboldt Core have been carefully cross-walked to other biodiversity data standards to assure compatibility where possible with other data dictionaries. Use case e.g. Map of Life.
Maintenance	
Documentation	https://mol.org/humboldtcare/
Citation	Guralnick R, Walls R, Jetz W (2018) Humboldt Core – toward a standardized capture of biological inventories for biodiversity monitoring, modeling and assessment. <i>Ecography</i> , 41: 713-725. https://doi.org/10.1111/ecog.02942
Further reading	https://www.tdwg.org/community/osr/humboldt-core/ https://github.com/mapoflife/humboldtcare

Name	Open Digital Specimen
Focus	Open Digital Specimen (openDS) is a new Standard for digital specimens and other natural science digital object types.
Maintenance	under active development in DiSSCo
Documentation	https://github.com/DiSSCo/openDS
Citation	-
Further reading	Hardisty A, Ma K, Nelson G, Fortes J (2019) 'openDS' – A New Standard for Digital Specimens and Other Natural Science Digital Object Types. Biodiversity Information Science and Standards 3: e37033. https://doi.org/10.3897/biss.3.37033 Addink W, Hardisty AR (2020) 'openDS' – Progress on the New Standard for Digital Specimens. Biodiversity Information Science and Standards 4: e59338. https://doi.org/10.3897/biss.4.59338

Name	Taxonomic Concept Transfer Schema
Focus	The Taxonomic Concept Transfer Schema (TCS) schema was conceived to allow the representation of taxonomic concepts as defined in published taxonomic classifications, revisions and databases. As such, it specifies the structure for XML documents to be used for the transfer of defined concepts. Valid transfer documents may either explicitly detail the defining components of taxon concepts, transfer GUIDs referring to defined taxon concepts (if and when these are available) or a mixture of the two.
Maintenance	TDWG Current (2005) standard, maintained by the TDWG Taxonomic Names and Concepts Interest Group
Documentation	https://www.tdwg.org/standards/tcs/
Citation	Taxonomic Names and Concepts Interest Group (2006) Taxonomic Concept Transfer Schema (TCS), version 1.01. Biodiversity Information Standards (TDWG) http://www.tdwg.org/standards/117
Further reading	https://github.com/tdwg/tcs

b. *Standards relevant for the description of collections*

Name	Collection Description Standard
Focus	The Collections Descriptions (CD) Standards describes groups of natural collection objects.
Maintenance	TDWG Draft Standard, maintained by the TDWG Collection Description Interest Group. Use case e.g. Synthesys+ Collections Description Dashboard.
Documentation	https://github.com/tdwg/cd/ https://www.tdwg.org/community/cd/
Citation	Collections Descriptions interest group (2019) Collection Descriptions (CD), in development. Biodiversity Information Standards (TDWG) http://www.tdwg.org/standards/
Further reading	-

Name	Minimum Information about a Digital Specimen
Focus	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. Use case e.g. Synthesys+ Collections Description Dashboard.
Maintenance	TDWG Draft Standard, maintained by the TDWG MIDS Task Group
Documentation	https://github.com/tdwg/cd/ https://www.tdwg.org/community/cd/mids/
Citation	See current draft at https://github.com/tdwg/mids/tree/working-draft/current-draft
Further reading	Hardisty A, Addink W, Dillen M, Groom Q, Haston E et al. (<i>Draft</i>) Minimum Information about a Digital Specimen (MIDS) v0.11, July 2020 . Borsch T, Stevens A-D, Häffner E, Güntsch A, Berendsohn WG et al. (2020) A complete digitization of German herbaria is possible, sensible and should be started now. Research Ideas and Outcomes 6: e50675 . MISC (2012) iDigBio MISC Data Element Catalog (Phase 1, V0, rev. 15 December 2012)

c. *Standards relevant for the description of tissues and sequences*

Name	Global Genome Biodiversity Network Data Standard
Focus	The Global Genome Biodiversity Network (GGBN) Data Standard is a set of vocabularies designed to represent tissue, DNA or RNA samples associated with voucher specimens, tissue samples and collections. Use e.g. GGBN Data Portal.
Maintenance	Under Ratification of TDWG, maintained by GGBN
Documentation	https://wiki.ggbn.org/ggbn/GGBN_Data_Standard_v1 https://wiki.ggbn.org/ggbn/Documents
Citation	Droege G, Barker K, Seberg O, Coddington J, Benson E, Berendsohn WG et al. (2016) The Global Genome Biodiversity Network (GGBN) Data Standard specification. <i>Database</i> , Volume 2016, baw125, https://doi.org/10.1093/database/baw125
Further reading	https://wiki.ggbn.org/ggbn/Documents

Name	Minimum Information about any (x) Sequence
Focus	Minimum Information about any (x) Sequence (MIxS) is the core standards of the Genomic Standards Consortium for describing genomes, metagenomes and gene marker sequences.
Maintenance	Genomic Standards Consortium
Documentation	https://gensc.org/mixs/ https://github.com/GenomicsStandardsConsortium
Citation	-
Further reading	https://gensc.org/projects/mixs-gsc-project/ Yilmaz P, Kottmann R, Field D (2011) Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. <i>Nat Biotechnol</i> 29 , 415–420, https://doi.org/10.1038/nbt.1823 Field D, Garrity G, Gray T et al. (2008) The minimum information about a genome sequence (MIGS) specification. <i>Nat Biotechnol</i> 26 , 541–547, https://doi.org/10.1038/nbt1360

d. Standards relevant for the description of digital multimedia assets

Name	Audubon Core
Focus	The Audubon Core (AC) is a set of vocabularies designed to represent metadata for biodiversity multimedia resources and collections. These vocabularies aim to represent information that will help to determine whether a particular resource or collection will be fit for some particular biodiversity science application before acquiring the media. Among others, the vocabularies address such concerns as the management of the media and collections, descriptions of their content, their taxonomic, geographic, and temporal coverage, and the appropriate ways to retrieve, attribute and reproduce them.
Maintenance	TDWG Current Standard, maintained by the TDWG Audubon Core Maintenance Group
Documentation	https://github.com/tdwg/ac
Citation	-
Further reading	https://tdwg.github.io/ac/introduction https://tdwg.github.io/ac/structure https://tdwg.github.io/ac/termlist https://tdwg.github.io/ac/guide Morris et al. 2013. Discovery and Publishing of Primary Biodiversity Data associated with Multimedia Resources: The Audubon Core Strategies and Approaches. Biodiversity Informatics, 8(2). https://doi.org/10.17161/bi.v8i2.4117

e. Standards related to localities

Name	Geography Markup Language
Focus	The Geography Markup Language (GML) is a XML grammar defined by the Open Geospatial Consortium (OGC) to express geographical features. GML serves as a modeling language for geographic systems as well as an open interchange format for geographic transactions on the Internet. Key to GML's utility is its ability to integrate all forms of geographic information, including not only conventional "vector" or discrete objects, but coverages and sensor data.
Maintenance	Open Geospatial Consortium
Documentation	http://www.opengis.net/spec/GML/3.3
Citation	-
Further reading	https://www.ogc.org/

Name	ISO 3166 – 1
Focus	Codes for the representation of names of countries and their subdivisions – Part 1: Country code. Standard defining codes for the names of countries, dependent territories, and special areas of geographical interest
Maintenance	International Organization for Standardization (ISO)
Documentation	https://isotc.iso.org/livelink/livelink?func=ll&objId=16944257
Citation	
Further reading	https://www.iso.org/iso-3166-country-codes.html

Name	ISO 3166 – 2
Focus	Codes for the representation of names of countries and their subdivisions – Part 2: Country subdivision code. Standard defining codes for country subdivisions (administrative territorial entity). Usually only the top levels of a country are assigned ISO 3166-2 codes, but sometimes there is a second level hierarchy.
Maintenance	International Organization for Standardization (ISO), actively maintained (2-4 updates a year)
Documentation	https://www.iso.org/obp/ui/#iso:std:iso:3166:-2:ed-4:v1:en Subdivisions of an individual country can be viewed using the ISO Online Browsing Platform with the ISO 3166-1 alpha 2 country code: <a href="https://www.iso.org/obp/ui/#iso:code:3166:<code>">https://www.iso.org/obp/ui/#iso:code:3166:<code> , e.g. https://www.iso.org/obp/ui/#iso:code:3166:DE
Citation	<i>Current Version:</i> ISO 3166-2:2020(en) Codes for the representation of names of countries and their subdivisions — Part 2: Country subdivision code, Committee: ISO/TC 46, Edition: 4, Publication date: 2020-08, ICS: 01.140.30.
Further reading	https://en.wikipedia.org/wiki/ISO_3166-2 (note that the lists provided by Wikipedia and Wikidata are in some cases outdated)

f. *Other relevant standards*

Name	Ecological Trait-data standard
Focus	This defined vocabulary aims at providing all essential terms to describe datasets of functional trait measurements and facts for ecological research. Use case e.g. GFBio (German Federation for Biological Data)
Maintenance	ETS Vocabulary is under continuous and open community development, hosted on GitHub. It has been ported to the GFBio Terminology Service.
Documentation	https://terminologies.gfbio.org/terms/ets/pages/
Citation	Schneider FD, Jochum M, Provost GL, Ostrowski A, Penone C, Fichtmüller D, Güntsch A, Gossner MM, König-Ries B, Manning P, Simons NK (2019). Towards an Ecological Trait-data Standard, <i>Methods in Ecology and Evolution</i> , 10:2006-2019
Further reading	https://github.com/EcologicalTraitData/ETS

Name	Bioschemas.org
Focus	Bioschemas aims to improve the Findability on the Web of life sciences resources such as datasets, software, and training materials. It does this by encouraging people in the life sciences to use Schema.org markup in their websites so that they are indexable by search engines and other services. Use case e.g. MNHN Paris, see also https://bioschemas.org/liveDeploys/
Maintenance	different types and profiles of the standard are maintained by different groups, active community
Documentation	https://bioschemas.org/types/ https://bioschemas.org/profiles/
Citation	https://bioschemas.org Gray AJG, Goble CA and Jimenez R (2017) Bioschemas: From Potato Salad to Protein Annotation. In International Semantic Web Conference (Posters, Demos & Industry Tracks)."
Further reading	https://bioschemas.org/ https://github.com/BioSchemas/specifications

Name	CloudEvents
Focus	CloudEvents is a specification for describing event data in common formats to provide interoperability across services, platforms and systems. CloudEvents has received a large amount of industry interest, ranging from major cloud providers to popular SaaS companies. Approved as a Cloud Native sandbox level project on May 15, 2018
Maintenance	hosted by the Cloud Native Computing Foundation (CNCF), actively developed
Documentation	https://github.com/cloudevents
Citation	
Further reading	https://cloudevents.io/

Name	Data Quality Vocabulary
Focus	“The Data Quality Vocabulary (DQV) provides a metadata model for expressing data quality. DQV was developed by the Data on the Web Best Practice (DWBP) Working Group of the World Wide Web Consortium (W3C) between 2013 and 2017.”
Maintenance	W3C Working Group Note 15 December 2016, maintained by the Data on the Web Best Practice (DWBP) Working Group of the W3C
Documentation	https://www.w3.org/TR/vocab-dqv/
Citation	Albertoni R and Isaac A (2021) Introducing the Data Quality Vocabulary (DQV). Semantic Web, Volume 12, 1 Jan. 2021 : 81 – 97. https://doi.org/10.3233/SW-200382
Further reading	https://github.com/w3c/dwbp

Name	Datacite Metadata Schema
Focus	The DataCite Metadata Schema is a list of core metadata properties chosen for an accurate and consistent identification of a resource for citation and retrieval purposes, along with recommended use instructions. It is used for DOI creation for research data via DataCite.
Maintenance	DataCite Metadata Working Group, actively maintained, current release 4.3 from 2019-08-16
Documentation	https://schema.datacite.org/
Citation	<i>For the documentation:</i> DataCite Metadata Working Group. (2019). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.3. DataCite e.V. https://doi.org/10.14454/7xq3-zf69 <i>For the schema:</i> DataCite Metadata Working Group. (2019). DataCite Metadata Schema for the Publication and Citation of Research Data. Version 4.3. DataCite e.V. https://doi.org/10.14454/f2wp-s162
Further reading	
Use Cases	DOI creation for research data

Name	Digital Object Interface Protocol
Focus	The Digital Object Interface Protocol (DOIP) is a conceptual protocol for software applications (“clients”) to interact with “services” which could be either the digital objects or the information systems that manage those digital objects.
Maintenance	DONA Foundation
Documentation	https://www.dona.net/sites/default/files/2018-02/Protocol_Specification.pdf
Citation	Reilly S (2009) Digital Object Protocol Specification, Version: 1.0 12 November 2009
Further reading	https://www.dona.net/doipv1doc https://www.dona.net/suggested-reading-documents

Name	FAIR Digital Object Framework
Focus	FAIR Digital Object Framework (FDOF) is guided by the FAIR Principles and combines the powerful features from both Digital Objects and Linked Data. The FAIR Digital Object has the potential to serve as the “center of the hourglass” for an emerging Internet of FAIR Data and Services. It is giving users a definite but minimal standard to build toward, yet also the freedom to choose their own implementation paths.
Maintenance	Draft Standard, maintained by an international consortium
Documentation	https://github.com/GEDE-RDA-Europe/GEDE/tree/master/FAIR%20Digital%20Objects
Citation	-
Further reading	https://www.go-fair.org/today/fair-digital-framework/Joint-Statement-on-FAIR-Digital-Object-Framework Kahn R and Wilensky R (2006) A framework for distributed digital object services. International Journal on Digital Libraries 6(2): 115–123, DOI: 10.1007/s00799-005-0128-x

Name	FAIR Implementation Profile
Focus	The FAIR Implementation Profile (FIP) is a collection of FAIR implementation choices made by a community of practice for each of the FAIR Principles. Community specific FAIR Implementation Profiles are themselves captured as FAIR datasets and are made openly available to other communities for reuse.
Maintenance	FIP Working Group, under active development
Documentation	-
Citation	Magagna, B, et al. 2020. Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. OSF Preprints. https://doi.org/10.31219/osf.io/2p85g
Further reading	https://www.go-fair.org/how-to-go-fair/fair-implementation-profile/

Name	Schema.org
Focus	<p>Schema.org is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond.</p> <p>Schema.org vocabulary can be used with many different encodings, including RDFa, Microdata and JSON-LD. These vocabularies cover entities, relationships between entities and actions, and can easily be extended through a well-documented extension model. These vocabularies are used by 10 mio. sites, incl. Google, Microsoft, Pinterest, Yandex and others.</p>
Maintenance	W3C Schema.org Community Group
Documentation	https://github.com/schemaorg/schemaorg https://schema.org/docs/documents.html
Citation	
Further reading	https://schema.org/

Name	Structured Descriptive Data
Focus	<p>The goal of the Structured Descriptive Data (SDD) standard is to allow capture, transport, caching and archiving of descriptive data (data describing a taxon or specimen), using a platform- and application-independent, international standard. Such a standard is crucial to enabling lossless porting of data between existing and future software platforms including identification, data-mining and analysis tools, and federated databases.</p>
Status	TDWG Current (2005) standard, maintained by the TDWG Biological Descriptions Interest Group
Documentation	https://github.com/tdwg/sdd
Citation	Hagedorn G, Thiele K, Morris R, Heidorn PB 2005. Structured Descriptive Data (SDD) w3c-xml-schema, Version 1.0. Biodiversity Information Standards (TDWG) http://www.tdwg.org/standards/116
Further reading	-

Name	W3C PROV
Focus	Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness. The PROV Family of Documents defines a model, corresponding serializations and other supporting definitions to enable the inter-operable interchange of provenance information in heterogeneous environments such as the Web. This document provides an overview of this family of documents. PROV is a set of recommended standards of the World Wide Web Consortium. These include its data model, an XML schema for that model, an OWL2 ontology mapping that model to RDF, and a mapping from that ontology to Dublin Core.
Maintenance	W3C, maintained by the Provenance Working Group
Documentation	https://www.w3.org/TR/prov-overview/
Citation	
Further reading	https://github.com/pgroth/PROVTutorial https://www.w3.org/2011/prov/wiki/Main_Page

Name	Web Annotation Data Model
Focus	The Web Annotation Data Model specification describes a structured model and format to enable annotations to be shared and reused across different hardware and software platforms. Implemented by AnnoSys (https://annosys.bgbm.fu-berlin.de/).
Maintenance	W3C Recommendation, The specification is an outcome from the Open Annotation Community Group and the document was published by the Web Annotation Working Group as a Recommendation, maintained by W3C.
Documentation	https://www.w3.org/TR/annotation-model/
Citation	
Further reading	https://github.com/w3c/web-annotation

Ontologies

Known relevant ontologies in a tabular format incl. information on Name, Focus, and Further reading. The ontologies are in alphabetical order. Although the list is already very comprehensive, further ontologies might be of interest or relevance during ongoing development of the DiSSCo Research Infrastructure.

Name	PROV-O
Focus	The PROV Ontology (PROV-O) expresses the PROV Data Model [PROV-DM] using the OWL2 Web Ontology Language (OWL2) [OWL2-OVERVIEW]. It provides a set of classes, properties, and restrictions that can be used to represent and interchange provenance information generated in different systems and under different contexts. It can also be specialized to create new classes and properties to model provenance information for different applications and domains.
Further reading	https://www.w3.org/TR/prov-o/

Name	Biological Collections Ontology
Focus	The Biological Collections Ontology (BCO) is being developed as an application ontology as part of the Biocode Commons project, within the OBO Foundry framework. The goal of the BCO is to support the interoperability of biodiversity data, including data on museum collections, environmental/metagenomic samples, and ecological surveys.
Further reading	http://www.obofoundry.org/ontology/bco.html

Name	Ontology for Biomedical Investigations
Focus	The Ontology for Biomedical Investigations (OBI) is an integrated ontology for the description of biological and clinical investigations. It helps to communicate clearly about scientific investigations by defining more than 2500 terms for assays, devices, objectives, and more.
Further reading	Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, et al. (2016) The Ontology for Biomedical Investigations. PLOS ONE 11(4): e0154556, https://doi.org/10.1371/journal.pone.0154556 , http://obi-ontology.org/ , https://github.com/obi-ontology/obi/wiki

Name	Information Artifact Ontology
Focus	The Information Artifact Ontology (IAO) is a new ontology of information entities, originally driven by work by the OBI digital entity and realizable information entity branch. It is an ontology based on a theory of document acts describing what people can do with documents.
Further reading	http://www.obofoundry.org/ontology/iao.html https://github.com/information-artifact-ontology/IAO

Name	Open Digital Specimen Ontology
Focus	The Ontology for open Digital Specimens (ODS) situates open Digital Specimens in the relevant OBO Foundry ontologies and extends from those roots to define the new concepts needed to support mass digitization and Digital Specimens on the Internet.
Further reading	https://github.com/DiSSCo/openDS/blob/master/ods-ontology/ods-ont-intro.md

Name	MarineTLO ontology
Focus	MarineTLO is a top-level ontology for the marine domain (also applicable to the terrestrial domain) to tackle the need for having integrated sets of facts about marine species, and thus to assist research about species and biodiversity. It provides a unified and coherent core model for schema mapping which enables formulating and answering queries which cannot be answered by any individual source
Further reading	Tzitzikas Y. et al. (2013) Integrating Heterogeneous and Distributed Information about Marine Species through a Top Level Ontology. In: Garoufallou E., Greenberg J. (eds) Metadata and Semantics Research. MTSR 2013. Communications in Computer and Information Science, vol 390. Springer, Cham. https://doi.org/10.1007/978-3-319-03437-9_29 http://www.ics.forth.gr/isl/MarineTLO/

04 OUTLOOK

MS Report 5.3 on “Documentation of PIDs relevant for DiSSCo technical infrastructure” due in Month 18 will further extend the list of relevant building blocks for the DiSSCo RI. Concepts to assess their technical readiness and those of other important services will be part of MS 5.2 (Implementation of concepts for sustainability of services, CMS, and overall TRL) likewise due in Month 18. The first version of the DiSSCo Knowledgebase is already running and will be demonstrated during the first DPP All Hands Meeting (Session on 19th January 2021). Feedback on the current version and further requirements will be collected. The relevant data standards, ontologies and PIDs will be available via the DiSSCo Knowledgebase, time of publication will however depend on the prioritization.