MEDIN Data Flow Case Studies

Introduction
The purpose of these case studies is to provide practical worked examples of key stages of data flow within the MEDIN Framework. In particular we identify how the following processes are carried out:

- Metadata
  - Generation (formats used)
  - Creation of Unique Resource Identifier for the Data Set
  - Validation
  - Publication

- Data
  - Generation (formats used)
  - Quality Control
  - Archival
  - Management of updates / corrections

Case Study 1. Habitat Survey by a Countryside Agency

A Countryside Agency (CA) commissions some habitat survey work to a consultancy. The Countryside Agency (e.g. Northern Ireland Environment Agency, The Countryside Council for Wales, Natural England, and Scottish Natural Heritage) will keep its own copies of the metadata for the data set and the raw data.

The data set consists of GIS files, raw species counts, photographs and video. From the outset the consultancy is made aware that the data will be submitted to a MEDIN DAC and therefore any work carried out in the contract or outputs from the contract should follow MEDIN standards.

At the conclusion of the project the contractors complete a MEDIN metadata discovery record for the data set (assuming it is considered as a single data set) using the online form on the MEDIN website (note 1) and allocated a Unique Resource Identifier. This record together with the data set is provided to the Countryside Agency.

The Countryside Agency has the responsibility to ensure this data set meets their requirements and that the data have been collected to MEDIN standards. Once the CA is satisfied the consultancy has fulfilled their objectives, the data and metadata are input into the CA systems.

The data set and metadata are then transferred to the appropriate MEDIN DAC, which performs quality assurance on the metadata and then publishes the metadata record on their server. If there are any errors found during QA then this is reported back to the CA and corrected accordingly. The data are archived at the DAC and made available according to an agreement with the CA. If updates or corrections are required then the CA should inform the DAC and, if required, resubmit the corrected data set.

Note a: If the CA has its own format for discovery metadata then the discovery metadata should be entered into their own system. A metadata conversion routine will then be run over the metadata to create MEDIN format metadata records. It will then be necessary to run the MEDIN *schematron* to check the content of the metadata.

Note b: This type of data flow is likely to work for a variety of data types from a CA to a DAC.

Case Study 1. Habitat Survey by a Countryside Agency

→	Flow of Data
→	Flow of Metadata
→	User

Data from consultancy

Metadata created using online tool, URI assigned

Updates

Countryside Agency

Copies of data and metadata stored at Countryside agency

Errors in data/metadata reported back to Countryside Agency

MEDIN DAC

Error(s) found

QA

QA

Metadata uploaded to server and available through portal

Data archived at DAC

Accesses data according to availability agreed by Countryside Agency

USER Searches for appropriate data through the metadata portal

Case Study 2a. An Environmental Characterisation Survey is commissioned by a Policy Agency / Government Department.

The Aggregate Levy Sustainability Fund (ALSF) commissions surveys for Regional Environmental Characterisation (REC) surveys in the Thames Estuary. On completion of the contracts, the raw and processed survey data, and the interpreted/derived data (e.g. GIS layers resulting from integrated analysis of the raw geophysical, biological and archaeological data) are submitted to the ALSF's contracted data management organisation (currently the Geodata Institute) who catalogue and provide a temporary archive for the data. The final arrangements for long term storage and management of the data within the MEDIN DAC network is currently under discussion. The ALSF has implemented the MEDIN data clause in both the survey /interpretation contract and the data management contract to ensure that all data are submitted to MEDIN DACs on conclusion of the project. Agreements are made in advance, with and between the appropriate MEDIN DACs, to decide which DAC will hold which data sets in the long term.
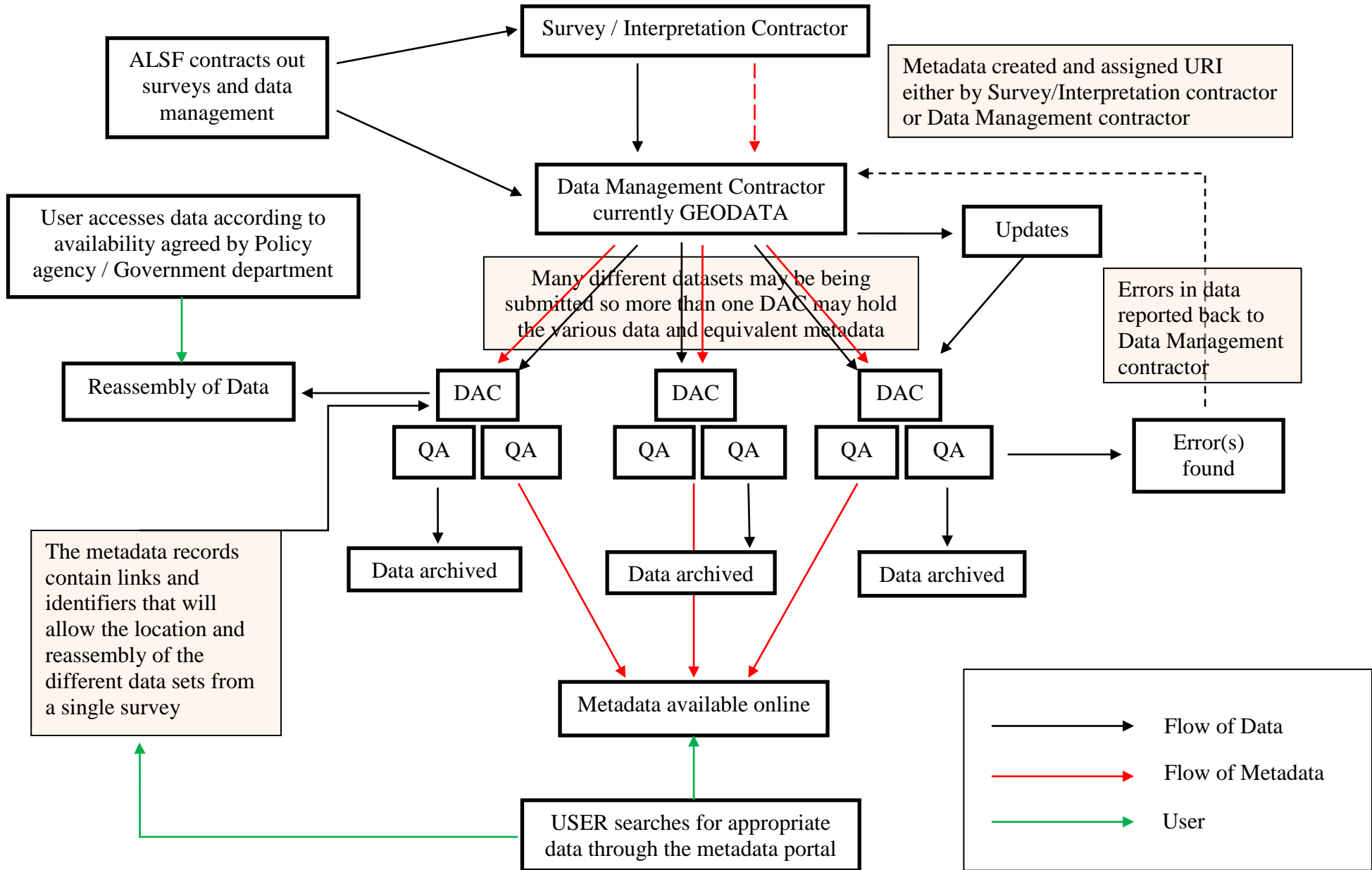
More than one data set will be produced, each of which may comprise many different types of data (grab samples, oceanographic measurements, species counts, videos, multi-beam surveys). The contractor will collect the data and prepare them according to standards agreed with the ALSF that are INSPIRE and MEDIN compliant. The MEDIN compliant metadata record is submitted by either the survey/interpretation contractor or the data management contractor (to be decided after further discussions between ALSF and MEDIN) using the MEDIN online tool (which automatically checks the content and format), and creates Unique Resource Identifiers for each data set.

Once the ALSF is satisfied the survey and data management contractors have fulfilled their objectives with regard to a particular commissioned survey the data sets and metadata are then transferred to the MEDIN DACs. The MEDIN DACs perform quality assurance on the data and metadata and then publish the metadata records on their servers. If there are any errors found during QA then this is reported back to the data management organisation and corrected accordingly. The data are archived at the DAC and made available according to an agreement with the policy agency.

The metadata records contain links and identifiers that will allow the location and reassembly of the different data sets (which may be archived in different DACS) from a single survey.

If subsequent analyses of the data identify the need for updates the agency responsible for managing the programme should be contacted with recommendations to support updates. If agreed the DAC now holding the data will then be contacted and the data will then be resubmitted.

Case Study 2a. An Environmental Characterisation Survey is commissioned by a Policy Agency / Government Department.

Survey / Interpretation Contractor

ALSF contracts out surveys and data management

Metadata created and assigned URI either by Survey/Interpretation contractor or Data Management contractor

Data Management Contractor currently GEODATA

Updates

User accesses data according to availability agreed by Policy agency / Government department

Many different datasets may be being submitted so more than one DAC may hold the various data and equivalent metadata

Errors in data reported back to Data Management contractor

Reassembly of Data

DAC

DAC

DAC

QA QA

QA QA

QA QA

Error(s) found

The metadata records contain links and identifiers that will allow the location and reassembly of the different data sets from a single survey

Data archived

Data archived

Data archived

Metadata available online

USER searches for appropriate data through the metadata portal

Flow of Data

Flow of Metadata

User

Case Study 2b. An Environmental Characterisation Survey is commissioned by a Policy Agency / Government Department.

A policy agency (e.g. DEFRA or the Scottish Government) commissions surveys to organisations such as ABPmer, Fugro, Royal Haskoning, WS Atkins. The policy agency does not wish to hold the actual data themselves but have implemented the MEDIN data clause to ensure that all data are submitted to MEDIN DACs on conclusion of the project. Agreements are made in advance with and between the appropriate MEDIN DACs to decide which DAC will hold which data sets in the long term.
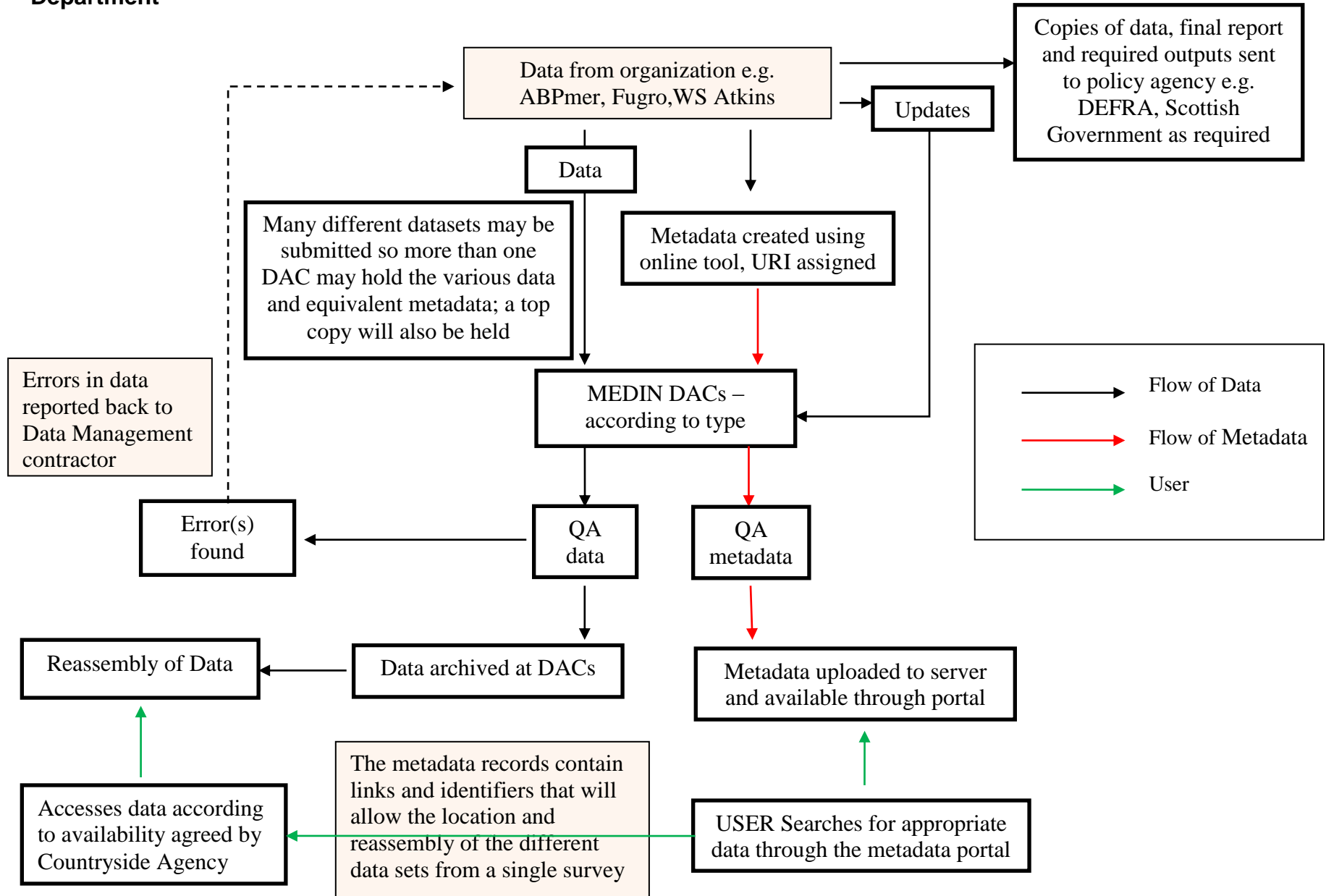
More than one data set will be produced, each of which may comprise many different types of data (grab samples, oceanographic measurements, species counts, videos, multi-beam surveys). The contractor will collect the data and prepare them according to standards agreed with the policy agency that are INSPIRE and MEDIN compliant. The MEDIN compliant metadata record is submitted by either the survey/interpretation contractor or the data management contractor (to be decided after further discussions between policy agencies and MEDIN) using the MEDIN online tool (which automaticaly checks the content and format), and creates Unique Resource Identifiers for each data set.

Once the policy agency is satisfied the survey and data management contractors have fulfilled their objectives with regard to a particular commissioned survey the data sets and metadata are then transferred to the MEDIN DACs. The MEDIN DACs perform quality assurance on the metadata and then publish the metadata records on their servers. If there are any errors found during QA then this is reported back to the commissioned contracted organisation and corrected accordingly. The data are archived with the DACs and made available according to an agreement with the policy agency. It is agreed with the policy agency beforehand whether they require copies of the data, outputs and final reports

The metadata records contain links and identifiers that will allow the location and reassembly of the different data sets (which may be archived in different DACS) from a single survey and a top copy of the survey will also be held.

If subsequent analyses of the data identify the need for updates the agency responsible for managing the programme should be contacted with recommendations to support updates. If agreed the DACs holding the data will then be contacted and the data will then be resubmitted.

Case Study 2b. An Environmental Characterisation Survey is commissioned by a Policy Agency / Government Department

Data from organization e.g. ABPmer, Fugro,WS Atkins

Copies of data, final report and required outputs sent to policy agency e.g. DEFRA, Scottish Government as required

Updates

Data

Metadata created using online tool, URI assigned

Many different datasets may be submitted so more than one DAC may hold the various data and equivalent metadata; a top copy will also be held

Errors in data reported back to Data Management contractor

MEDIN DACs – according to type

Error(s) found

QA data

QA metadata

Reassembly of Data

Data archived at DACs

Metadata uploaded to server and available through portal

Accesses data according to availability agreed by Countryside Agency

The metadata records contain links and identifiers that will allow the location and reassembly of the different data sets from a single survey

USER Searches for appropriate data through the metadata portal

Flow of Data

Flow of Metadata

User

Case Study 3. A Research Cruise by a Fisheries Laboratory (Non Fish Data)

A fisheries laboratory runs a research cruise to study the physical, chemical and biological processes off NW Scotland. The fisheries laboratory will carry out a detailed scientific study, and may hold the collected data itself for potential future internal use. However, it does not wish to be responsible for maintaining a long term archive to allow third parties to access the data – and so will archive the data with DAC(s) within the MEDIN framework.
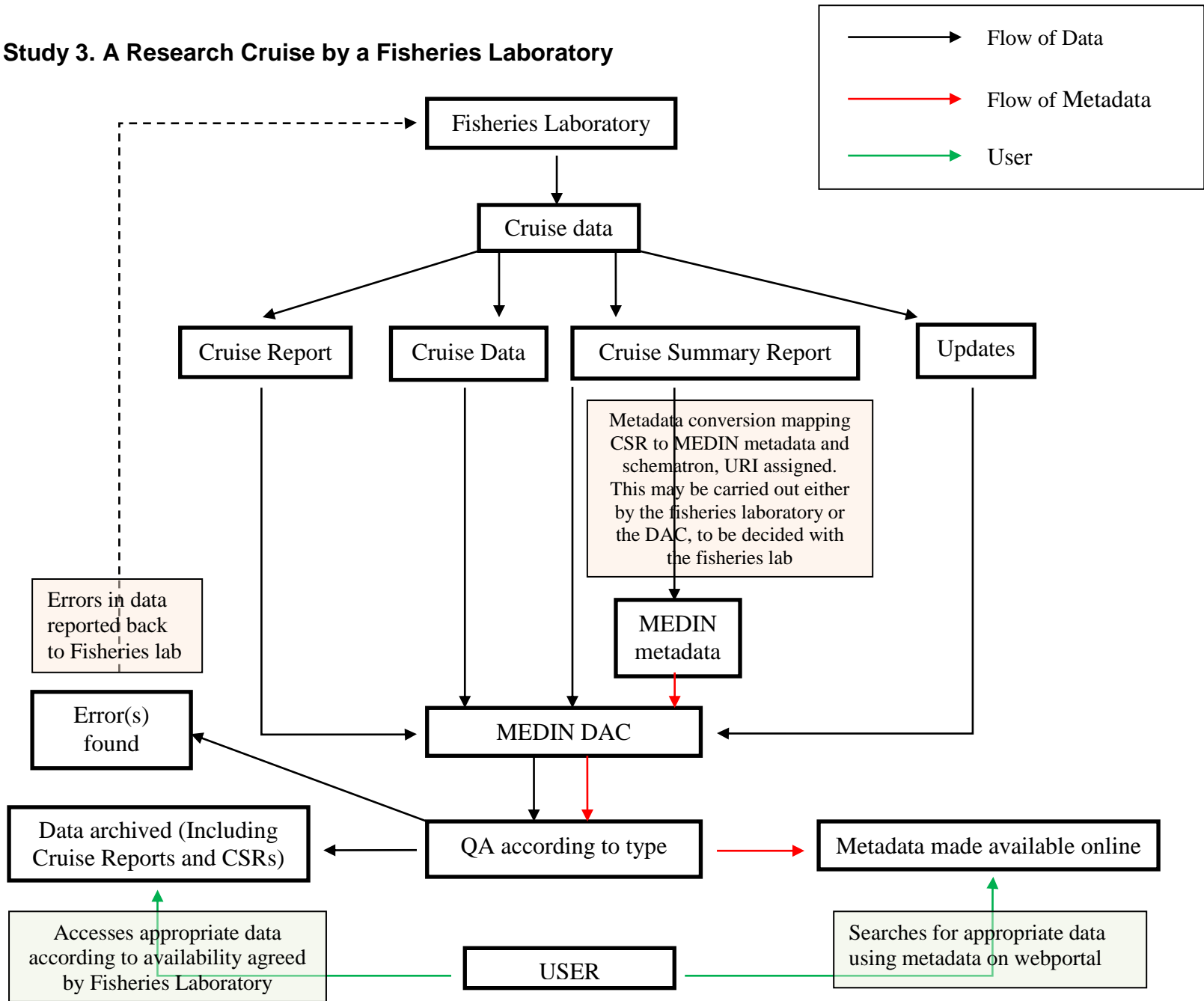
The Fisheries Laboratory already has in place an agreement with a MEDIN DAC(s) that they will supply worked up cruise data, cruise summary reports and cruise reports to the DAC(s) in specified formats. Prior to the cruise the Lab will confirm with the DAC that they are willing to archive the data and advise when these data are expected to be available.

On the cruise the Fisheries Laboratory collects a suite of underway data, water samples, and station data. A summary of cruise particulars and the data collected are recorded in an IOC cruise summary report (CSR) during or post cruise. The CSR is then mapped to the MEDIN metadata schema and an application is run to generate a MEDIN format discovery metadata record from the CSR (To be done either by the Fisheries lab or DAC (to be decided) with mapping provided by MEDIN, some manual input may be required). The application will also verify the format and the contents and will generate a Unique Resource Identifier for the data set. Once data are worked up, a cruise report is produced.

When the Fisheries Laboratory is content that the data have been worked up and checked and are ready for dissemination the data sets and metadata are then transferred to the MEDIN DAC(s). The MEDIN DAC(s), perform quality assurance on the metadata and then publish the metadata records on their servers. If there are any errors found during QA then this is reported back to the Fisheries Lab. and corrected accordingly. The cruise data, cruise summary reports and cruise reports are archived at the DAC and made available according to an agreement with the Fisheries Laboratory.

If updates or corrections are required then the Laboratory should inform the DAC and, if required, resubmit the corrected data set. How this is done is dependent on whether the Fisheries lab or the DAC converts the CSR to MEDIN metadata, ie whether the update is re-run through the mapping application and submitted to the DAC or the applciation is re-run by the DAC. The old record will be replaced with the updated record either manually by the DACs or using a tool to check the URI and replace the old record with the updated record,

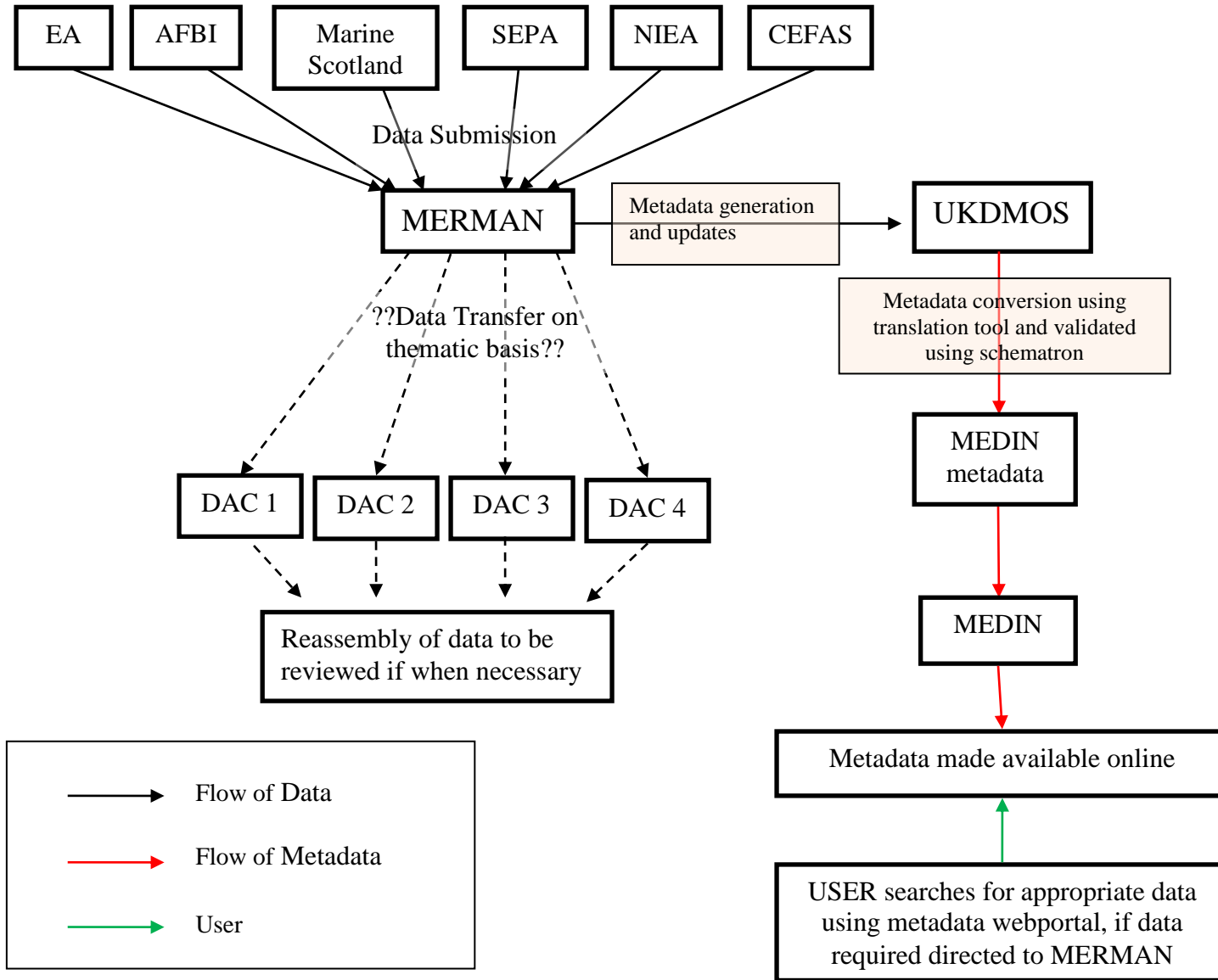# Case Study 3. A Research Cruise by a Fisheries Laboratory

Legend:
- → Flow of Data
- → Flow of Metadata
- → User

Fisheries Laboratory

Cruise data

- Cruise Report
- Cruise Data
- Cruise Summary Report
- Updates

Metadata conversion mapping CSR to MEDIN metadata and schematron, URI assigned. This may be carried out either by the fisheries laboratory or the DAC, to be decided with the fisheries lab

MEDIN metadata

Errors in data reported back to Fisheries lab

Error(s) found

MEDIN DAC

Data archived (Including Cruise Reports and CSRs)

QA according to type

Metadata made available online

Accesses appropriate data according to availability agreed by Fisheries Laboratory

USER

Searches for appropriate data using metadata on webportal

Case Study 4. An agency collecting monitoring data under the Clean and Safe Seas Monitoring Programme submits data to MERMAN

Under the Clean and Safe Seas Monitoring programme, an agency collects water, sediment and biota samples and analyses them for contaminants. The data are quality assured using internal and external programmes (including through QUASIMEME and BEQUALM). The analysed and QC'd data are submitted to the MERMAN database, which then provides an annual submission of Quality Controlled data to ICES.

Subject to ongoing discussions with the data owners, the data held in MERMAN may also be lodged in a DAC and initial thoughts suggest that the benthic invertebrate data would be submitted to DASSH, the fish disease data would be submitted to the Fisheries DAC, and the oceanographic data be submitted to BODC. It has yet to be decided where the contaminant data could be archived. It is anticipated that the archiving process would take place annually and updates performed where necessary.

Using the data in MERMAN, UKDMOS entries for the CSEMP monitoring programmes by each organisation are derived. MEDIN discovery metadata for these programmes are then generated automatically from UKDMOS by MEDIN using a translational tool, creating a URI and validating the metadata entry using a schematron. If data are to be submitted to DACs the process of data flow and reassembly of data will be reviewed.

Case Study 4. An agency collecting monitoring data under the Clean and Safe Seas Monitoring Programme submits data to MERMAN
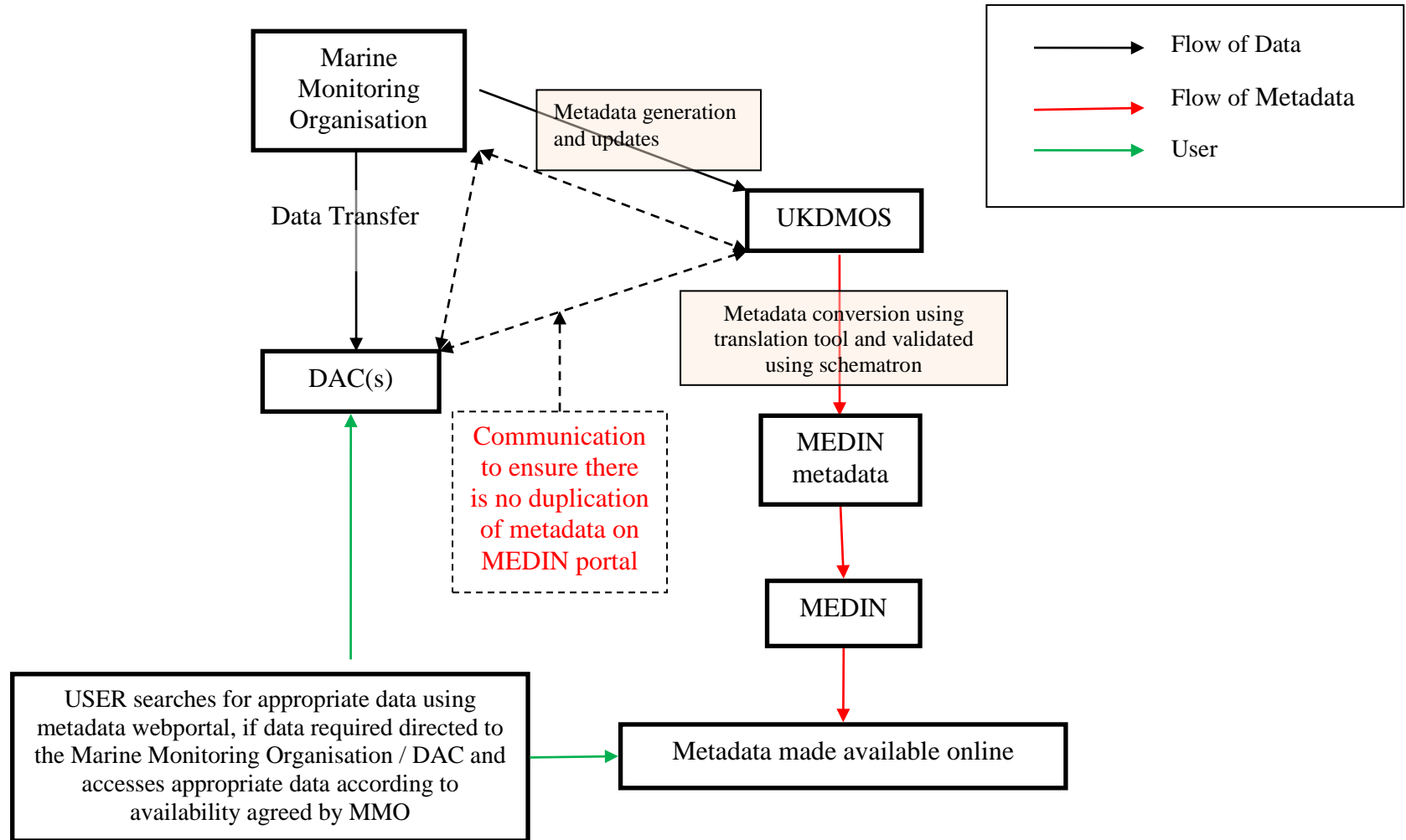
EA	AFBI	Marine Scotland	SEPA	NIEA	CEFAS

Data Submission

MERMAN

Metadata generation and updates

UKDMOS

??Data Transfer on thematic basis??

Metadata conversion using translation tool and validated using schematron

DAC 1	DAC 2	DAC 3	DAC 4

MEDIN metadata

Reassembly of data to be reviewed if when necessary

MEDIN

Metadata made available online

| Flow of Data |
| Flow of Metadata |
| User |

USER searches for appropriate data using metadata webportal, if data required directed to MERMAN

Case Study 5.   An organisation already submits metadata on its routine monitoring operations to UKDMOS.

An organisation runs a number of routine Marine Monitoring Programmes, which contribute to the UK Marine Monitoring and Assessment Strategy. The data sets so described may be archived in a MEDIN DAC, or within the organisation, to be decided with each organisation. The metadata provides information on this. The organisation has generated and provided metadata records describing these programmes to the UK Directory of Marine Observing Systems (UKDMOS), which it reviews and updates each year. MEDIN discovery metadata for these programmes are then generated automatically from UKDMOS by MEDIN using a translational tool, creating a URI and validating the metadata entry using a schematron. The MEDIN metadata entry is then uploaded onto the MEDIN portal.

If the organisation is submitting data to the DACs, then the DACs are instructed that they are not required to generate or serve a metadata record to the MEDIN portal.

If the organisation has already appointed a Unique Resource Identifier for the dataset then this should be retained by the UKDMOS entry and the DAC if required.

Case Study 5.   An organisation already submits metadata on its routine monitoring operations to UKDMOS.

Note 1: The MEDIN online Metadata creation tool
This tool is being developed by DASSH, and will create MEDIN format discovery metadata records and automatically check the content and format, prompting the user to make corrections as necessary.
Questions:
Can the tool have an option to produce the Unique Resource Identifier?
Can the tool have an option to offer publication of the Metadata Record on the DASSH OAI server?

(NB Seazone have also developed a stand-alone metadata production tool, which they have modified to meet the new MEDIN discovery standard)

Note 2: Could MEDIN have a simple stand alone online tool to generate a Unique Resource Identifier? The tool should be able to check if a matching (or closely matching record exists), to display the closely matching record and ask the user to confirm this is a new record.

Note 3: This would require an add-on to UKDMOS that checks the MEDIN discovery portal for an existing matching entry, translates the UKDMOS metadata record into MEDIN discovery format, and offers options as to where the new MEDIN metadata record should be held and published.

Table indicating the key processes involved in creating MEDIN standard metadata and where these processes may be most appropriately carried out

	1	2	3	4	5
	Habitat Survey by Countryside Agency	**Environmental Survey by Policy Agency**	**Research Cruise by Fisheries Laboratory**	**Data Collected for CSEMP and submitted to MERMAN**	**Metadata on Routine Monitoring programmes submitted to UKDMOS**
Unique Resource Identifier	Contractor / MEDIN URI tool	Data collection contractor or interim data management contractor/ MEDIN URI tool	Fisheries Laboratory / MEDIN URI tool	Organisation/UKDMOS	Monitoring Agency / MEDIN URI tool
Metadata Generation	Contractor - MEDIN standard / MEDIN mdata tool	Data collection contractor or interim data management contractor: MEDIN standard / MEDIN mdata tool	Created through mapping from CSR by DAC or Fisheries Lab	MERMAN/UKDMOS	Monitoring Agency to UKDMOS and transfer to MEDIN
Validation	MEDIN Tool	MEDIN tool	MEDIN tool	UKDMOS to MEDIN metadata transfer	MEDIN translation tool
Publication	MEDIN (DAC)	MEDIN (DAC)	MEDIN (DAC)	MEDIN (DAC)?	MEDIN
Data Generation	MEDIN standards	MEDIN standards	MEDIN / IOC standards	WQ agency	Monitoring Agency
Quality Control	Countryside Agency /DAC on upload	Data collection contractor / interim data management contractor / DAC	Fisheries Laboratory /DAC on upload	WQ agency (Quasimeme, Bequalm)	Monitoring Agency / DAC on upload
Archival	MEDIN DAC	Intermediate storage by intermediate data management contractor then MEDIN DAC	MEDIN DAC	MEDIN DAC	Monitoring Agency / or MEDIN DAC
Update management	Countryside agency advises MEDIN DAC		Fisheries Laboratory advises MEDIN DAC	Via updates to MERMAN	Monitoring Agency advises MEDIN DAC
Tools used?	MEDIN online metadata tool MEDIN URI tool	MEDIN online metadata tool MEDIN Data Clause applied MEDIN URI tool	CSR metadata mapping / MEDIN URI tool	MEDIN URI tool MERMAN	MEDIN URI tool MEDIN Translation Tool