DASSH
The archive for marine species and habitats data

International Standards for National Data Management Excellence

Anna Luff, Data Officer

THE MARINE BIOLOGICAL ASSOCIATION

MEDIN
marine environmental data & information network

DASSH
The archive for marine species and habitats data

Who are we?

- ✓ UK Data Archive Centre
- ✓ Core funded by the UK and Scottish Government
- ✓ Hub for accessing, safeguarding, managing and disseminating data
- ✓ Support for monitoring and reporting frameworks
- ✓ Committed to applying the FAIR Data Principles

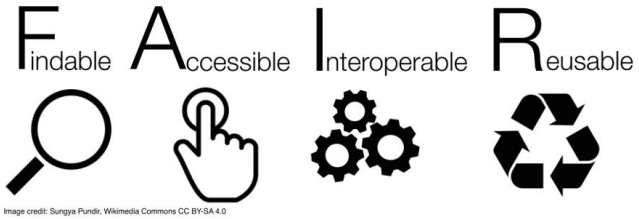
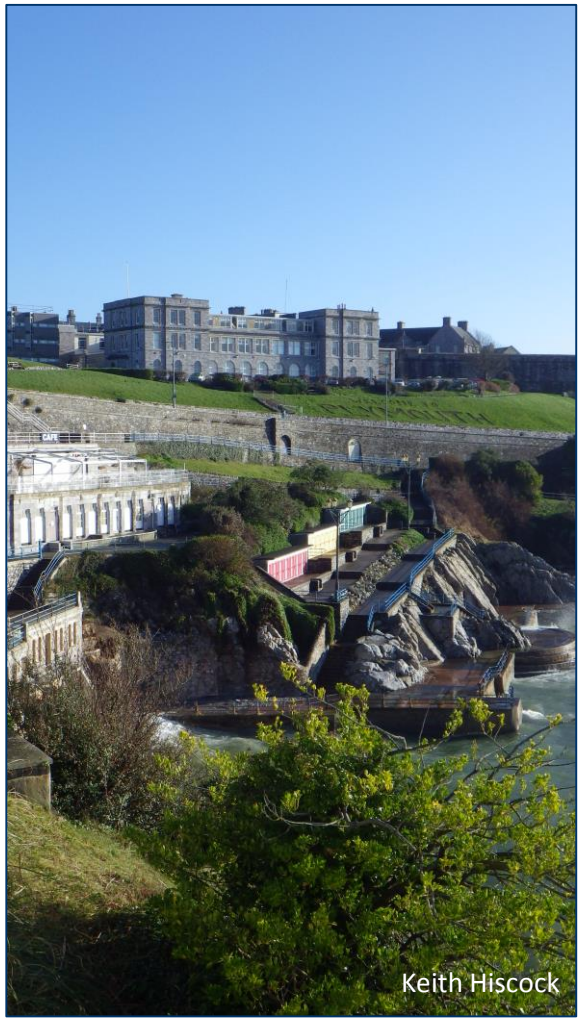


Image credit: Sungya Pundir, Wikimedia Commons CC BY-SA 4.0


Keith Hiscock

THE MARINE BIOLOGICAL ASSOCIATION

MEDIN
marine environmental data & information network

DASSH
The archive for marine species and habitats data

The Role of DASSH

National role

- Accredited MEDIN Data Archive Centre since 2005
- Marine Node of the National Biodiversity Network
- Data provider to the Atlas of Living Scotland

EU role

- EMODnet Biology partner

International role

- UK Node of OBIS
- Accredited Data Unit of IODE

NBN National Biodiversity Network

NBN atlas SCOTLAND

OBIS OCEAN BIOGEOGRAPHIC INFORMATION SYSTEM

EMODnet European Marine Observation and Data Network

IODE International Oceanographic Data and Information Exchange

THE MARINE BIOLOGICAL ASSOCIATION

MEDIN marine environmental data & information network

DASSH
The archive for marine species and habitats data

QUESTIONS

1) How does the work at DASSH contribute to the development of a global marine biodiversity observation network?

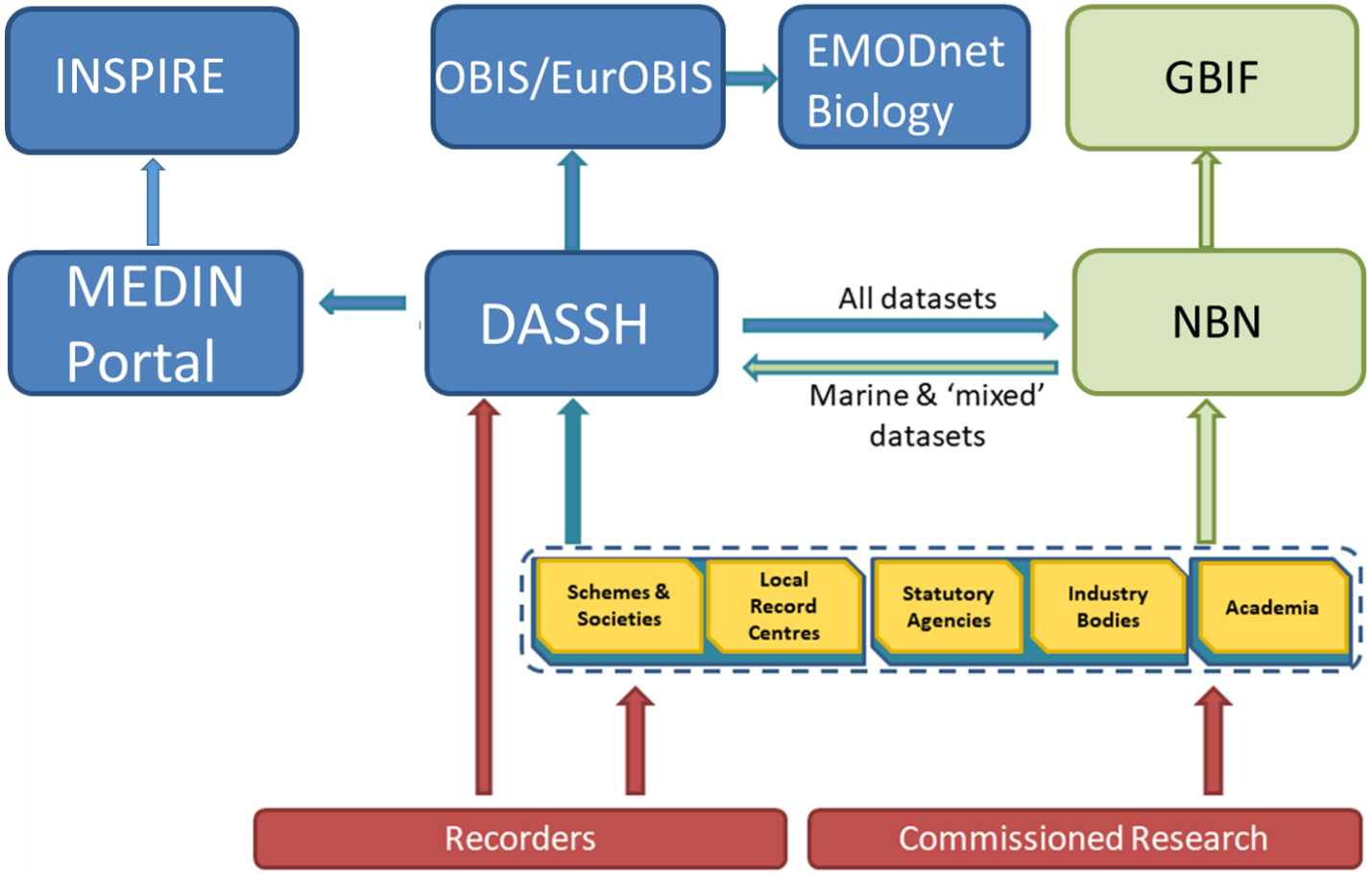
2) How do automated QA and Ingestion tools streamline data at a National and Global scale?

THE MARINE BIOLOGICAL ASSOCIATION

MEDIN
marine environmental data & information network

DASSH
The archive for marine species and habitats data

Data Flow

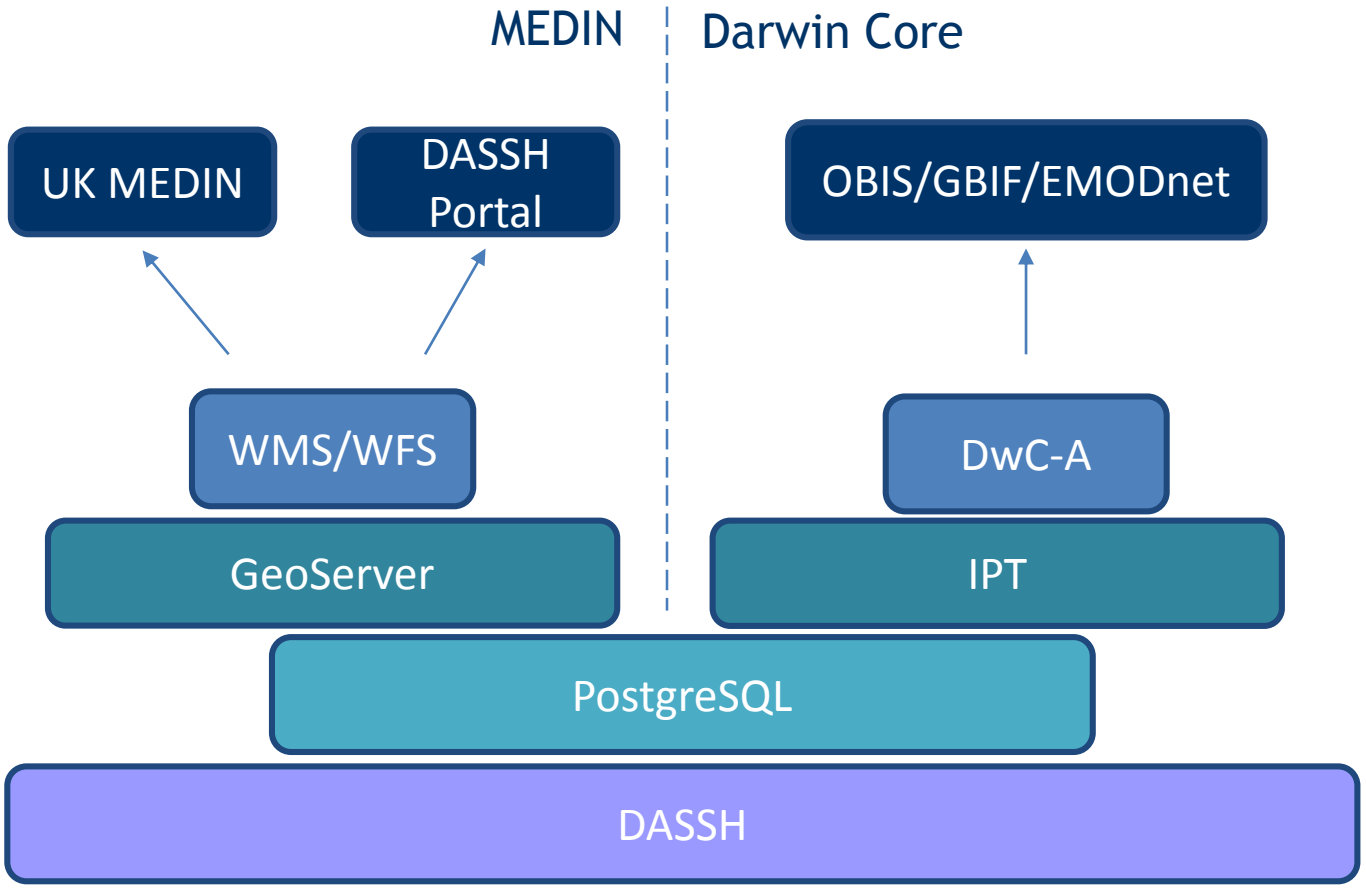


THE MARINE BIOLOGICAL ASSOCIATION

MEDIN
marine environmental data & information network

DASSH
The archive for marine species and habitats data

Data Dissemination

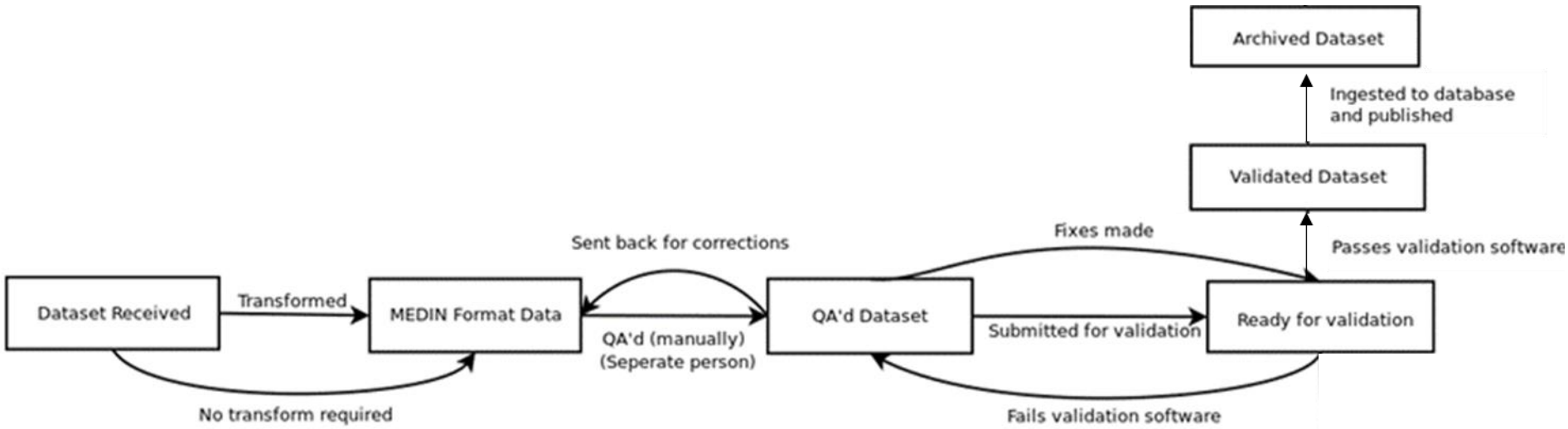


THE MARINE BIOLOGICAL ASSOCIATION

MEDIN
marine environmental data & information network

DASSH
The archive for marine species and habitats data

Data Life Cycle



THE MARINE BIOLOGICAL ASSOCIATION
MEDIN marine environmental data & information network
PostgreSQL

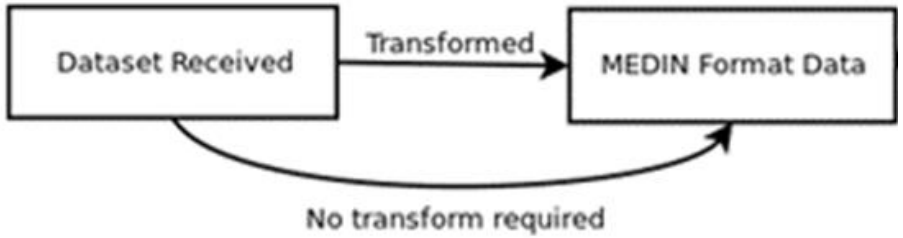
DASSH
The archive for marine species and habitats data

Data Standardisation

- Data are manually standardised to increase interoperability and re-use of marine data

- Use of MEDIN (Marine Environmental Data and Information Network) guidelines to standardise and input data

MEDIN
marine environmental data & information network



THE MARINE BIOLOGICAL ASSOCIATION

MEDIN
marine environmental data & information network

DASSH

The archive for marine species and habitats data

MEDIN Guidelines

General Metadata Form

	Field Title	Your Data and Information	Add additional columns as required
Project Metadata (Link if applicable)	projectName		
	projectCode		
	projectStartDate		
	projectEndDate		
	projectMetadataURL		
Survey Metadata	surveyName	2018 Plymouth Sound Grab Survey	
	surveyType	Benthic Grab Survey	
	surveyAbstract	A grab survey was conducted by the Marine Biological Association of the United Kingdom (MBA), in Plymouth Sound to survey benthos.	
	surveyCode	DASHDT000000456	
	originator	Marine Biological Association	
	owner	Marine Biological Association	
	surveyStartDate	01/01/2018 00:00	
	timeZone	UTC	
	spatialCRS	WGS84: EPSG:: 4326	
	positionFix	DGPS	
	horizontalAccuracy	10	
	surveyMetadataURL		
	surveyEndDate		
	originalCRS		
	transformation		
	depthCRS		
	verticalAccuracy		
	platformType		
	platformName		
	marineRecorderSurveyId		
	emodnetSeabedHabitatsGui		
	cruiseReportReference		
	surveyReportReference		
	confidentiality	public	

PRIMARY KEY

Sample Event Form

Field Title	sampleEventID	surveyCode	methodID	stationID	replicateID	sampleDate	dateRange	sampleLatitude	sampleLongitude	originalSampleLatitud	originalSampleLongitud	sampleTime	waypointNumber	eventName
Your Data	DASSHDT00000456_SE_001	DASSHDT00000456	METHOD_01	ST_01		2018-01-01		50.342612	-4.150171			13:00:00		
Add additional rows as required														

Species Form

	Field Title	sampleEventID	replicateID	taxonID	taxonName	aphiaID	originalName	qualifier	abundance	abundanceUnits	determin
Species Data	Your Data	DASSHDT00000456_SE_001		DASSHDT00000456_SE_001_103741	*Corella borealis*	103741			1	count	
	Add additional rows as required	DASSHDT00000456_SE_001		DASSHDT00000456_SE_001_126440	*Pollachius pollachius*	126440			P	presence/absence	

PostgreSQL

THE MARINE BIOLOGICAL ASSOCIATION

MEDIN marine environmental data & information network

DASSH
The archive for marine species and habitats data

Data Standardisation

MEDIN marine environmental data & information network

Discovery Metadata Editor

Home | Create | Update | Export | Import | My account | My records | Delete Records | Help | Contact | Logout

Dashboard

Create

Enter the title

1992 Centre for Environment, Fisheries and Aquaculture Science (Cefas) North Sea :

What type of resource do you want to describe

dataset

I am happy for my metadata to be passed to data.gov.uk

Create

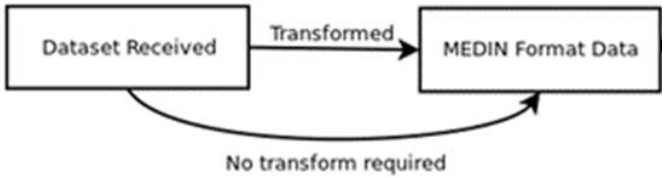
Update

Choose a record to update

Edit this record

Export

Search for a record



THE MARINE BIOLOGICAL ASSOCIATION

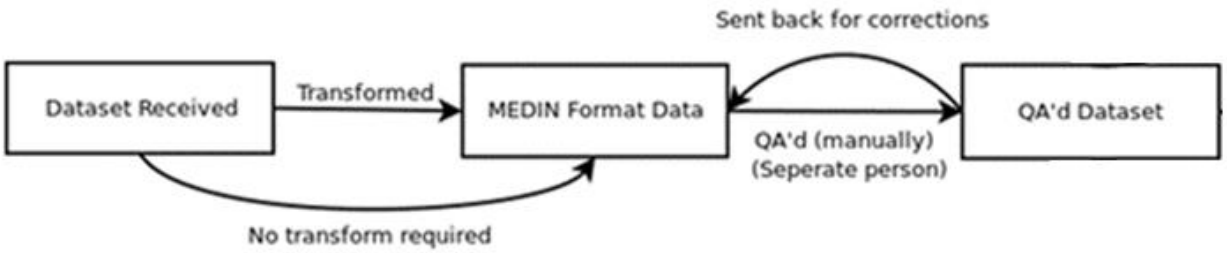
MEDIN marine environmental data & information network

DASSH
The archive for marine species and habitats data

Quality Assurance

Manual Quality Assurance check of data and metadata

- ✓ Abstract, title, start/end dates
- ✓ Do we have the necessary Detailed Metadata? Is it correct?
- ✓ Check of species data
- ✓ QA of Discovery Metadata on MEDIN- are all of the mandatory fields completed? Does the metadata match the raw data?



THE MARINE BIOLOGICAL ASSOCIATION

MEDIN
marine environmental data & information network

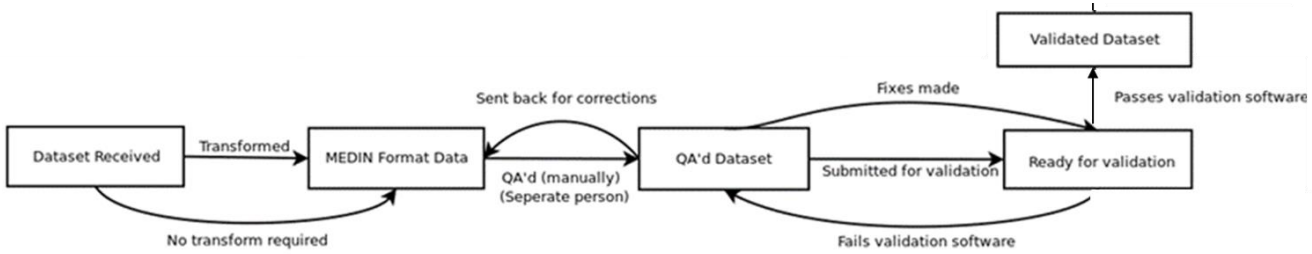
DASSH
The archive for marine species and habitats data

Validation

Validation Tool used to check data in MEDIN guidelines

Benefits?

- ✓ Speeds up data processing
- ✓ Increased accuracy
- ✓ Development of validation tool as a web-service
- ✓ Potential to be embedded in other applications



THE MARINE BIOLOGICAL ASSOCIATION

MEDIN
marine environmental data & information networ

DASSH
The archive for marine species and habitats data

Ingestion and Archival

Ingestion Tool

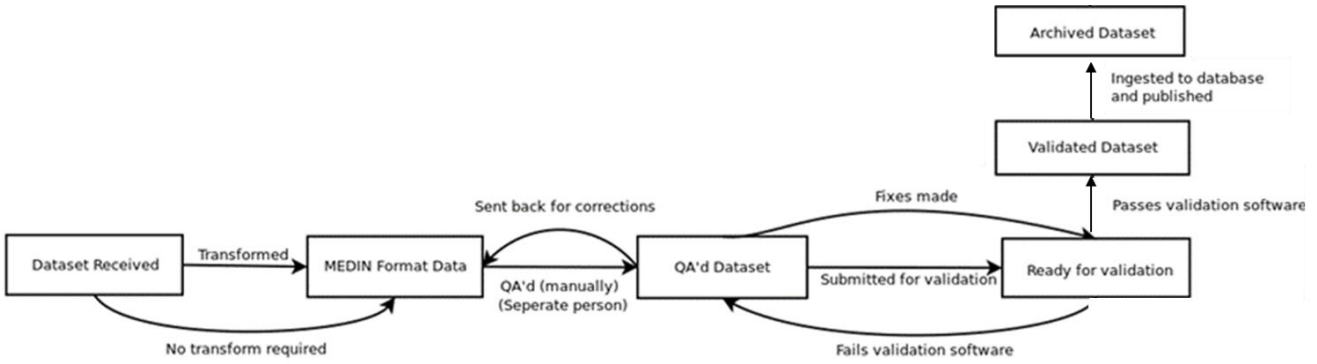
- Takes each MEDIN guideline sheet in turn, and ingests into the database.
- Extracts tables, reads and unpivots matrices
- Data quickly made available via GeoServer

```
#!/usr/local/bin/python3
# This is a python3 script. You will need to install python3 to run it.

# This is a script to put MEDIN spreadsheets into the database

import sys
import os
import argparse
import datetime
import traceback
import ingest_excel  # this reads the excel files into python
import parse_medin_data   # this extracts MEDIN data
import dbossh_library as db_lib   # this interacts with the postgres db
from parse_medin_data import customException
from medin_spreadsheets_shared import getWorksheetInfo

## Parse command line arguments (and set defaults)
parser = argparse.ArgumentParser(description='Script to put MEDIN spreadsheets into the database')
parser.add_argument("-s",
                    "--spreadsheet",
                    required=True,
                    dest='filePath',
                    help="The spreadsheet to be read into DBOSSH")
parser.add_argument('-l',
                    '--logfile',
                    dest='logFile',
                    help="Specify a log file to output to")
parser.add_argument("-d",
                    "--database",
                    dest='dbName',
                    help="The target database name if want to override DBOSSH default")
parser.add_argument('-t',
                    "--tableList",
                    dest='tableListLocation',
                    help="The file containing the list of tables in DBOSHH, "
                        "their column names, their datatypes and their primary key",
                    default="data/table_list.txt")
parser.add_argument("--update",
                    action="store_true",
                    help="Instruct the script to overwrite already existing"
                        "entries, rather than skipping")
parser.add_argument("-m",
                    "--metadataID",
                    dest='metaID',
                    help="The metadata identifier (used to link with the "
                        "metadata file held by MEDIN). Either provide a single"
                        "ID which will be applied to all lines or a string"
                        "matching IDs to survey codes in the form "
                        "surveyCode:metadataID,surveyCode:metadataID,...",
                    default="0f0f0f0f0f0f0f0f0f0f0f0f0f0f0f0f")
parser.add_argument("-D",
                    "--date",
                    dest='dateAdded',
                    help="The date uploaded to DBOSSH in format YYYY-MM-DD (default is today's date)")
parser.add_argument("-S",
                    "--sheets",
                    dest='sheets',
                    nargs='+',
                    help="Only ingest the specified sheet(s) (all others will be omitted). "
```



THE MARINE BIOLOGICAL ASSOCIATION

MEDIN marine environmental data & information network

DASSH
The archive for marine species and habitats data

Conclusion

1) How does the work at DASSH contribute to the development of a global marine biodiversity observation network?

- Data flow feeds into European and Global marine biological data infrastructure
- Contribute to data products on a global scale
- Influential in development of UK standards in accordance to INSPIRE specifications

THE MARINE BIOLOGICAL ASSOCIATION

MEDIN
marine environmental data & information network

DASSH
The archive for marine species and habitats data

Conclusion

2) How do automated QA and Ingestion tools streamline data at a National and Global scale?

- Data published faster- freeing up staff time
- Increased process transparency, and user confidence
- Tools can be used by other organisations
- Accept data in wide range of formats

THE MARINE BIOLOGICAL ASSOCIATION

MEDIN
marine environmental data & information network

DASSH
The archive for marine species and habitats data

Thank you for listening, do you have any questions?

Anna Luff, Data Officer annluf@mba.ac.uk
DASSH enquiries dassh.enquiries@mba.ac.uk

THE MARINE BIOLOGICAL ASSOCIATION

MEDIN
marine environmental data & information network