University of Tasmania
GPO Box 367
Hobart 7001
Australia

imos.org.au

April 2026

IMOS NESP 5.9

Kelp Squidle+ Annotations Data Product

Product Technical Document

Version 1.1

Thomas Galindo
Australian Ocean Data Network AODN / IMOS

Australia's Integrated Marine Observing System (IMOS) is enabled by the National Collaborative Research Infrastructure Strategy (NCRIS). It is operated by a consortium of institutions as an unincorporated joint venture, with the University of Tasmania as Lead Agent.

IMOS acknowledges the Traditional Custodians and Elders of the land and sea on which we work and observe, and recognise them as Australia's first marine scientists and carers of sea Country. We pay our respects to Aboriginal and Torres Strait Islander peoples past and present.

1. Version History

Version	Date	Comments	Author
v1.0	January-2026	Initial Release	Galindo, T
v1.1	April-2026	Minor Changes	Galindo, T

Citation

Galindo, T. (2026). Kelp Squidle+ Annotations Data Product Technical Document. Version 1.1 Integrated Marine Observing System.

Copyright/Creative Commons Licence

Table of Contents

2. Background

The Kelp Squidle+ Annotations Data Product extracts kelp annotations from Squidle and translates them to Darwin Core.

2.1 Squidle+

Squidle+ is a centralised marine image data management, discovery, and annotation platform with vocabulary translation.

Scientists and the general public use Squidle+ to host and annotate photo based marine surveys.

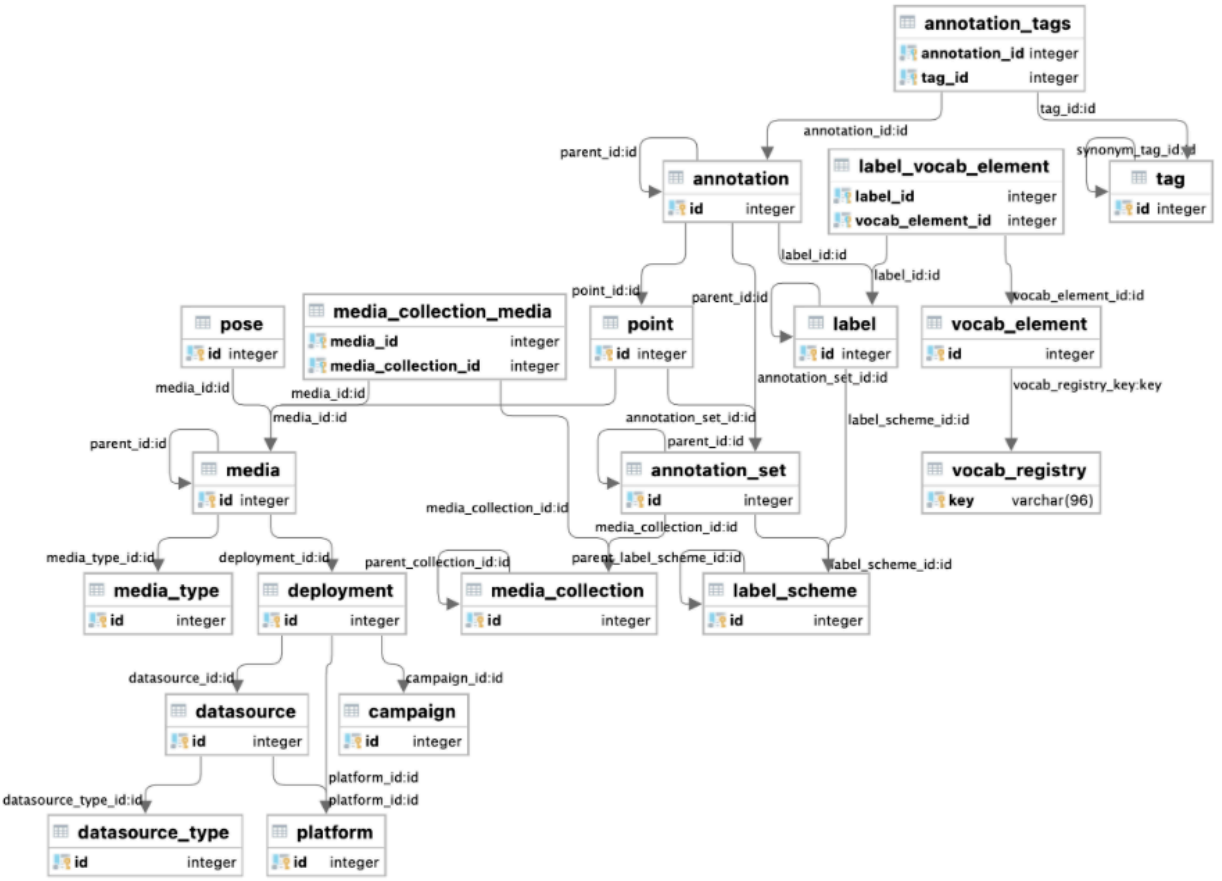
2.2 Darwin Core Mapping

The Kelp Annotations are mapped to the Darwin Core scheme to increase interoperability of Squidle+ data with existing Darwin Core tooling and systems (eg OBIS).

2.3 Basis of the Kelp Squidle Annotations Data Product

Bulk exportation of Squidle+ data requires experience working with REST APIs (see API reference here) as well as a good understanding of relational databases.

The complexity of the underlying Squidle+ relational database further complicates the extraction and aggregation of Squidle+ data into aggregated data products suitable for scientific analysis, as exemplified by the simplified Entity-Relationship Model (ERM):



The goal of the aggregated data product is to flatten the Kelp survey photo annotations into a single aggregated data product that is ready for scientific analysis and interoperable with key marine biological data systems.

3. Methodology

The Kelp Squidle Annotations Data Product (Data Product) relies on the standard Extract, Transform, and Load (ETL) methodology.

3.1 Extract

Data Source

The underlying data are sourced from public Squidle+ annotations and label schemes.

Additionally, the relevant WoRMS vocabulary was extracted from the WoRMS marine species website.

Filtering and Label Schemes

Squidle+ annotations are labeled with user specified controlled vocabularies.

The user controlled aspect of these vocabularies makes identifying kelp by text difficult as the label text evolves over time as per user modifications.

To reduce potential inclusion of non-kelp labels, the following ID list was generated to filter Squidle+ annotations by the Kelp Label Allow List

Extraction

Annotations and related translation metadata are exported from Squidle+ using the Squidle+ API.

This is a three step process:

1. Filter to kelp containing public annotation sets using the Kelp Label Allow List
2. Export all kelp containing annotation sets
3. Export all vocab elements for the kelp labels

Additionally, the Australia geometry shapefile is downloaded from the Australian Bureau of Statistics (ABS):

- URL: `https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/access-and-downloads/digital-boundary-files/AUS_2021_AUST_SHP_GDA2020.zip`
- This geometry is used downstream in the transform stage for land filtering.

The raw annotations, related WoRMS taxonomic controlled vocabulary elements, a mapping between Squidle+ labels and WoRMS labels, and the Australia geometry are passed on to the transform stage.

3.2 Transform

Summary

Phase	Task	Description
Setup	Workspace Prep	Ensures the temporary directory `transform_dir` exists for ephemeral data.
Cleaning	Data Refinement	Filters for kelp-specific records and resolves overlapping or conflicting media metadata.
DWC Translation	Standardisation	Maps the local schema to the global **Darwin Core** standard using WoRMS taxonomy.
Temporal-Spatial	Time Pruning, Geofencing & Land Filtering	Filters out all records occurring before **September 28, 2007**, constrains data to the Australian region, and removes points falling within the Australian mainland.
Geospatial Enrichment	Spatial Indexing	Adds H3 spatial index and Australian Marine Region tags to each record.
Validation	Quality Control	Validates the final Polars DataFrame and exports it as a **PyArrow Table**.

Cleaning

The cleaning phase prunes annotations to kelp and resolves label conflicts:

Kelp Pruning

Filters the dataset to retain only records classified as **Kelp**, removing unrelated biological observations.

Media Conflict Resolution

Squidle+ allows overlapping annotation layers on a single image. The pipeline identifies media with multiple annotation sets and **rejects** any media where labels are contradictory, ensuring high data integrity.

DWC Translation

This step converts raw annotations into a standardized occurrence format through several key operations:

Taxonomic Alignment

The local `label.id` is joined with the **World Register of Marine Species (WoRMS)** vocabulary. This enriches the data with authoritative taxonomic hierarchies (Kingdom through Genus) and provides a globally unique `scientificNameID` (LSID).

Identifier Generation

A unique `occurrenceID` is constructed by concatenating the internal annotation ID with the deployment ID, ensuring traceability back to the original survey.

Temporal Normalization

Diverse time formats are parsed into a unified `eventDate` (GMT timestamp).

Spatial Normalization

Raw pose data (latitude/longitude) is renamed to DwC standard `decimalLatitude` and `decimalLongitude`.

Observation Metadata

Sets the `basisOfRecord` to `MachineObservation` and hardcodes the `occurrenceStatus` to `present`, as the input consists of photographs and positive labels, respectively.

Media Linkage

Constructs a direct URL for `associatedMedia` by formatting the Squidle+ media ID into an API-accessible iframe link for visual verification.

Temporal-Spatial

Filters out all records occurring before **September 28, 2007** and constrains data to the Australian region (lon [67.054, 171.801], lat [-58.449, -8.471]).

Land Filtering

After spatial extent filtering, occurrence records where the point falls **within** the Australian mainland polygon are removed. This ensures only marine observations are retained.

The mechanism:

- The ABS Australia shapefile (downloaded during extract) provides the mainland polygon
- Shapely's `within` predicate is used to identify and exclude points falling on land

---

Validation

Validates the final Polars DataFrame conforms to the expected output schema and confirms null constraints are valid.

Geospatial Enrichment

The output dataset is enriched with the following geospatial columns:

- `h3Index` — A hexadecimal string representing an H3 polygon at resolution 5, derived from `decimalLatitude` and `decimalLongitude`
- `australianMarineRegionsTags` — A `|` separated tag column featuring common Australian marine regions

4. Load

Finally, the transformed Parquet is loaded to the `processing-stored-bucket` at path `datauplift/kelp/kelp.parquet`.

5. Flow Advantages

Reproducibility the original Kelp Squidle+ data requires experience with REST APIs and relational databases to extract and aggregate. The ETL procedure makes this process reliable and replicable.

Space Efficiency raw original data is reduced to **~2MB** with modern compression and file formats (Parquet), optimising storage and access

Time Efficiency scheduled updates allow users to access up-to-date aggregated data without re-implementing the aggregation steps and validation

Cloud-native workflows reduce I/O overhead

6. Appendix

6.1 Kelp Label Allow List

```
[
    {
        "id": 7327,
        "catalogue": "Australian_Morphospecies_Catalogue",
        "name": "Phyllospora comosa"
    },
    {
        "id": 29052,
        "catalogue": "Australian_Morphospecies_Catalogue",
        "name": "Seirococcus axillaris"
    },
    {
        "id": 14983,
        "catalogue": "Australian_Morphospecies_Catalogue",
        "name": "Cystophora spp"
    },
    {
        "id": 29051,
        "catalogue": "Australian_Morphospecies_Catalogue",
        "name": "Perithalia caudata"
    },
    {
        "id": 12678,
        "catalogue": "Australian_Morphospecies_Catalogue",
        "name": "Durvillaea potatorum"
    },
    {
        "id": 7325,
        "catalogue": "Australian_Morphospecies_Catalogue",
        "name": "Brown Other Canopy Forming spp"
    },
    {
        "id": 7326,
        "catalogue": "Australian_Morphospecies_Catalogue",
        "name": "Ecklonia radiata"
    },
    {
        "id": 28751,
        "catalogue": "Australian_Morphospecies_Catalogue",
        "name": "Scaberia agardhii"
    },
    {
        "id": 583,
        "catalogue": "Australian_Morphospecies_Catalogue",
        "name": "Brown"
    },
    {
        "id": 13379,
        "catalogue": "Australian_Morphospecies_Catalogue",
        "name": "Carpoglossum confluens"
    },
    {
        "id": 28462,
        "catalogue": "Australian_Morphospecies_Catalogue",
        "name": "Xiphophora gladiata"
    },
```

```
    {
        "id": 13407,
        "catalogue": "Australian_Morphospecies_Catalogue",
        "name": "Scytothalia dorycarpa"
    },
    {
        "id": 12680,
        "catalogue": "Australian_Morphospecies_Catalogue",
        "name": "Lessonia corrugata"
    },
    {
        "id": 28465,
        "catalogue": "Australian_Morphospecies_Catalogue",
        "name": "Macrocystis pyrifera"
    },
    {
        "id": 28883,
        "catalogue": "Australian_Morphospecies_Catalogue",
        "name": "Carpophyllum spp"
    },
    {
        "id": 29037,
        "catalogue": "Australian_Morphospecies_Catalogue",
        "name": "Platythalia angustifolia"
    },
    {
        "id": 29038,
        "catalogue": "Australian_Morphospecies_Catalogue",
        "name": "Platythalia quercifolia"
    },
    {
        "id": 29036,
        "catalogue": "Australian_Morphospecies_Catalogue",
        "name": "Caulocystis spp"
    },
    {
        "id": 28945,
        "catalogue":
"Benthic_habitat_and_relief_schema_for_BRUV__BOSS_and_DOV_Imagery",
        "name": "Scytothalia dorycarpa"
    },
    {
        "id": 28943,
        "catalogue":
"Benthic_habitat_and_relief_schema_for_BRUV__BOSS_and_DOV_Imagery",
        "name": "Brown"
    },
    {
        "id": 28946,
        "catalogue":
"Benthic_habitat_and_relief_schema_for_BRUV__BOSS_and_DOV_Imagery",
        "name": "Phyllospora comosa"
    },
    {
        "id": 28944,
        "catalogue":
"Benthic_habitat_and_relief_schema_for_BRUV__BOSS_and_DOV_Imagery",
```

```
        "name": "Ecklonia radiata"
    },
    {
        "id": 583,
        "catalogue": "Catami_1_4",
        "name": "Brown"
    },
    {
        "id": 10905,
        "catalogue": "Hunter_-_Extended_RLS_Scheme",
        "name": "Desmarestia and Himantothallus"
    },
    {
        "id": 10901,
        "catalogue": "Hunter_-_Extended_RLS_Scheme",
        "name": "Durvillaea"
    },
    {
        "id": 10906,
        "catalogue": "Hunter_-_Extended_RLS_Scheme",
        "name": "Large brown laminarian kelps"
    },
    {
        "id": 28227,
        "catalogue": "Hunter_-_Extended_RLS_Scheme",
        "name": "Lessonia corrugata"
    },
    {
        "id": 10903,
        "catalogue": "Hunter_-_Extended_RLS_Scheme",
        "name": "Phyllospora"
    },
    {
        "id": 14948,
        "catalogue": "Hunter_-_Extended_RLS_Scheme",
        "name": "Sargassum"
    },
    {
        "id": 13553,
        "catalogue": "Hunter_-_Extended_RLS_Scheme",
        "name": "Carpoglossum"
    },
    {
        "id": 10904,
        "catalogue": "Hunter_-_Extended_RLS_Scheme",
        "name": "Macrocystis"
    },
    {
        "id": 14950,
        "catalogue": "Hunter_-_Extended_RLS_Scheme",
        "name": "Xiphophora"
    },
    {
        "id": 14947,
        "catalogue": "Hunter_-_Extended_RLS_Scheme",
        "name": "Cystophora"
    },
```

```
    {
        "id": 28201,
        "catalogue": "Hunter_-_Extended_RLS_Scheme",
        "name": "Perithalia caudata"
    },
    {
        "id": 10902,
        "catalogue": "Hunter_-_Extended_RLS_Scheme",
        "name": "Ecklonia radiata"
    },
    {
        "id": 10904,
        "catalogue": "RLS_Australian_Coral_Species_List",
        "name": "Macrocystis"
    },
    {
        "id": 10903,
        "catalogue": "RLS_Australian_Coral_Species_List",
        "name": "Phyllospora"
    },
    {
        "id": 10905,
        "catalogue": "RLS_Australian_Coral_Species_List",
        "name": "Desmarestia and Himantothallus"
    },
    {
        "id": 10906,
        "catalogue": "RLS_Australian_Coral_Species_List",
        "name": "Large brown laminarian kelps"
    },
    {
        "id": 10902,
        "catalogue": "RLS_Australian_Coral_Species_List",
        "name": "Ecklonia radiata"
    },
    {
        "id": 10901,
        "catalogue": "RLS_Australian_Coral_Species_List",
        "name": "Durvillaea"
    },
    {
        "id": 10904,
        "catalogue": "RLS_Catalogue",
        "name": "Macrocystis"
    },
    {
        "id": 10903,
        "catalogue": "RLS_Catalogue",
        "name": "Phyllospora"
    },
    {
        "id": 10901,
        "catalogue": "RLS_Catalogue",
        "name": "Durvillaea"
    },
    {
        "id": 10906,
```

```
        "catalogue": "RLS_Catalogue",
        "name": "Large brown laminarian kelps"
    },
    {
        "id": 10902,
        "catalogue": "RLS_Catalogue",
        "name": "Ecklonia radiata"
    },
    {
        "id": 10905,
        "catalogue": "RLS_Catalogue",
        "name": "Desmarestia and Himantothallus"
    },
    {
        "id": 285,
        "catalogue": "SQUIDLE_1_0",
        "name": "Ecklonia radiata"
    },
    {
        "id": 583,
        "catalogue": "SQUIDLE_1_0",
        "name": "Brown"
    }
]
```