Antimicrobial resistance surveillance pilot site selection and database extension

Chief Scientist's Group report

July 2022

SC210012

We are the Environment Agency. We protect and improve the environment.

We help people and wildlife adapt to climate change and reduce its impacts, including flooding, drought, sea level rise and coastal erosion.

We improve the quality of our water, land and air by tackling pollution. We work with businesses to help them comply with environmental regulations. A healthy and diverse environment enhances people's lives and contributes to economic growth.

We can't do this alone. We work as part of the Defra group (Department for Environment, Food & Rural Affairs), with the rest of government, local councils, businesses, civil society groups and local communities to create a better place for people and wildlife.

Published by:

Environment Agency
Horizon House, Deanery Road,
Bristol BS1 5AH

www.gov.uk/environment-agency

Further copies of this report are available from our publications catalogue: www.gov.uk/government/publications or our National Customer Contact Centre: 03708 506 506

Email: research@environment-agency.gov.uk

Author(s):
Andrew Hargreaves, Vera Jones, Andrew Singer, Matt Shipton, Isobel Stanton, Holly Tipper

Checkers: Glenn Castle, Pau Juan Garcia

Reviewers: Mark Blackmore

Keywords:
antimicrobial resistance, AMR, surveillance,

Research contractor:
Atkins (lead contractor); UK Centre for Ecology and Hydrology

Environment Agency's team:
Wiebke Schmidt, Jono Warren, Alwyn Hart

Project number: SC210012

This work was carried out as part of on-going Environment Agency research into AMR in the environment and has been funded by the PATH-SAFE programme which has been funded by HMT through the Shared Outcomes Fund.

Research at the Environment Agency

Scientific research and analysis underpins everything the Environment Agency does. It helps us to understand and manage the environment effectively. Our own experts work with leading scientific organisations, universities and other parts of the Defra group to bring the best knowledge to bear on the environmental problems that we face now and in the future. Our scientific work is published as summaries and reports, freely available to all.

This report is the result of research commissioned by the Environment Agency's Chief Scientist's Group.

You can find out more about our current science programmes at https://www.gov.uk/government/organisations/environment-agency/about/research

If you have any comments or questions about this report or the Environment Agency's other scientific work, please contact research@environment-agency.gov.uk.

Dr Jo Nettleton
Chief Scientist

Contents

Executive summary

Antimicrobial resistance (AMR) represents a major concern for human, animal and plant populations, and the wider environment. The Government's 20-year Vision for AMR and the 5-year National Action Plan outline how the UK will address the AMR challenge. This includes specific reference to the importance of better understanding the potential spread, transmission, and risk of AMR in the environment. The current project follows on from two earlier studies (Environment Agency, 2020; Environment Agency, 2021), which set out the framework for AMR in the environment and created a database of information relating to AMR in the environment, respectively. The aim was to take this earlier work further, proposing a selection of sample locations for a pilot AMR surveillance study, expanding the existing AMR database, and mapping relevant AMR datasets. The project was divided into four key tasks.

Task 1 set out to develop selection criteria for the identification of suitable river catchments for environmental AMR surveillance. After consultation with academics, this selection focused on the presence/ proximity of combined sewer overflows (CSOs) and the frequency of CSO spills. Following a deductive approach, the following four Water Framework Directive (WFD) catchments were proposed for selection: Croal Irwell, Ellen and West Coast, Goyt Etherow Tame, and Nottingham Urban. The Torridge WFD catchment was also added to the list, following discussion with the Environment Agency team.

Task 2 set out to expand the existing AMR geodatabases. A number of new datasets were added to the updated databases, which included: datasets that were utilised in Task 1 of this study, datasets identified in an earlier study (Environment Agency, 2020) as "Priority 2" and "Priority 3", and highest priority ("Priority 1"; Environment Agency, 2020) datasets which represent 'areas' for which no geospatial datasets were previously processed, identified from a high-level gap filling activity.

Task 3 focused on identifying gaps in the existing AMR datasets compiled in Task 2 and proposing ways to infill them. Datasets were assessed in terms of spatial and temporal coverage. All of the 48 databases identified were considered spatially complete. There were 25 datasets that were considered temporally "up-to-date", 21 that were considered "out of date" and one that had no information on the date range at which it was conducted. Suggestions for infilling data gaps included primary data collection and estimation based on existing datasets.

Task 4 focused on mapping environmental AMR datasets. For each WFD operational catchment across England a score for environmental AMR abundance and catchment sensitivity to AMR was calculated using agreed variables. These were then combined to produce an overall score for AMR exposure for each WFD operational catchment. It is noted that this assessment of environmental AMR risk represents a starting point for this type of analysis, and is based on one approach for scoring risk (in this case with a focus on coastal areas); different results would have been produced if the approach was adapted and different weightings were assigned to the agreed variables. The site selection from Task 1 was finally reviewed considering the exposure map developed in Task 4.

This study has provided a key comprehensive resource for future work to improve our understanding of AMR in the environment across England.

Introduction

Antimicrobial Resistance (AMR) arises when microorganisms evolve in ways to be resistant to antimicrobial substances, such as antibiotics, antifungals, antivirals, etc. AMR exists in natural environments, but as human, animal, and agricultural activities are increasing, so does its prevalence; therefore, posing a potential health risk to people, animals, food sustainability, and ecosystems. The Government's 20-year Vision for AMR and the 5-year National Action Plan outline how the UK will address the AMR challenge. This includes specific reference to the importance of better understanding the potential spread, transmission, and risk of AMR in the environment.

The current project follows on from two earlier studies (Environment Agency, 2020; Environment Agency, 2021), which set out the framework for AMR in the environment and created a database of information relating to AMR in the environment, respectively. The Environment Agency (2021) study also explored visualisation options for the datasets contained in the database.

The current study encompasses four key tasks, which will provide information for an upcoming environmental AMR surveillance pilot study:

- Task 1: Co-development of selection criteria and identification of suitable river catchments for environmental AMR surveillance;
- Task 2: Extension of the existing AMR database;
- Task 3: Identification of data gaps and suggestions for how to address them; and
- Task 4: Mapping the environmental AMR hazard in England, as relevant to a future pilot freshwater AMR surveillance programme.

The report sections below outline the key outcomes from these tasks.

Task 1: Co-development of selection criteria & identification of suitable river catchments for environmental AMR surveillance

Background

The following aspects are included in this section:

- Background and objective of the task;
- Methodology for selecting catchments, including consultations with experts in the field; and
- Deliverables for Task 1: key catchments selected as suitable for AMR surveillance.

As part of this task, the project team and the Environment Agency co-developed selection criteria to identify suitable river catchments for piloting a surveillance programme for environmental AMR. Selection focused on sources of AMR-driving chemicals and AMR genes.

The determination of high-priority river catchments for AMR surveillance was based on the following assumptions about the role of point and diffuse pollution sources:

1. Point pollution source: Treated, and particularly untreated, sewage discharge would dominate the AMR signal found in freshwaters. Numerous studies highlight/ confirm this assumption (e.g. Singer et al., 2021; House of Commons Environmental Audit Committee, 2022). Inputs of treated sewage discharges and, in many cases, untreated sewage from combined sewer overflows (CSOs), can be constant or at the very least frequent (Hammond et al., 2021). Untreated sewage, such as that released from CSOs, will have a significantly higher load of AMR genes, antimicrobials and total microorganisms in comparison to that of treated wastewater by approximately two to three orders of magnitude (Phillips et al., 2012; Eramo et al., 2017; Honda et al., 2020; Singer et al., 2021).
2. Diffuse pollution source: In addition to point sources, agricultural inputs will be important for some river catchment and river stretches. These inputs are diffuse and irregular in their frequency and discharge quantity, and occur most frequently and with greater impact during heavy rains. This pollution can include: chemicals (such as antimicrobials and other co-selecting agents) and bacteria (including both pathogenic and antimicrobial resistant bacteria; House of Commons Environmental Audit Committee, 2022).

Criteria for the selection of catchments for a pilot surveillance study has developed after consultation with experts in the field of AMR in the environment, including William Gaze and Anne Leonard of the University of Exeter, Daniel Read of the UK Centre for Ecology &

Hydrology, Nicole Stoesser and Kevin Chau of the University of Oxford, Dov Stekel of the University of Nottingham, and Alexander Corbishley of the University of Edinburgh. Their feedback can be found in the Methodology & Outputs Section. The experts were in agreement in the attribution of CSOs as one of the most influential drivers of AMR in freshwaters. Other point and diffuse pollution sources (e.g., land use) were identified and incorporated into the Hazard Assessment component of Task 4, which developed a map of high-hazard catchments for AMR based on these additional drivers (see the Section describing Task 4).

Methodology & Outputs

Expert Consultation

Emails were sent to the experts identified: William Gaze, Dov Stekel, Nicole Stoesser and Alexander Corbishley. They, in turn, recruited their colleagues: Anne Leonard, Daniel Read and Kevin Chau.

The initial email described our reasons for proposing CSOs as the main criteria for selection of catchments. Three questions were posed to the experts:

1. Do you think that it is justified to prioritise CSO discharges when looking for highly impacted catchments with respect to AMR?
2. What other datasets would you use? If a dataset does not exist for what you find important, please also share this.
3. Do you have any suggestions or comments on future refinements?

Of the seven experts, four explicitly agreed and two implicitly agreed that prioritising CSOs were the right choice. The final expert noted that the use of CSO data for informing hazard assessment can be better informed if water use data was also available (e.g. bathing water status, or water user data), thereby allowing the assessment to include a measure of exposure risk to untreated sewage. Another expert suggested that faecal pollution is highly correlated with exposure risk of harbouring antibiotic-resistant *E. coli* and, therefore, choosing catchments with less polluting CSO sites would be choosing catchments with a reduced AMR exposure risk to humans (e.g. lower abundance of AMR genes relative to the total population of microorganisms). Based on the opinion from all experts contacted, it was justified to continue with prioritising catchments based on CSO events and duration.

In addition to commenting on the CSO prioritisations, six of the seven experts provided additional datasets that they deemed relevant for predicting the pollution of downstream catchments with AMR. These can be summarised in five broad categories:

1. Treated wastewater (volume, size of treatment plant, dilution effect of downstream rivers, healthcare effluent, and application of biosolids to land).
2. Bathing water (history of poor water quality, catchments that terminate at bathing waters that regularly fail bathing water standards).
3. Animal faeces (animal contribution and density of grazing animals).
4. Chemical pollution (use of chemical contaminants as proxies for anthropogenic pollution/ impact).

5. Healthcare (local clinical AMR rates, prescribing pattern and hospital wastewater effluent).

These suggestions were used to guide the hazard mapping conducted in Task 4. Pristine locations upstream of diffuse and point sources were recommended for use as control sites, with the caveat that they might be unknowingly impacted by additional sources. Regions devoid of CSOs could still be impacted by faecal pollution originating from septic tanks, but at the time of reporting, there is no data on the location of and relative contribution of septic tanks to freshwater. As such, the absence of some CSOs and all septic tanks represents a potentially significant gap that would benefit from greater confidence in future iterations of this procedure.

Overview of process for identifying locations

The process for determining the location within England for future AMR surveillance in freshwaters employed: the Water Framework Directive (WFD) operational catchment shapefiles, the CSO discharge data in England in 2020 and the feasibility of sampling downstream of CSO sites. The following steps were taken, in series, with greater detail provided in the Methodology & Outputs section:

1. Mapping of 422 WFD operational catchments with active CSOs.
2. CSOs were retained if they reported >100 releases per year. A threshold of >100 was needed owing to the very high number of CSOs with a high number of discharges per year. Catchments containing CSOs that do not contain event duration monitors or collected incomplete data in 2020 were omitted from the analysis.
3. Catchments containing CSOs retained from Stage 2 were further assessed for CSOs that discharged a total of over >10,000 minutes per year. This criteria was chosen as it was desirable to have sampling locations that could be assured of sewage impact, on average, more than once a week.

   Both criteria (events and duration) also assured that regardless of the sampling schedule undertaken by the Environment Agency, a CSO discharge would likely occur within days preceding the sampling time point.
4. Consideration was subsequently given to the logistical challenge of accessing the identified locations in Stage 3. To this end, the Environment Agency team provided a list of potential suitable sampling sites across England that were used to further refine the catchments to those containing an Environment Agency preferred sampling location <1 km downstream of a high priority CSO identified in Stage 3. The broad selection criteria included:

a) Removal of catchments without Environment Agency selected sites;
b) Removal of catchments that only had Environment Agency sites >1 km away from a high risk CSO (more likely to find high levels of AMR the closer to CSO);
c) Removal of catchments which only had sampling sites upstream of a high risk CSO (higher levels of AMR found downstream of CSOs);

d) Removal of catchments with only one sampling location close (<1 km) to a high risk CSO (offering catchments with more relevant sampling locations over the catchment close to (both up and downstream) of CSO sites); and
e) removal of catchments where Environment Agency sites are >100 m away from high risk CSOs (to reduce the number of catchments further as they suggested in a meeting they wanted the closest possible sites to CSOs). This threshold was established to ensure that a sampling location would be maximally impacted by sewage and display a significant AMR signature, typical of a sewage-impacted freshwater ecosystem.

First cut of catchments (422 to 48 catchments):

ArcMap 10.6.1 was used with the layers Event Duration Monitoring – Storm Overflows – 2020 (England and Wales) (to plot CSO sites) and WFD Surface Water Operational Catchments Cycle 2 (to plot England's operational catchments). Figure 1 shows the CSOs from this data layer in England.

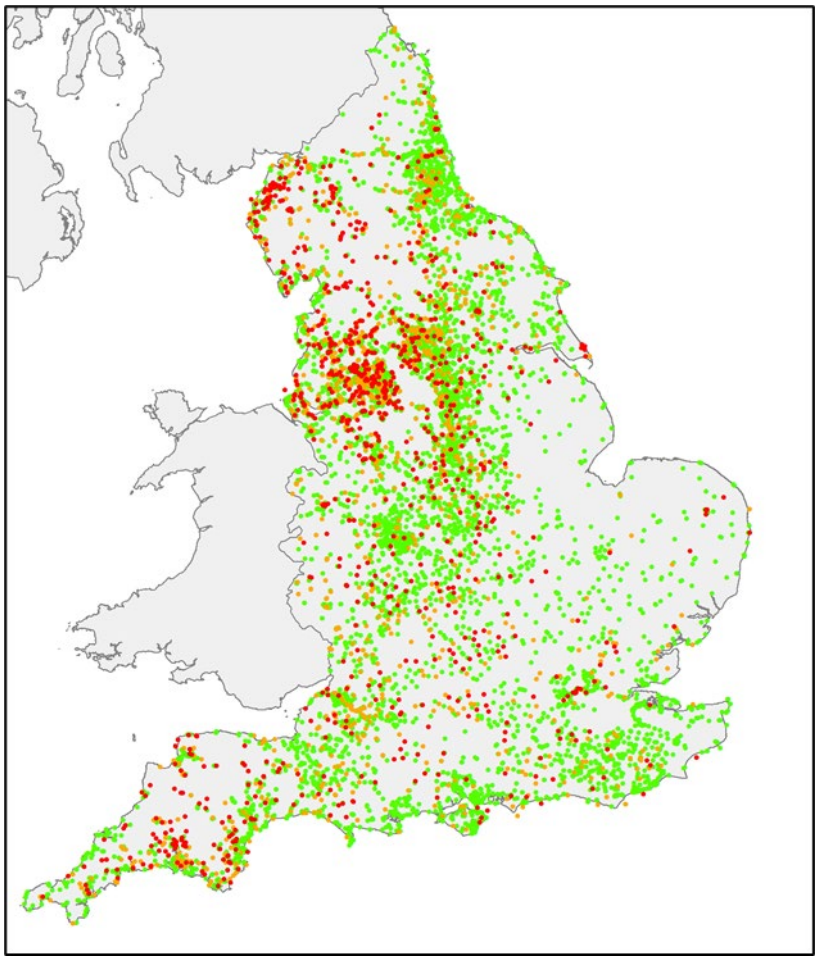


Figure 1 - CSO sites in England. Data provided from "Event Duration Monitoring – Storm Overflows - 2020 (England and Wales)". Sites are colour coded by number of discharges per year: green = 1 to 60 discharges; orange = 61 to 100 discharges; red >100 discharges.

The "Intersect" tool in ArcMap was used to group CSO sites by operational catchments. The resulting data was exported to Excel. Excel was used to filter the data to remove CSO sites

that had equal to or less than 100 discharges a year (resulting in 247 operational catchments remaining).

Subsequently, from the 247 operational catchments, the CSO sites that discharged over 100 times a year were investigated for their total discharging duration (minutes) per catchment using the =sumif() function in Excel. Catchments with >10,000 discharging minutes per year at their CSO sites and with over 100 spills a year were deemed the most polluted in terms of faecal pollution and were selected for further investigation. This resulted in 48 catchments.

Figure 2 shows the locations of the 48 operational catchments containing CSO sites discharging over 100 times per year.

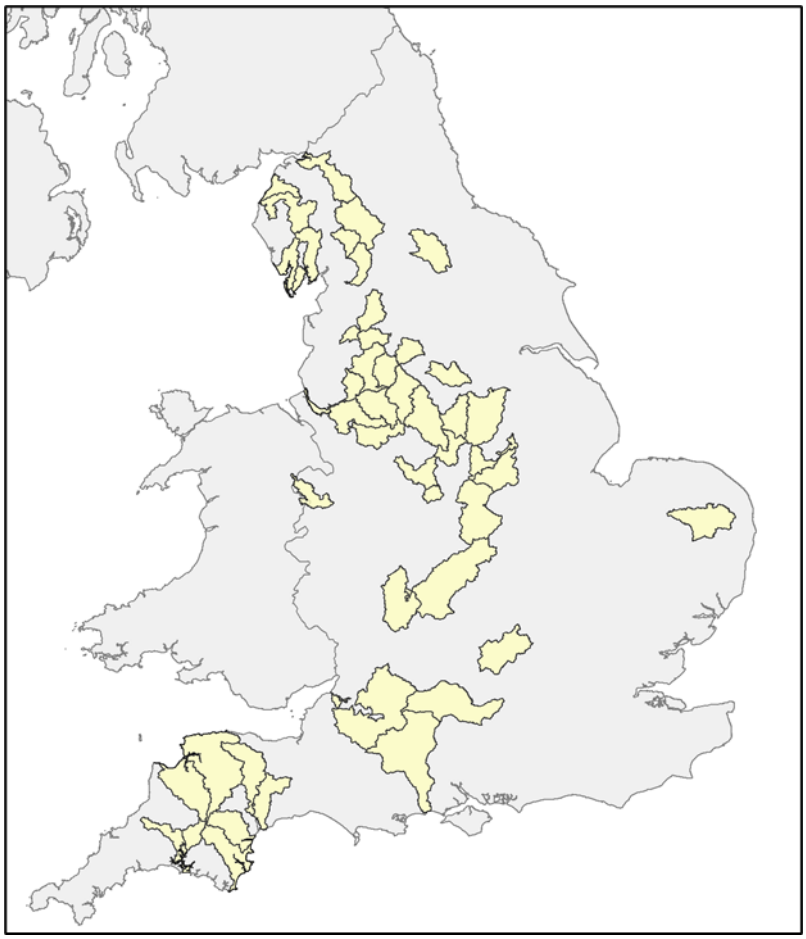


Figure 2 - 48 operational catchments (yellow highlighted areas) identified in first cut of catchments by CSO severity.

These 48 catchments were handed to the Environment Agency team to identify feasibility of sampling sites within these catchments.

Second cut of catchments (48 to 4 catchments):

Figure 3 shows the sampling sites provided by the Environment Agency team that are suitable to be sampled under an environmental AMR surveillance pilot study.

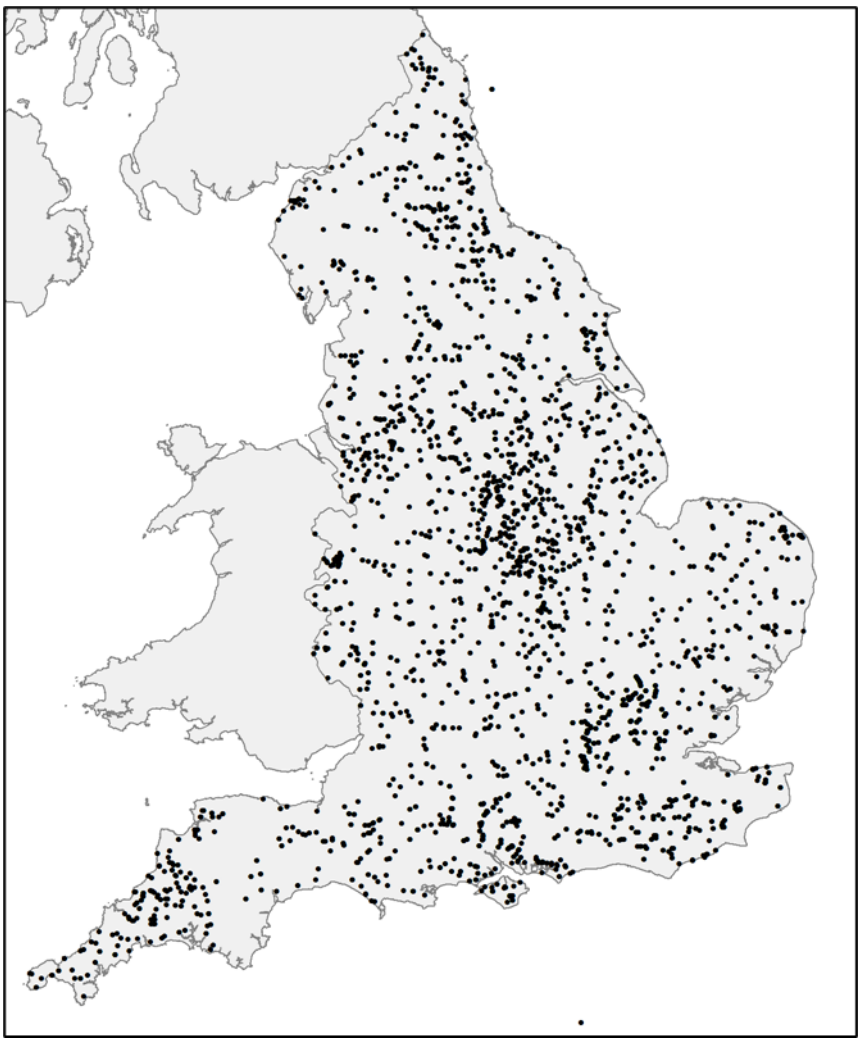


Figure 3 - Potential field site locations identified by the Environment Agency team (black points).

By using the locations of these sites, we were able to reduce the 48 catchments selected in the first cut to four catchments (see Figure 4 below) by using a further set of defined steps (presence of Environment Agency sites in catchment, proximity of Environment Agency site to CSO site, Environment Agency site downstream of CSO, more than one Environment Agency site in close proximity downstream of CSO). Details of the operational catchments included at each stage of the selection are described below and can be found in a supplementary file provided (task1.xml).

Eight catchments were removed as they did not contain any sampling points the Environment Agency deemed practicable to sample at. A total of 40 catchments progressed to the next stage.

A further 25 catchments were eliminated that did not contain sampling points that were within 1 km of high priority CSO (>100 spills/year) sites. This was as a result of discussions with the Environment Agency, agreeing that the preference was for river catchments where sampling sites were closest to CSO (>100 spills/year) sites. Five catchments were removed where the Environment Agency sites were found only upstream of high priority CSO (>100 spills/year) sites. This was established by evaluating the direction of the flow of the river

using the WFD River Water Bodies – Cycle 2 data layer which displays rivers. This reduced the list to 10 potential catchments.

In addition, catchments that contained only one suitable Environment Agency sampling site were removed (based on the aforementioned criteria), removing three catchments from further consideration.

The final seven catchments (listed in Table 1) were all deemed suitable for sampling based on the aforementioned criteria.

Table 1 - List of seven potential operational catchments for sampling. Number of sites = the number of sites that meet the criteria: have an Environment Agency sampling site within 1 km of a CSO that discharges over 100 times a year, represents some downstream sampling sites within the catchment and where more than one sampling site per catchment has been identified. Down = downstream, up = upstream.

WFD Operational Catchment ID	WFD Operational Catchment name	Number of Environment Agency sampling sites	Distance between Environment Agency site and high priority CSO site along WFD river (upstream/ downstream/ same site)
3039	Bollin Dean Mersey Upper	4	420 m (down), 560 m (up), 400 m (down), 650 m (down)
3116	Croal Irwell	7	150 m (up), 200 m (up), 20 m (same site), 50 m (down), 90 m (up), 20 m (same site), 630 m (down)
3168	Ellen and West Coast	3	120 m (down), 4 m (same site), 60 m (downstream)
3206	Goyt Etherow Tame	4	40 m (down), 220 m (up), 570 m (down), 380 m (up)
3341	Nottingham Urban	2	40 m (down), 150 m (down)
3382	Roch Irk Medlock	3	450 m (up), 650 m (down), 800 m (up)
3407	Soar River	4	880 m (down), 550 m (up), 100 m (up), 930 m (up)

To refine these further, catchments with the closest sampling sites to CSO (>100 spills/year) sites were chosen. To do this, we excluded catchments where the sites were more than 100 m downstream of the CSO sites.

This resulted in the final selection of four operational catchments: Croal Irwell (3116), Ellen and West Coast (3168), Goyt Etherow Tame (3206), and Nottingham Urban (3341). The location of these catchments can be seen in Figure 4 .

Figure 4 - Final selection of operational catchments (yellow highlighted areas). These are Croal Irwell, Ellen and the West Coast, Goyt Etherow Tame, and Nottingham Urban.

Deliverables: Final catchments

Figure 5 shows maps of the final operational catchments (grey) with high risk CSOs (red points) and Environment Agency sampling sites (black points) on the rivers (blue) for catchments 3116, 3168, 3206 and 3341, respectively. Whilst these catchments have been deemed suitable for sampling at relevant locations within the specific catchment, the 48 catchments presented at the end of the first cut are all suitable for sampling in terms of AMR risk from human faecal pollution.

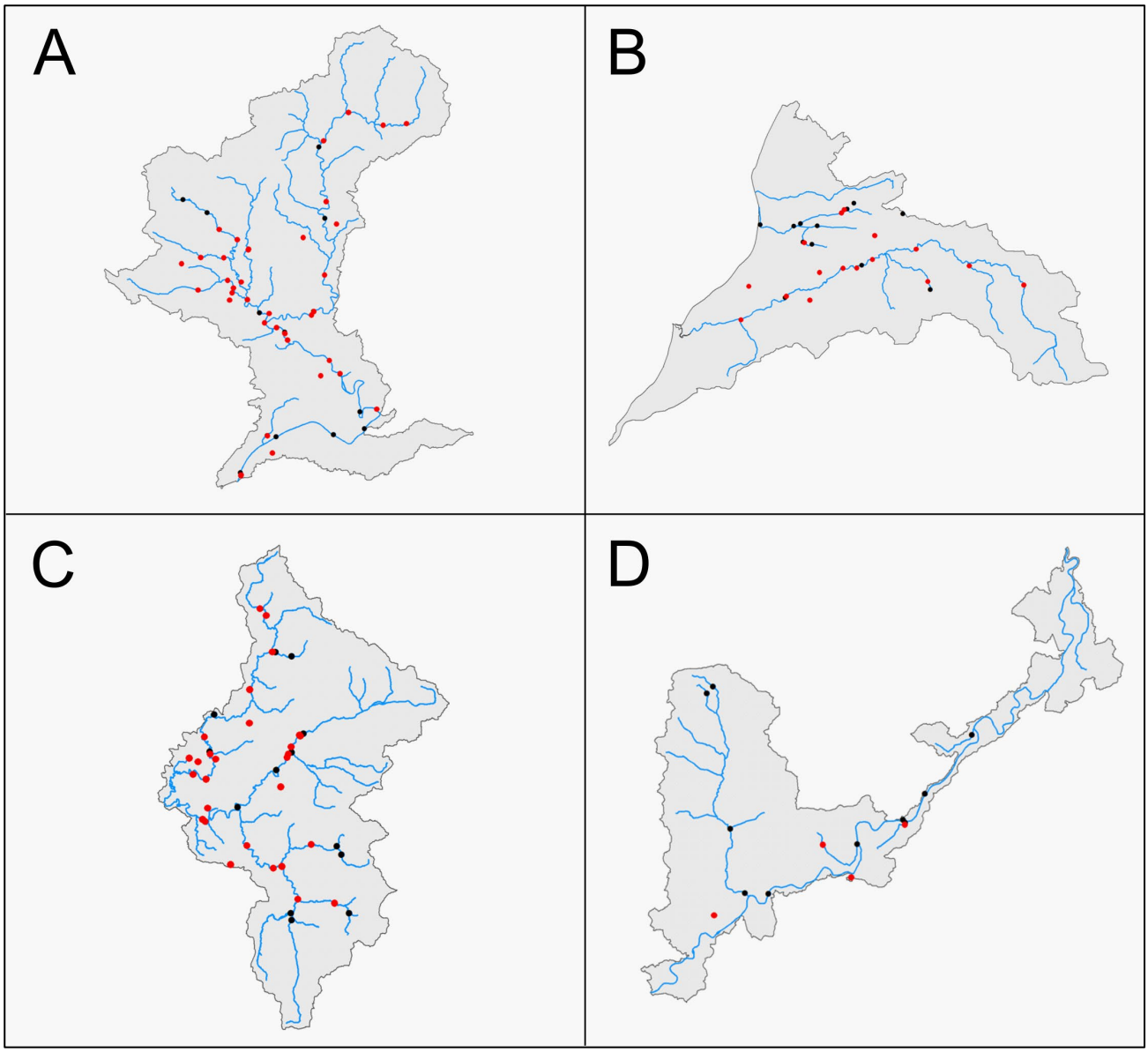


Figure 5 - (A) Croal Irwell (3116), (B) Ellen and West Coast (3168), (C) Goyt Etherow Tame (3206) and (D) Nottingham Urban (3341). Grey highlights the operational catchment, blue highlights the rivers, CSO sites (>100 spills per year) are shown by the red points and Environment Agency sampling sites are shown by the black points.

Additional catchment

The Environment Agency asked the project team to include operational catchment 3468 (Torridge) in the final catchment selection (Figure 6). The stated methodology outlined above initially removed it as it had only one Environment Agency sampling site upstream of a CSO site and no downstream sites within 1 km; the remaining sampling sites were all upstream. In addition, this catchment had seven high risk CSO sites (>100 discharges a year), with the most polluting CSO site discharging 326 times in 2020, and eight medium risk CSO sites (between 61-100 discharges a year). It is clear that this catchment would be highly relevant to sample in terms of AMR surveillance from faecal pollution. In addition, this catchment has both bathing water and shellfish sites and is, therefore, deemed important in

terms of human health exposure risk. However, constraints of sampling sites make it logistically difficult to capture the regions of the catchment with the highest faecal pollution.

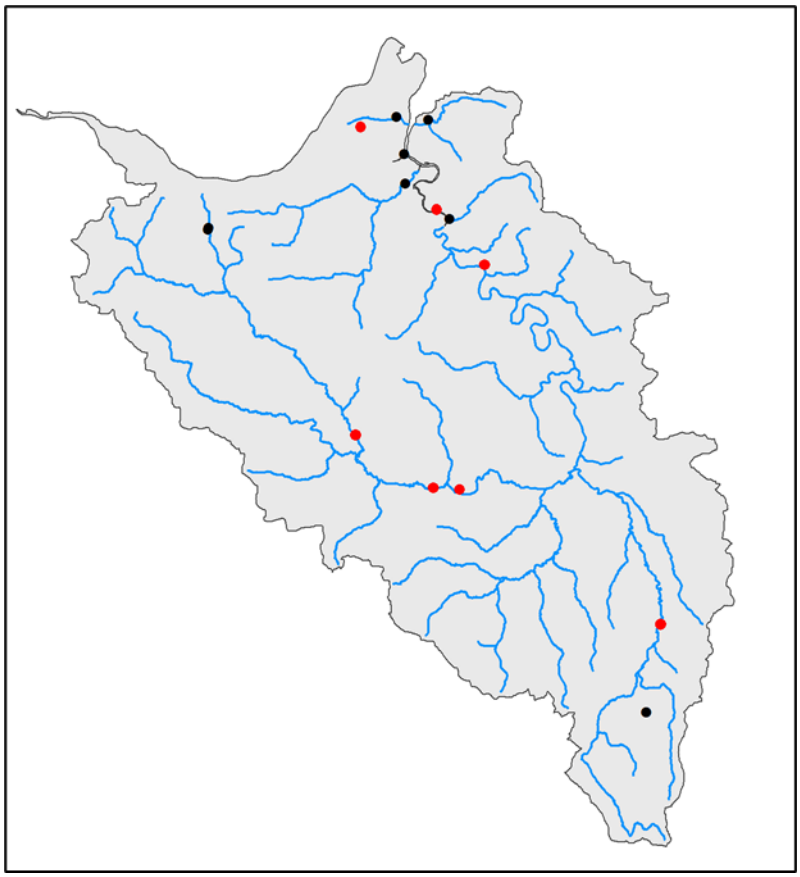


Figure 6 - Torridge (3468). Grey highlights the operational catchment, blue highlights the rivers, CSO sites (>100 spills per year) are shown by the red points and Environment Agency sampling sites are shown by the black points.

Limitations

Recognised limitations of the approach are set out below:

- One limitation of the methodology used in Task 1 is that it prioritises catchments that have a significant urban impact, while offering little insight into catchments that could have significant agricultural impact (e.g. runoff from farms, land spreading, grazing). Moreover, the focus on CSOs does not imply it is a risk to human health, however, it is implicit that a CSO discharge will present an AMR hazard to the receiving water body. This limitation of Task 1 has been addressed in Task 4 where a more extensive hazard map was developed, inclusive of these wider AMR drivers and exposure routes (e.g., bathing waters). In addition, a review of catchment selection based on the hazard map in Task 4 and a comparison with catchments selected in this task has been undertaken.
- A limitation of the methodology used in Task 1 is that it was constrained by the location of existing Environment Agency sampling sites. Many very heavily CSO impacted sites were ruled out because of this requirement.

Task 2: Extension of AMR geodatabases

Background

The following aspects are included in this section:

- Background on the previous project in which the AMR geodatabases were developed;
- Datasets that were investigated for inclusion in the AMR geodatabases in this study;
- Data processing and geodatabase updates made; and
- Deliverables from Task 2.

The overarching aim of previous work (Environment Agency, 2021) was to bring together various datasets about potential sources of AMR in the environment into one database, as an easy-to-access resource that allows the inclusion of new data as and when available. Datasets relating to AMR in the environment that were identified as part of the recent Environment Agency/ Defra report (hereafter referred to as 'AMR report'; Environment Agency, 2020) were available in an Excel spreadsheet and had been classified as Priority 1, 2, or 3 based on the report author's expert opinion. In summary, previous work located and compiled Priority 1 datasets that were categorised as 'freely available' or 'available through UKCEH' from the AMR report into file geodatabases (hereafter referred to as 'AMR geodatabases').

ESRI's proprietary format, a file geodatabase is a collection of files in a folder that can store, query, and manage spatial and nonspatial data and to improve the user's ability to work and share the AMR datasets they were split into four AMR geodatabases based on key themes:

- **Environment** – Water environment and environmental designations.
- **Land Cover** – Farming, agriculture and other land uses.
- **Anthropogenic** – Medical, veterinary and household pets, wastewater and solid waste.
- **Regulatory** – Environmental permits and pollution.

Datasets

Datasets that were investigated in this study can be separated into the following categories:

- Priority 1 datasets from the AMR report that were requested but unavailable in time for the previous project;
- Fish Health Inspectorate (FHI) data for aquaculture businesses;
- Priority 2 and Priority 3 datasets from the AMR report;
- Datasets that were utilised in Task 1 of this study; and
- Priority 1 datasets representing 'areas' for which no geospatial datasets were previously processed – identified from a high-level gap filling activity.

Datasets, for which more detail is provided in the sub-sections below, are summarised by category (above) in Table 2, which also includes identified status (e.g. available, unavailable), action taken i.e. whether the dataset was processed as part of this project and, where applicable, the location of a dataset in the deliverables from Task 2 (see 'Deliverables' section for more detail).

Priority 1 datasets requested, but unavailable in time for the previous project

UK Lakes Portal

Regarding the UK Lakes Portal, it was identified, in the previous project (Environment Agency, 2021), that the polygons of the lakes are freely available to anyone and the dataset (hereafter referred to as 'Lake Polygons' dataset) was requested though was not delivered in time to be processed as part of the project. The original request for the polygons was submitted to the UK Centre for Ecology and Hydrology (UKCEH) contact for this data (Philip Taylor; philor@ceh.ac.uk) on 15/03/2021 and a follow up email requesting the data was submitted during the initial stages of this project (04/11/2021), with no reply received (as of 10/12/2021). The polygons available from the UK Lakes Portal are for waterbodies > 1 ha, whereas those from the alternative 'WFD Lake Water Bodies Cycle 2 Classification 2019' dataset processed (during the previous project) are less refined, covering waterbodies > 5 ha (in protected) and > 50 ha in non-protected areas. Therefore, it is recommended that the Environment Agency submit a further request for this data for inclusion in this database.

In the online version of the UK Lakes Portal (UKCEH, n.d.), data for a number of attributes are available for each lake. It is, however, uncertain which attribute data could be made available to the Environment Agency. It was identified in the previous project (Environment Agency, 2021) that attribute fields are gathered from numerous sources, such that it would be untenable for UKCEH to identify all attributes in the underlying UK Lakes Portal dataset for which data is freely available to the Environment Agency. An initial request was submitted to the UKCEH contact for this data (Oliver Robertson; oliv@ceh.ac.uk) for a smaller list of four attributes[1] on 15/03/2021. One attribute from a selection of the headings on the UKCEH Lakes Portal website was selected as it is a possibility that underlying data (freely available to the Environment Agency) may be found for multiple attributes of similar type (although this unknown). A follow up email requesting an update as to whether the data for the list of attributes would be freely available to the Environment Agency was submitted during this project (04/11/2021), with no reply received (as of 10/12/2021). It was envisaged that should data be available it could be included in the attribute table for lake polygons.

[1] Catchment-to-lake-ratio, mean alkalinity, catchment strahler 1 length, catchment area (agreed with the Environment Agency)

Table 2 – Datasets investigated in this study; identified status, action taken in this project and location in the deliverables from Task 2

Dataset Type	Dataset ID[2]	Title	Resource Status	Action taken in this project	Location of the Dataset – Task 2 deliverable[3]
Priority 1 datasets requested but unavailable in time for the previous project	20	UK Lakes Portal – polygons	Data requested, not yet received	N/A[4]	N/A
	20	UK Lakes Portal – attribute data	Data requested, not yet received	N/A	N/A
FHI data for aquaculture businesses	72	Aquaculture businesses register	Unavailable	N/A	N/A
Priority 2 and Priority 3 datasets	03	Public Rights of Way (Council / open datasets)	Unavailable	Not Processed	N/A
	04	OS MasterMap Highways Network - Paths	Available[5]	Not Processed	Raw data repository
	05	OS Detailed Path Network	Freely Available[6]	Processed	AMR geodatabase
	06	OS Vectormap Local	Available	Not Processed	N/A
	07	OS Vectormap District – Roads	Available	Processed	AMR geodatabase
	22	Estimates of manure volumes by livestock type and land use for England and Wales	Freely Available	Processed	AMR geodatabase

[2] Code used in the Excel spreadsheet that identified the relevant AMR datasets in the previously mentioned AMR report (Environment Agency, 2020)
[3] Described in more detail in 'Deliverables' section
[4] Not applicable
[5] Available to the Environment Agency under an existing licence/ agreement (applicable where 'Available' used thoughout the table)
[6] Freely available to the Environment Agency (applicable where 'Freely Available' used thoughout the table)

Dataset Type	Dataset ID[2]	Title	Resource Status	Action taken in this project	Location of the Dataset – Task 2 deliverable[3]
	23	Methane gas detection from Airborne Visible and Infrared Imaging Spectrometer Next Generation (AVIRIS-NG)	Unavailable	N/A	N/A
	25	Traditional Orchards HAP (Provisional) (England)	Freely Available	Processed	AMR geodatabase
	31	Liquid chromatography-mass spectrometry (LCMS) Target and Non-Targeted Screening	Freely Available	Processed	AMR geodatabase
	33	Environment Agency Water quality archive	Freely Available	Not Processed	N/A
	37	Important Bird Areas UK	Freely Available	Processed	AMR geodatabase
	38	Sites of Special Scientific Interest (England)	Freely Available	Processed	AMR geodatabase
	40	Pets UK	Freely Available	Processed	Non-spatial data repository
	43	Permitted Waste Sites Authorised Landfill Site Boundaries	Freely Available	Processed	AMR geodatabase
	44	Historic Landfill Sites	Freely Available	Processed	AMR geodatabase
	53	Environmental Permitting Regulations – Waste Sites	Freely Available	Processed	AMR geodatabase
	58	Waste Data Interrogator 2017	Freely Available	Processed	AMR geodatabase
	59	Permitted Waste Sites – Animal Disposal Site Boundaries (AfA076)	No link provided	N/A	N/A
	60	Environmental Permitting Regulations – Industrial Sites	Freely Available	Processed	AMR geodatabase
	61	Bio-solid use rates from British Survey Fertiliser Practices	Freely Available	Processed	Non-spatial data repository
	62	1996 Environment Agencyreport on sewage sludge	Freely Available	Processed	Non-spatial data repository

Dataset Type	Dataset ID[2]	Title	Resource Status	Action taken in this project	Location of the Dataset – Task 2 deliverable[3]
	63	Ofwat report appendix on 'Sludge treatment, transport and disposal – supporting evidence and design options'	Freely Available	Processed	Non-spatial data repository
	67	Scientific journal paper: Summary of current knowledge of the size and spatial distribution of the horse population within Great Britain	Freely Available	Processed	Non-spatial data repository
	68	Animal and Plant Health Agency (APHA) Livestock Demographic Data Group population density	Available	Processed	AMR geodatabase
	69	APHA Livestock Demographic Data Group Enhanced Demographic reports	Freely Available	Processed	Non-spatial data repository
Datatsets from Task 1	N/A	Event Duration Monitoring - Storm Overflows – 2020 (England and Wales)	Freely Available	Processed	AMR geodatabase
	N/A	Environment Agency Catchment Data (England)	Freely Available	Processed	AMR geodatabase
Priority 1 datasets from high-level gap filling	46	List of Composting Sites (England) (Geocoded and operating at some point during 2005 - 2014)	Freely Available	Processed	AMR geodatabase
	73	Consented Discharges to Controlled Waters with Conditions – Fish + Aquaculture/Fish Farm/Cress Farm	Freely Available	Processed	AMR geodatabase
	02	Wildswim.com	Unavailable	N/A	N/A
	14	Effluent point discharge data	Freely Available	Not Processed	N/A

Fish Health Inspectorate (FHI) data for aquaculture businesses

In the previous project (Environment Agency, 2021) the only 'freely available' Priority 1 dataset that could not be processed was the online register that the Centre for Environment, Fisheries and Aquaculture Science (CEFAS) provides of aquaculture businesses in England and Wales. It was noted that despite being available as an online register (CEFAS, n.d.) it is not possible to download this dataset, and provision of the underlying dataset was not possible, on request. Nevertheless, previous work identified a potential alternative source of this information; the FHI require information to be provided in fish farm applications (as well as when importing fish; Department for Environment, Food & Rural Affairs (Defra) and CEFAS, 2021), and therefore, it was deemed plausible that they might hold records that are accessible. This potential source was explored during this project and contact was made with the FHI. It was indicated that the FHI are unable to provide any further information to that currently included in the online register (provided by CEFAS)[7]. The FHI were not able to provide the underlying dataset for the same reason given by CEFAS; data that is published on the Register is what is required by the Aquatic Animal Health Regulations (2009) and because of this providing the underlying dataset would breach General Data Protection Regulations (GDPR)[8].

Priority 2 and Priority 3 datasets

Datasets relating to AMR in the environment that were identified as part of the AMR report (Environment Agency, 2020) were available in an Excel spreadsheet. A total of 25 datasets were listed as Priority 2 or Priority 3 in the spreadsheet[9] and their status with regards to availability were identified to be: 'freely available to the Environment Agency', 'available to the Environment Agency under an existing licence/ agreement', 'unavailable' and 'no link provided'. A breakdown of their availability is shown in Figure 7.

[7] Email received from Tracey Bull (FHI) on 09/11/2021

[8] Email received from Sue Dale (CEFAS) on 24/02/2021 (during previous project)

[9] Following removal of duplicate datasets and datasets which included duplicated data

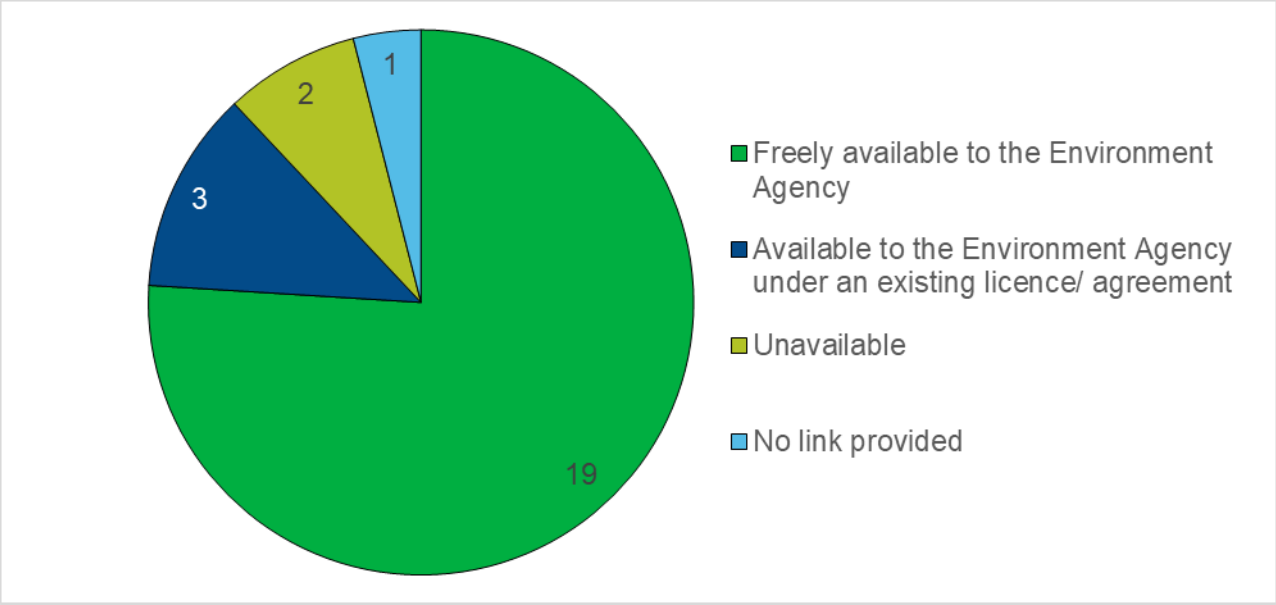


Figure 7 – Priority 2 and Priority 3 datasets identified resource status

Datasets that were either 'freely available to the Environment Agency' or 'available to the Environment Agency under an existing licence/ agreement' were processed[10] in this project. Key notes relating to several individual datasets (ID as included in the Excel spreadsheet that identified the relevant AMR datasets in the AMR report; Environment Agency, 2020) are detailed below.

Liquid Chromatography Mass Spectrometry (LCMS) Target and Non-Targeted Screening (ID31) – it was agreed with the Environment Agency[11] that given the size of this dataset (data for samples collected from 2005 – 2021 for hundreds of compounds) and time constraints of this project a smaller, sub-set of the data (i.e. only data collected during 2020, the last 'full' year for which data is available), would be processed. It is recommended that this dataset is revisited in the future with a view to including data for compounds of specific interest/ relevance to AMR.

WFD River, Canal and Surface Water Transfer Water Bodies Cycle 2 Classification 2019 (ID33) – as agreed with the Environment Agency[11] this dataset, which identifies the river waterbodies managed under the WFD and any related programmes (and includes canals and surface water transfers) and their Cycle 2 (2019) classifications, was processed as an alternative to data from the Environment Agency water quality data archive (Defra and Environment Agency, 2021a; the 'original' ID33 dataset). Though it is recognised that the data archive could be interrogated for pesticides, herbicides, fungicides, and pharmaceuticals data collected at sampling stations across England, such a task would require significant time inputs (as this dataset contains data for thousands of determinands

[10] Except 'OS MasterMap® Highways Network – Paths', 'OS Vectormap Local', and 'Environment Agency Water quality data archive' – reasons for which are described below.
[11] In a progress meeting on 16/11/2021

and sites) and due to the time constraints of this project was not investigated further. This is an action that it is recommended the Environment Agency undertake in the future.

Animal and Plant Health Agency (APHA) Livestock Demographic Data Group population reports (Cattle, Sheep, Goat, Poultry) (ID68) – Datasets that include population density of cattle, poultry, goat, and sheep across the UK at 1 km scale (averaging over a radius of 15 km) based on reporting from farms on each animal type were requested and received from APHA (the contact was Dan Brown, Daniel.Brown@apha.gov.uk). A separate request was made to the Agriculture & Horticulture Development Board (AHDB; contact for the data was Jennifer Newman, Jennifer.Newman@ahdb.org.uk) to access this data for pigs. It was indicated[12] that an official request for this data would be required from the Environment Agency, whilst discussion with the traceability team within Defra may be a prerequisite, as the primary use of the data is for recording animal movements for legislative and disease prevention purposes. Following this Wiebke Schmidt (Environment Agency) contacted her Defra colleagues to seek further advice. They confirmed that 'this route' to accessing the data is difficult and not achievable within the timeframe of this project. An alternative 'route' that could also be explored in the future and which could provide estimates of population density was suggested using the Environment Agency Environmental Permitting Regulations (EPR) inspections (which used to be called Intensive Pig and Poultry Inspections). These are carried out on all intensive pig producers (over 2000 pigs or 750 sows) and given that most pig farms are now specialist and large in size, it is likely that data collected would account for 80-90% of the total pig population in England.

Animal and Plant Health Agency (APHA) Livestock Demographic Data Group Enhanced Demographic reports (Cattle, Pigs) (ID69) – These reports were available and were processed. A request was made to APHA for the raw datasets of any geospatial data presented in the reports. Nevertheless, it was indicated[13] that these data are compiled from different sources for each report (i.e. Cattle and Pigs) and there are different suppliers who require restrictions from APHA on the further sharing of these data. The maps and any accompanying tables are generated from spatial point location data but APHA are very unlikely to be able to share these point locations given the aforementioned restrictions. Given the short timeframe of this project no further investigation into the availability of raw datasets was requested.

OS Vectormap Local (ID6) – Following submission of a request for this data from the Defra Data Services Platform Team, their GIS team stated[14] that obtaining data for the extent of England is untenable as it would take weeks to extract that much data and split it into smaller areas. It has 180 million features, and the total size will most likely be well over 100 gigabytes

[12] In an email received from Jenifer Newman (AHDB) on 8/11/2021

[13] In an email received from the APHA contact (Dan Brown, Daniel.Brown@apha.gov.uk) for this data on 11/11/2021

[14] Email received from Benjamin Knibbs (Defra Data Services Platform Team) on 8/11/2021

therefore the largest the GIS can process is around the size of two counties. Given that the spatial scale of interest in this project was England, no further data request was made.

OS VectorMap District (ID07) and OS MasterMap® Highways Network – Paths (ID04) – Both of these datasets are very large in size and subsequently their inclusion (or even the inclusion of a sub-set of the 'OS MasterMap® Highways Network – Paths' data) would increase any geodatabase to which they were added to a size that is likely to cause problems when trying to deliver them to the Environment Agency, and will inevitably significantly reduce the speed in which the geodatabase can be loaded and viewed/ manipulated. It was noted that the main reason the 'OS MasterMap® Highways Network – Paths' dataset was identified in the AMR report was because the runoff from highways could introduce a substantial quantity of metals and other pollutants into waterways, and we can acknowledge that this source could be important where they intersect rivers. As part of the 'OS VectorMap District' dataset, geospatial data is available for roads which could potentially be used in conjunction with the 'OS MasterMap Water Network Layer' (processed during the previous project) to identify intersections. It was agreed with the Environment Agency that inclusion of the data for roads from the 'OS VectorMap District' was a suitable alternative to including the full 'OS VectorMap District' dataset and the 'OS MasterMap® Highways Network – Paths' dataset. It is also worth noting that at a later stage the Environment Agency may choose to process other data from these datasets for inclusion in the AMR geodatabases (or include the data in a separate repository) as they are available to the Environment Agency under an existing licence/ agreement, for example 'OS MasterMap® Highways Network – Paths' includes data on pathways and rights of way (but only for towns and cities).

Datasets from Task 1

The following datasets that were obtained and utilised in Task 1 were processed:

- Event Duration Monitoring – Storm Overflows – 2020 (England and Wales); and
- WFD Surface Water Operational Catchments Cycle 2

Numerous datasets were available from the link at which the 'WFD Surface Water Operational Catchments Cycle 2' dataset was sourced (Defra and Environment Agency, 2021b) and several were processed, these include those which define WFD Cycle 2 lake, transitional, groundwater, and coastal water bodies and artificial and groundwater catchments.

Priority 1 datasets from high-level gap filling

Following the inclusion of Priority 2 and 3 datasets and datasets from Task 1 (described above), a high-level gap filling activity was undertaken in order to improve the diversity of spatial datasets included in the AMR geodatabases. This involved taking a high-level look at the Priority 1 datasets that were not included as part of previous project work (i.e. those which were not categorised as 'freely available' or 'available through UKCEH'; Environment Agency, 2021); identifying 'areas' for which no geospatial datasets were previously

processed and investigating whether these 'gaps' could be filled (by accessing and taking further Priority 1 datasets forward). The 'areas' for which gaps were identified and the 'action' taken to fill these gaps are described in Table 3.

Note: this was only a 'high-level' gap filling activity and it is recommended that in the future the Environment Agency undertake a more detailed investigation of the datasets from the AMR report (Environment Agency, 2020) that have not been processed as part of this and the previous project (Environment Agency, 2021). It is plausible that in some instances datasets that were previously found to be unavailable may now have become accessible or an accessible alternative may exist (upon investigation).

Table 3 – High-level gap filling

Area for which gap identified	Action
Composting	A list of composting sites[15] (in England) dataset was sourced[16] and processed. Note: It is possible that data (such as bioaerosol emissions) may be available (from Williams et al. 2019 and Health and Safety Executive, 2010; datasets ID45 and ID47 from AMR report, Environment Agency, 2020) and could be included in the attribute table for composting sites. It is recommended that this is explored as part of future work.
Aquaculture	'Consented Discharges to Controlled Waters with Conditions' dataset (Environment Agency, 2022) was accessed and data for active discharge consents categorised as 'Fish+Aquaculture/Fish Farm/Cress Farm' was processed. Note: It is possible that additional data from the 'Consented Discharges to Controlled Waters with Conditions' dataset could be processed for inclusion in the AMR geodatabases; in particular, sewage discharges (water company and not water company) in line with effluent discharges 'area' for which gap identified below.
Wild swimming	The Wildswim.com map plotted a crowd-sourced dataset of 'wild swimming' locations and was identified as being unavailable in the previous project (Environment Agency, 2021), though it was noted that the Open Swimming Society were reviewing licence conditions to see if the dataset could be available for research projects. Further investigation (as part of this project) found that in June 2020 the Outdoor Swimming Society took down wildswim.com over concerns about popular bathing spots being overwhelmed (The Guardian, 2020); though wildswim.co.uk now exists (Wild Swimming, n.d.) raw data were not found to be accessible and although geospatial data for some locations are reported by region on the website the map itself still can't be loaded. Future exploration (in particular contact with the Outdoor Swimming Society) is recommended to fill this data gap.

[15] Geocoded and operating at some point during 2005 - 2014

[16] Received in an email from Wiebke Schmidt on 30/11/2021 (who was provided with this data by Philippa Douglas, UK Health Security Agency)

Area for which gap identified	Action
Effluent discharges	It was noted the Environment Agency stores daily discharge time series for >3100 effluent discharge points (many sites have data from 2005 to present) and it is suggested that the daily mean discharge is provided to the Environment Agency and stored in their WISKI hydrometric archive[17]. It is also possible that discharges may be monitored by water companies at higher resolution[17]. Should this data have been made available (daily or higher resolution), it would not have been achievable to process this within the timeframe of this project. As such, it is recommended that this data is accessed and processed for inclusion in the AMR geodatabases in the future.

[17] Source - Excel spreadsheet that identified the relevant AMR datasets in the previously mentioned AMR report (Environment Agency, 2020)

Data processing and geodatabase updates

Datasets identified for processing were downloaded and saved in a 'Raw data repository' (more details on which are provided in the next section)[18]. The data formatting varied across the datasets that were processed and a full list of data types found are shown in Table 4.

Table 4 – Description of data formats encountered in the raw data

Data type	Description
Spatial data	
Access	Microsoft Access database proprietary filetype
CSV file	Comma-separated value text file
Shapefile	Open geospatial vector data format for geographic information system software
FGDB	ESRI's proprietary format, a file geodatabase is a collection of files in a folder that can store, query, and manage spatial and nonspatial data
Raster	Pixelated (or gridded) data where each pixel is associated with a specific geographical location (Data Carpentry, n.d.)
XLSX	Standard extension for the modern Microsoft Excel spreadsheet files
GPKG	A GeoPackage (with '.gpkg' extension) is an SQLite Database file (Open Geospatial Consortium, n.d.)
Non-spatial data	
PDF	Portable Document Format (PDF), standardized as ISO 32000, is a file format developed by Adobe in 1993 to present documents
ODS	Open source spreadsheet format, commonly created by Calc, a spreadsheet program included in the Apache OpenOffice suite

Non-spatial datasets were in PDF or ODS format (Table 4) and were saved within a separate 'Non-spatial data repository' (more details are provided in the next section).

Datasets with spatial data were transformed and included within the AMR geodatabases created as part of the previous project (see 'Background' section above for more detail). Datasets included in the geodatabases were saved with their ID number from the Excel

[18] Note that the complete 'OS VectorMap District' and 'OS MasterMap® Highways Network – Paths' datasets are included in the 'Raw data repository'

spreadsheet that identified the list of relevant AMR datasets (from the previously mentioned AMR report; Environment Agency, 2020) as a prefix, for example, 'ID60_Active_EPR_Ind'. Some datasets (when accessed and downloaded) included more than one feature class and/ or associated table; these were added to the geodatabases where considered appropriate[19] under the 'existing' dataset ID (for example, table 'ID60_Active_EPR_Ind_ASR' was added as it contains a description of the activity relevant to sites with a permit number for which geospatial data is included within the feature class, 'ID60_Active_EPR_Ind'). It is recommended that in the future the Environment Agency review the feature classes and tables included within the AMR geodatabases (in this and the previous project; Environment Agency, 2021) as it may be possible to revise the content and structure of the geodatabases to better suit their needs, dependent on how it is intended that the AMR geodatabases will be used in the future.

The AMR geodatabases are in FGDB format and consist of a collection of files in a folder that can store, query, and manage spatial and non-spatial data. This is the most common type of geodatabase, which is compatible with any of ESRI's products. Datasets were converted into FGDB format and were added to one of the four AMR geodatabases (Environment, Land Cover, Anthropogenic, Regulatory) depending on the 'theme' which describes the dataset[20]. The general approach to data conversion was to transform the data from its original tabular form and convert it to a table in the FGDB with its corresponding spatial representation, if any. When a dataset was provided in a format that included a geographical coordinate system, this has been maintained. Otherwise, National British Grid has been used.

Three methods were used to transform the data:

1) ArcGIS Pro in-built tools to read shapefile, CSV, GPKG and FGDB;
2) In the case of Access and Excel files, these were opened using their respective proprietary software and exported or saved to CSV format, respectively. Geospatial data was converted to a feature class using the 'XY Table to Point' in-built ArcGIS Pro tool prior to import into a geodatabase; and
3) Raster data were loaded into a geodatabase using the 'Raster to Geodatabase' in-built ArcGIS Pro tool.

All the data consisted of official, open-source or licensed datasets. No scripting or heavy data manipulation was necessary to read the data. The quality and formatting of the data corresponded to what was expected of public/ commercial datasets, and typically little

[19] Based on professional opinion

[20] These were assigned to all datasets in the previous project (Environment Agency, 2021). 'New' datasets from Task 1 – 'Event Duration Monitoring - Storm Overflows - 2020 (England and Wales)' and 'WFD Surface Water Operational Catchments Cycle 2' were themed as Anthropogenic and Environment, respectively.

formatting correction or data manipulation was needed. The most notable exceptions, where 'dataset-specific' processing was required, are included in Table 5.

Table 5 – Dataset-specific processing

Dataset name (and ID[2])	Dataset-specific processing
OS VectorMap District – Roads (ID07)	Only data for 'Roads' located across (at least) the full extent of England were included in the Anthropogenic geodatabase[21] (reason why is detailed in 'Priority 2 and Priority 3 datasets' sub-section above). The road shapefiles that were downloaded (for, at least, the full extent of England) were merged into a single feature class using the in-built ArcGIS Pro 'Merge' tool.
LCMS Target and Non -Targeted Screening (ID31)	Annual average concentrations for 2020 were calculated[22] for substances with data available (within the 'LCMS Target and Non-Targeted Screening' Excel spreadsheet downloaded) and values were included in the attribute table alongside coordinates and sample location which were taken from the Excel spreadsheet (reasoning as to why the complete dataset was not included is detailed in 'Priority 2 and Priority 3 datasets' sub-section above).
Environmental Permitting Regulations – Waste Sites (ID53)	Sites where waste licence status was defined as 'Closure', 'Expired', 'Surrendered' or 'Revoked' and those for which no coordinates were provided (in the 'EPR_waste' table in 'EPR_Waste' Access database downloaded) were not included in the Regulatory geodatabase. Description of licence type ('lic_LType' table in 'EPR_Waste' Access database downloaded) associated with sites was however included in the attribute table.
Waste Data Interrogator 2017 (ID58)	For Incinerator data (waste returns and waste removed[23]), the Batch Convertor Tool (UK Grid Reference Finder, n.d.) was used to convert National Grid References provided into X and Y coordinates so that the data could be converted to a feature class using the 'XY Table to Point' in-built ArcGIS Pro tool.

[21] Data for extent of Great Britain can be found in the 'Raw data repository' (see next section for more detail)

[22] Where recorded concentration was limit of detection it was divided by 2 and no minimum number of values was required in the calculation of annual average concentrations.

[23] Included within the '2017_Incinerator_Waste_Returns_with_waste_removed (B)' Excel spreadsheet downloaded

Dataset name (and ID[2])	Dataset-specific processing
Environmental Permitting Regulations – Industrial Sites (ID60)	Data included within the Regulatory geodatabase is for 'active' sites with permit numbers (i.e. data from 'Active_EPR_Ind' table in 'EPR_Industry' Access database downloaded rather than 'All_EPR_Ind' table which was not included in the geodatabase). Activity description relevant to sites with a permit number (in 'Permit_Num' field) for which geospatial data is included within the feature class 'ID60_Active_EPR_Ind' is included in table 'ID60_Active_EPR_Ind_ASR' (data from 'Active_EPR_Ind_ASR' table in in 'EPR_Industry' Access database downloaded) in the Regulatory geodatabase.
Consented Discharges to Controlled Waters with Conditions – Fish + Aquaculture/Fish Farm/Cress Farm (ID73)	Data included within the Environment geodatabase is that for active discharge consents (i.e. from 'consents_active' table in 'Consented Discharges to Controlled Waters with Conditions' Access database downloaded rather than 'consents_all' table which was not included in the geodatabase) that are categorised as Fish + Aquaculture/Fish Farm/Cress Farm. The Batch Convertor Tool (UK Grid Reference Finder, n.d.) was used to convert National Grid References provided (for Outlet) into X and Y coordinates so that the data could be converted to a feature class using the 'XY Table to Point' in-built ArcGIS Pro tool.

Once converted into FGDB format, a check has been carried out to ensure each of the datasets could be opened and contained the correct formatting. In those cases where the original format was suited to geospatial data (i.e. shapefile and FGDB), ArcGIS Pro was able to recognise the coordinate system and create a geometry ('Shape') field in the underlaying dataset. For CSV files coordinates were encoded in separate fields (labelled X and Y) so that the data could be converted to a feature class using the 'XY Table to Point' in-built ArcGIS Pro tool and, upon conversion, a geometry ('Shape') field was created in ArcGIS Pro.

Deliverables

The deliverables from Task 2, which are described in more detail below, include:

- Raw data repository which includes all information that has been downloaded for a dataset;
- Processed data including geospatial data within revised AMR geodatabases and Non-spatial data; and
- Data register updated (from previous project; Environment Agency, 2021) with datasets processed in this study.

The number that identifies a dataset within each deliverable is the same code used in the Excel spreadsheet that identified the relevant AMR datasets in the previously mentioned AMR report (Environment Agency, 2020). The two 'new' datasets that were utilised in Task 1 of the study ('Event Duration Monitoring - Storm Overflows - 2020 (England and Wales)' and 'WFD Surface Water Operational Catchments Cycle 2') and were processed and

included in the geodatabases were also assigned a number (ID80 and ID81[24], respectively) for use as identity.

Raw data repository

Raw data repository contains individual folders with all the files that have been found per dataset. Each folder is saved as the number which identifies a dataset (described above). The purpose of the raw data repository is to ensure that all information that has been downloaded for a dataset is adequately logged, which in turn ensures the reproducibility of the data processing and serves as a back-up in case the databases were lost.

Processed data

Spatial data (file geodatabases) repository

Spatial data (file geodatabases) repository contains revised versions (16/12/2021) of the four AMR geodatabases (Environment, Land Cover, Anthropogenic, Regulatory) that were previously developed (see 'Background' section for more detail). The revised versions include the geospatial datasets (and any associated feature classes/ tables) included within this study.

Updated webmaps

Datasets from the spatial data (file geodatabases) repository (see above) have been added, as appropriate, to the Phase 1 data viewer webmap (see Environment Agency, 2021) that is now accessible to all Environment Agency staff. For access details see Environment Agency (2021) report. A record of which datasets have been displayed online is included within the data register, see below.

Non-spatial data repository

Non-spatial data repository contains individual folders that contain documents without spatial data that have been found per dataset. Each individual folder is saved as the number which identifies a dataset (described above). It is recommended that the folders containing non-spatial data delivered to the Environment Agency as part of the previous project (Environment Agency, 2021) are added to this repository to create a 'central' location where all non-spatial datasets are collated.

[24] Note that 'WFD Surface Water Operational Catchments Cycle 2' dataset is included 'within' dataset ID81 (as numerous datasets were available from the link at which this dataset was sourced and several were processed; see 'Datasets from Task 1' sub-section for more detail) which was given the 'collective' name 'Environment Agency Catchment Data (England)'.

It was noted that a more recent UK Veterinary Antibiotic Resistance and Sales Surveillance Report (VARSS) report (UK-VARSS, 2021) than that included in the previous project non-spatial folder (sub-folder '71 – Veterinary'; delivered to the Environment Agency) was available. This has been included in the non-spatial data repository delivered for this project and the data register (described below has been updated accordingly). It is recommended that this file is included in a 'central' repository folder for the associated dataset ID (71).

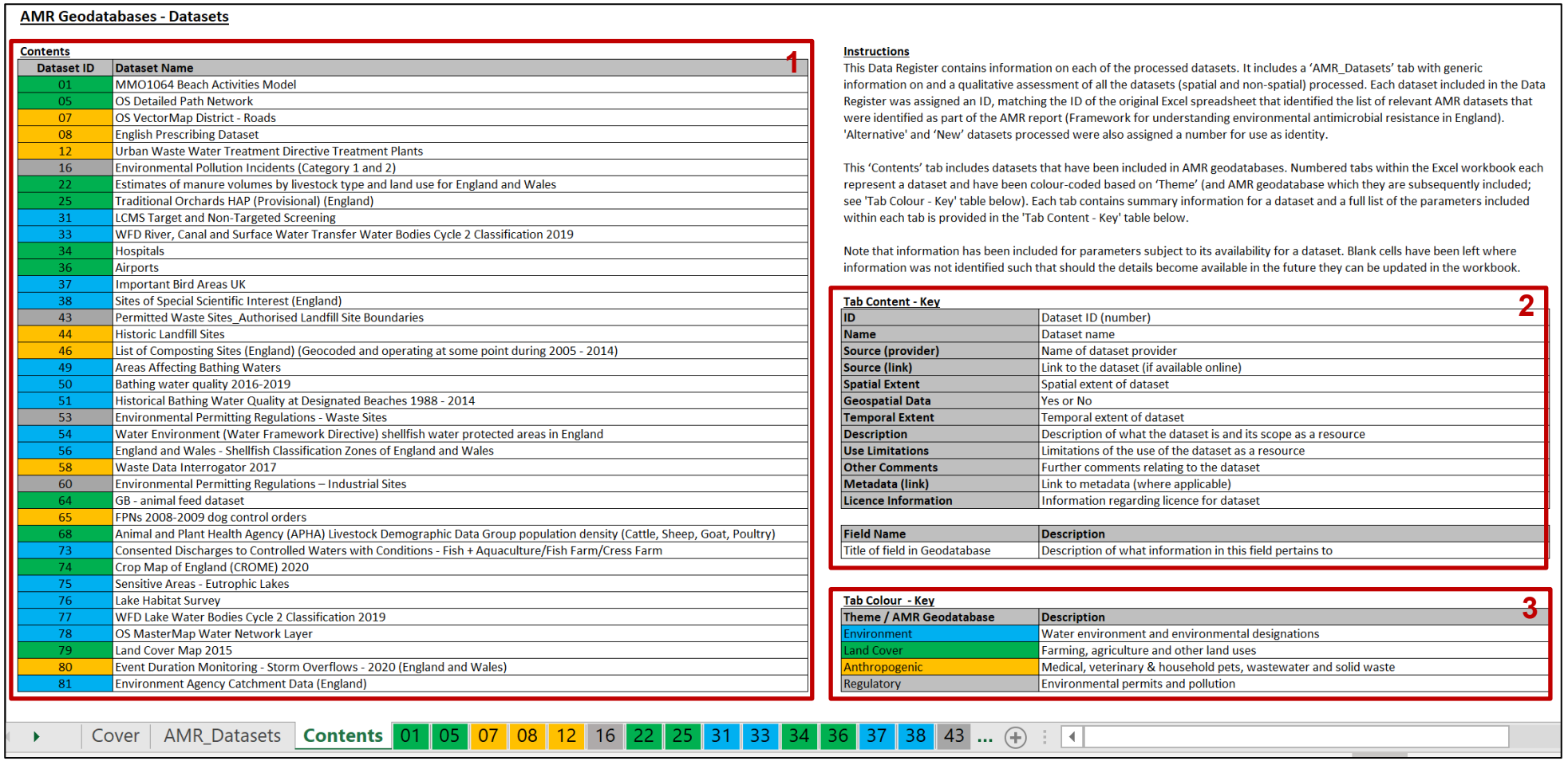
Note: spatial data collected when compiling the more recent VARSS report may be provided in aggregated form in the future.

Data register

Data register is an Excel workbook that contains information on each of the processed datasets. An 'original' version was created during the previous project (Environment Agency, 2021) and a revised version (v0.3) has been delivered as part of this project. The data register includes an 'AMR_Datasets' tab with generic information on all the datasets (spatial and non-spatial) processed, a qualitative assessment (completed as part of Task 3, see next section for more detail) and a record of the datasets included on the webmap (see section above for more detail).

The 'Contents' tab (a screenshot of which is shown in Figure 8 –) contains the datasets that are included in the revised versions (16/12/2021) of the AMR geodatabases (Figure 8 – , Box 1). Numbered tabs within the Excel workbook each represent a dataset and have been colour-coded based on 'Theme' (and the geodatabase in which they have been subsequently included; see Figure 8 – , Box 2 for tab colour key). Each tab contains summary information for a dataset; this includes source, a description of the dataset and the names and descriptions of fields included in the geodatabase tables. For some datasets the 'Other comments' parameter reports additional datasets that were noted to be available (in the metadata for the existing dataset) which may complement the data already included in or could possibly enhance the geodatabase. It is recommended that these are explored as possible additions to the AMR geodatabases in the future. A full list of the parameters included within each tab is provided in the 'Contents' tab (Figure 8 – , Box 3).

It is important to note that information has been included for parameters subject to its availability for a dataset. As in the previous version of the data register, blank cells have been left where information was not identified in this study such that should the details become available in the future they can be updated in the workbook.



AMR Geodatabases - Datasets

Contents 1

Dataset ID	Dataset Name
01	MMO1064 Beach Activities Model
05	OS Detailed Path Network
07	OS VectorMap District - Roads
08	English Prescribing Dataset
12	Urban Waste Water Treatment Directive Treatment Plants
16	Environmental Pollution Incidents (Category 1 and 2)
22	Estimates of manure volumes by livestock type and land use for England and Wales
25	Traditional Orchards HAP (Provisional) (England)
31	LCMS Target and Non-Targeted Screening
33	WFD River, Canal and Surface Water Transfer Water Bodies Cycle 2 Classification 2019
34	Hospitals
36	Airports
37	Important Bird Areas UK
38	Sites of Special Scientific Interest (England)
43	Permitted Waste Sites_Authorised Landfill Site Boundaries
44	Historic Landfill Sites
46	List of Composting Sites (England) (Geocoded and operating at some point during 2005 - 2014)
49	Areas Affecting Bathing Waters
50	Bathing water quality 2016-2019
51	Historical Bathing Water Quality at Designated Beaches 1988 - 2014
53	Environmental Permitting Regulations - Waste Sites
54	Water Environment (Water Framework Directive) shellfish water protected areas in England
56	England and Wales - Shellfish Classification Zones of England and Wales
58	Waste Data Interrogator 2017
60	Environmental Permitting Regulations – Industrial Sites
64	GB - animal feed dataset
65	FPNs 2008-2009 dog control orders
68	Animal and Plant Health Agency (APHA) Livestock Demographic Data Group population density (Cattle, Sheep, Goat, Poultry)
73	Consented Discharges to Controlled Waters with Conditions - Fish + Aquaculture/Fish Farm/Cress Farm
74	Crop Map of England (CROME) 2020
75	Sensitive Areas - Eutrophic Lakes
76	Lake Habitat Survey
77	WFD Lake Water Bodies Cycle 2 Classification 2019
78	OS MasterMap Water Network Layer
79	Land Cover Map 2015
80	Event Duration Monitoring - Storm Overflows - 2020 (England and Wales)
81	Environment Agency Catchment Data (England)

Instructions

This Data Register contains information on each of the processed datasets. It includes a 'AMR_Datasets' tab with generic information on and a qualitative assessment of all the datasets (spatial and non-spatial) processed. Each dataset included in the Data Register was assigned an ID, matching the ID of the original Excel spreadsheet that identified the list of relevant AMR datasets that were identified as part of the AMR report (Framework for understanding environmental antimicrobial resistance in England). 'Alternative' and 'New' datasets processed were also assigned a number for use as identity.

This 'Contents' tab includes datasets that have been included in AMR geodatabases. Numbered tabs within the Excel workbook each represent a dataset and have been colour-coded based on 'Theme' (and AMR geodatabase which they are subsequently included; see 'Tab Colour - Key' table below). Each tab contains summary information for a dataset and a full list of the parameters included within each tab is provided in the 'Tab Content - Key' table below.

Note that information has been included for parameters subject to its availability for a dataset. Blank cells have been left where information was not identified such that should the details become available in the future they can be updated in the workbook.

Tab Content - Key 2

ID	Dataset ID (number)
Name	Dataset name
Source (provider)	Name of dataset provider
Source (link)	Link to the dataset (if available online)
Spatial Extent	Spatial extent of dataset
Geospatial Data	Yes or No
Temporal Extent	Temporal extent of dataset
Description	Description of what the dataset is and its scope as a resource
Use Limitations	Limitations of the use of the dataset as a resource
Other Comments	Further comments relating to the dataset
Metadata (link)	Link to metadata (where applicable)
Licence Information	Information regarding licence for dataset

Field Name	**Description**
Title of field in Geodatabase	Description of what information in this field pertains to

Tab Colour - Key 3

Theme / AMR Geodatabase	**Description**
Environment	Water environment and environmental designations
Land Cover	Farming, agriculture and other land uses
Anthropogenic	Medical, veterinary & household pets, wastewater and solid waste
Regulatory	Environmental permits and pollution

Cover | AMR_Datasets | Contents | 01 | 05 | 07 | 08 | 12 | 16 | 22 | 25 | 31 | 33 | 34 | 36 | 37 | 38 | 43 | ...

Figure 8 – Data register (version 0.3) 'Contents' tab (screenshot). ***Box 1: Datasets that are included in the AMR geodatabases (revised versions, 16/12/2021), Box 2: Tab colour key, Box 3: Full list of parameters included in each 'numbered' tab.***

Task 3: Assessment of data gaps

Background

The following aspects are included in this section:

- Background and objective of the task; and
- Deliverables including qualitative assessment of datasets and high priority gaps table.

The objectives in this task were to:

- Undertake analysis of the datasets identified in the data registry in Task 2 for their spatial and temporal completeness.
- Of these datasets, create a table of high priority datasets that are critical to this research area that have clear spatial and/ or temporal gaps.
- Provide a proposed method of how to best update the datasets in the high priority gaps table (e.g., data collection, gap fill modelling).

Two tables were produced as deliverables for Task 3: “Qualitative Assessment” for completeness (included within the data register, see above) and “High Priority Gaps” (Table 6, within this report) which included the proposed gap filling solution.

Qualitative Assessment table

The datasets from the data register in Task 2 were investigated for temporal and spatial completeness with respect to their application as a data layer in an AMR risk/ hazard map. Information was collected on each dataset’s spatial coverage, the last date they were updated, and how long it had been since they had last been updated (months and years).

Datasets were deemed spatially complete if they were found to span the full extent of England (as deemed appropriate for each dataset). Datasets were deemed temporally complete only if they had been updated since the beginning of 2020. In addition, a narrative assessment was undertaken to provide context to the datasets and the role they play in assessing AMR risk/ hazard, both to the environment and, subsequently, to humans from the environment.

Finally, priority assignments were given according to the confidence that this dataset might play in risk assessments for environmental AMR. The framework rankings were re-assessed based on expert opinion from the previous AMR report (Environment Agency, 2020) and their new scores are noted in column ‘S’ of the data register. Where there are changes in framework ranking scores, these reflect the changes in our perception of the main drivers of AMR in the environment since the publication of the previous report.

Those with clear relevance and utility were given a 1, the highest score. Those datasets that could potentially be of use but of a lower priority were given a score of 2. Finally, those datasets that were unlikely to be of use for informing the AMR risk map were given a score of 3. Only those datasets scored as highest priority (1) were considered for further prioritisation after considering temporal and spatial completeness and suitability to the task of site selection for AMR surveillance in freshwaters.

High priority gaps table

Future efforts to conduct a hazard/ risk assessment for environmental AMR will require access to temporally and spatially complete datasets. The most useful of these datasets for risk analysis should be prioritised for future data collection efforts or, where appropriate, gap filling through modelling. To this end, a high priority gaps table was created to identify these highest priority datasets that are lacking spatially or temporally, but with the greatest utility for risk analysis for AMR. A scoring system was used to articulate those where there is the greatest urgency to fill the gaps (Priority 1), those datasets that are likely to be useful (Priority 2) and those datasets that could help but are lower priority (Priority 3).

We considered three solutions to addressing the spatial and temporal gaps in each of these datasets:

1) mining existing (e.g., unpublished) datasets or newly published datasets;

2) model and extrapolate/ interpolate the spatial gaps in a dataset; and

3) collect new data (e.g., sampling campaign or inquire with stakeholders for updated information as in the water industry's wastewater treatment works).

Deliverables

Qualitative assessment

Qualitative assessment of all datasets identified within the data register (tab 'AMR_Datasets') can be found in columns 'P', 'Q', 'R', 'S', and 'T', alongside the assessment (column 'G') as part of the previous AMR report (Environment Agency, 2020).

All of the 48 datasets identified were spatially complete, i.e. their data covered England. In addition, a number of datasets had data from other geographical locations, including United Kingdom, Wales/ Scotland/ Northern Ireland, and around the world.

There were 25 datasets that were considered temporally "up-to-date" (updated since the beginning of 2020), 21 that were considered "out of date" (last updated prior to 2020) and one that had no information on the date range at which it was conducted.

High priority gap datasets

The high priority gaps table developed is presented below (Table 6) and includes six datasets that were identified as high priority (i.e. Priority 1) and out-of-date (i.e. last updated prior to 2020). The time since the datasets were last updated and the gap filling solutions proposed are also shown in the table. There are no examples of datasets that would easily or even feasibly allow for interpolation. As such, modelling was not a viable solution to any spatial or temporal gaps identified in the high priority datasets. Similarly, there were no datasets that would benefit from having temporal gaps filled via modelling, as most of the utility of these datasets is how they might change over time. We, therefore, have recommended primary data collection as the gap filling solution for these six databases.

Table 6 – High priority datasets with data gaps

Dataset title (unique ID from data register)	Time since last updated	Gap filling solution
MMO1064 Beach Activities Models (ID1)	7 years and 8 months	Primary data will need to be collected to help identify if beach activity levels remain similar to previous described levels (seven years ago).
Urban Waste Water Treatment Directive Treatment Plants (ID12)	5 years and 8 months	Primary data will need to be sought from water companies to determine if new urban wastewater treatment plants have been constructed or upgraded since the data was last updated. However, this is unlikely to have changed drastically in the time since it was last updated.
Bathing water quality (ID50)	2 years	Due to the impact of the COVID-19 pandemic on the sampling programme no classifications were made for the 2020 season. However, the bathing water classification report for 2021 it is expected to be published in early 2022, providing updated status of Intestinal Enterococci and *E. coli* as faecal indicators.
Land Cover Map 2015 (ID79)	6 years	Primary data collection would need to occur for updated information on land cover. Although it is unlikely that there will have been drastic land cover changes in 6 years, continuing urbanisation may have changed boundaries and density of different land covers.

Dataset title (unique ID from data register)	Time since last updated	Gap filling solution
Estimates of manure volumes by livestock type and land use for England and Wales (ID22)	11 years	Estimations could be made from up-to-date data on type/ distribution/ quantity of livestock in England. These datasets do exist (dataset id 68 and 69), however they are also out of date. Primary data collection would need to occur to update these datasets before estimates of manure volume could be calculated.
1999 Environment Agency report on sewage sludge (ID62)	22 years	Primary data collection would need to occur as a result of the age of this report.

NHS temporal data analysis

In the UK, national prescription data is provided by the National Health Service (NHS). This data is freely accessible and consists of individual files for each month. The large files contain over 10 million records every month. The data cannot be used for the direct calculation of the prescription levels of different active pharmaceutical ingredients (APIs). Re-organisation and processing of the files is required before any exploration or analysis and to speed up the data reading.

An R package to calculate and visualise England NHS primary care prescribing data is now available[25,26]. This includes a data analysis tool for the 2015 to 2018 period. The tool can calculate prescribed quantity of an API which could be used to facilitate spatio temporal and long-term prescription trends for wider usage. For example, by enabling calculation of the total prescribed quantity of an API or a group of APIs, specified to a postcode or region. Application of this tool to examine how this data could inform understanding of environmental AMR has not been possible within the scope of this project. It is suggested that this is picked up in future phases of work.

[25] GitHub - PrAnaViz/PrAna: An R package to aggregate and visualize England's prescription data (Accessed 15th December 2021)

[26] 67eca268-a776-48cc-a5e7-290504452f6b.pdf (researchsquare.com) (Accessed 15th December 2021)

Task 4: Mapping data

Background

The primary objective of Task 4 was to produce national maps showing potential AMR exposure at a catchment scale.

For each WFD operational catchment a score for environmental AMR abundance and catchment sensitivity to AMR (i.e. quantifying the scale of potential impact pathways) was calculated using agreed selected priority 1 variables, see Table 7 and Table 8. These were then combined to produce an overall score for AMR exposure for each WFD operational catchment.

Table 7 - Variables used in exposure mapping - abundance

Dataset ID	Environmental abundance of AMR
80	CSO spill count & duration
33	River density (length of WFD river per km^2) - proxy for dilution potential[27]
73[28]	Count of on-site private wastewater treatment discharges in a catchment (Environment Agency active consented discharges)
73	Count of water company wastewater treatment work (WwTW) discharges (Environment Agency active consented discharges)
49/ 50[29]	Proximity of bathing water sites recently failing water quality criteria (is there one within 500 m?)
34	Count of hospitals
22	Estimated Livestock manure and slurry loads

[27] Used instead of dataset ID18 on the recommendation of the EA.

[28] Full dataset not included in the AMR geodatabase but is used in the exposure mapping.

[29] See 'AMR_Exposure_map_workbook_v1.0' for further detail.

Table 8 - Variables used in exposure mapping - sensitivity

Dataset ID	Catchment sensitivity to AMR
54	Shellfish waters (Water Environment (WFD) Regulations 2017 shellfish water protected area designations in England)
N/A[30]	Landcover – (arable landcover from Corine)
73	Aquaculture (Environment Agency active consented discharges)
49	Areas affecting bathing waters (ZOI)

Methodology

To obtain AMR exposure scores for each WFD operational catchment, the priority 1 data from Table 7 and Table 8 were processed in GIS to spatially join each variable to the operational catchments. Each variable was spatially joined in turn to the WFD catchments and saved and stored in a geodatabase as "Exposure_map_vX", with X being a number 1 to 11 denoting the addition of a new variable. On joining, each variable was summarised per catchment and normalised for area, typically using a metric such as count per km^2 or sum per km^2. A more detailed description of the processing carried out for each layer, and where applicable any proxies used or assumptions made, can be found in the 'AMR_Exposure_map_workbook_v1.0'. Once all variables had been processed, the final file "Exposure_map_v11" was exported from GIS to Excel for calculation of the scores.

In Excel, the 0, 20, 40, 60, 80 and 100^{th} percentiles were calculated for each summarised variable. From this, a new column was added adjacent to the variable where each variable was scored 1 to 5 based upon which percentile group the value for each catchment fell within (Table 9 and Table 10). For all variables with the exception of river density, a proxy for stage or dilution potential was measured as length of WFD river per km^2, the higher the number, the higher the risk score. For river density this was reversed as a higher number equated to high dilution potential and therefore lower environmental abundance. All variables were weighted equally, with the exception of the CSO variable, which was weighted by a factor of 2 to align with the logic presented in Task 1, i.e. reflecting its dominant role in controlling the abundance of environmental AMR.

[30] This dataset has not been included in the AMR geodatabases. It is recommended that it should be considered for inclusion in the future.

Table 9 - Scoring based on percentile ranges

Scoring	Percentile
5 (Very High)	80 – 100
4 (High)	60 – 80
3 (Medium)	40 – 60
2 (Low)	20 – 40
1 (Very Low)	0 - 20

Table 10 - Example excerpt of the scoring methodology

Catchment name	**Variable 1:** **No. hospitals/km^2**	**Variable 1 score**	**Variable 2a** **Sum slurry (L/km^2)**	**Variable 2a score** (2a&b weighting: *0.5 as two attributes used from the same dataset)	**Variable 2b Sum manure (kg/km^2)**	**Variable 2b score** (2a&b weighting: *0.5 as two attributes used from the same dataset)	**…Etc. for other variables** (NB CSO weighting *2)	**Abundance score** (sum of variable scores)	**Normalised abundance score** (scored based on percentiles so exposure and abundance given equal weight)
Catchment 1	40	4	600	1.5	955	2.5	-	35	5
Catchment 2	35	3	900	2.5	455	1	-	20	3
Catchment 3	85	5	200	1	300	0.5	-	15	2
Etc.	16	2	50	0.5	800	2	-	20	3

Table note: table contains dummy data to provide processing example

Using the scores for each variable an overall score for environmental abundance and catchment sensitivity were calculated for each catchment. Similarly, to the variables, the scores for environmental abundance and catchment sensitivity were normalised to give a score of 1 (very low) to 5 (very high) for each catchment. An overall AMR exposure score was calculated as the product of the normalised environmental score and the normalised catchment sensitivity score, as below:

Potential AMR exposure = Environmental abundance * Catchment sensitivity

Deliverables

Figure 9 and Figure 10 show environmental abundance and catchment sensitivity, respectively. Table 11 shows the combination of these, the potential AMR exposure score for each catchment. For the overall potential AMR exposure map, the AMR exposure scores relating to the five categories are shown in Table 11. Maps for each of the input variable listed in Table 7 and Table 8 can be found in Appendix A: AMR exposure mapping variables. These maps include coastal and transitional operational WFD catchments.

The results as well as maps of all input variables have been included within the Phase 1 webmaps (see Environment Agency, 2021). Here the input variables can be tracked through for each catchment to see the output sensivity and abundance scores they produce as well as the ultimate catchment exposure output (see Figure 12 and Figure 13).

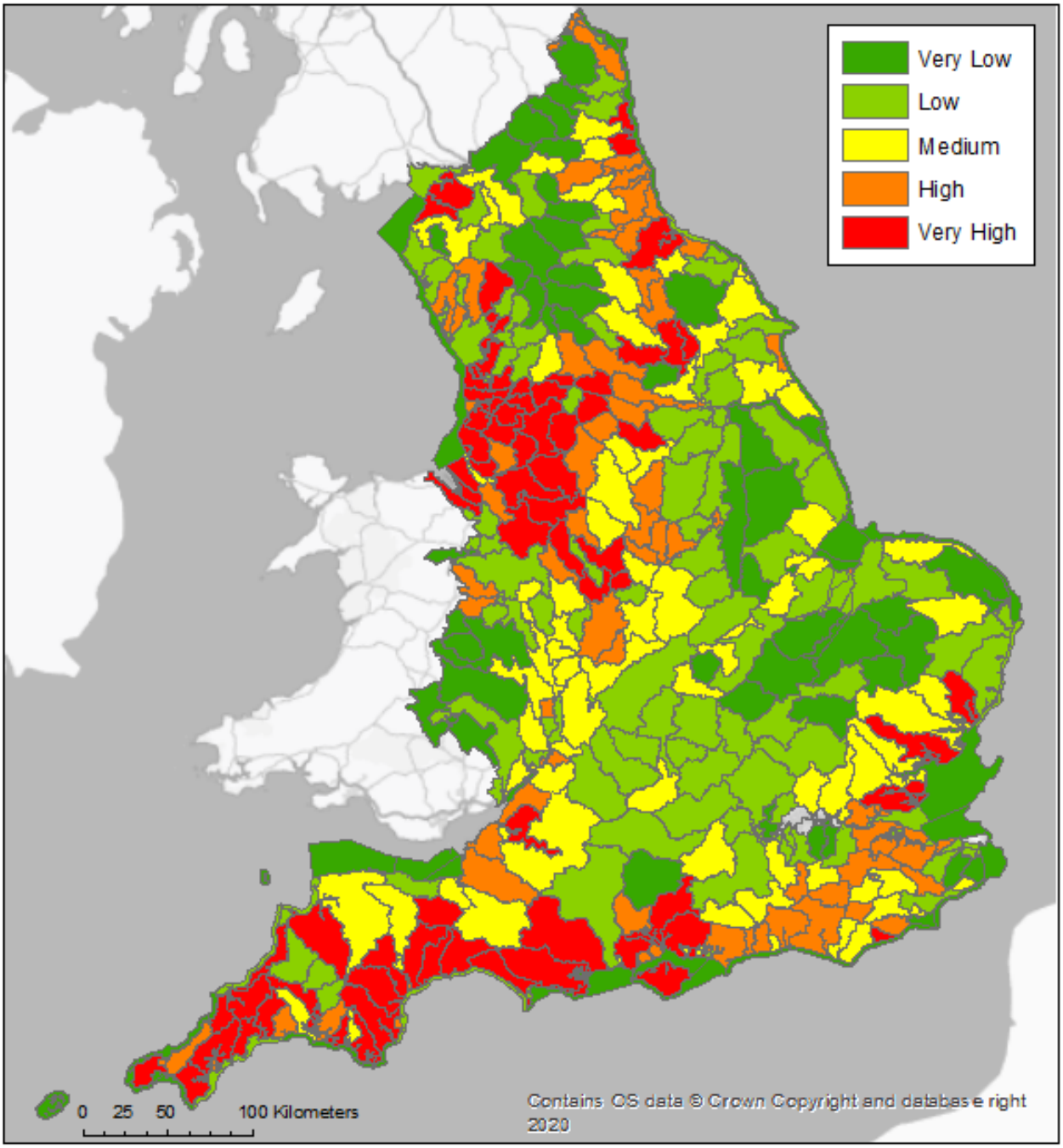


Figure 9 - Environmental abundance map

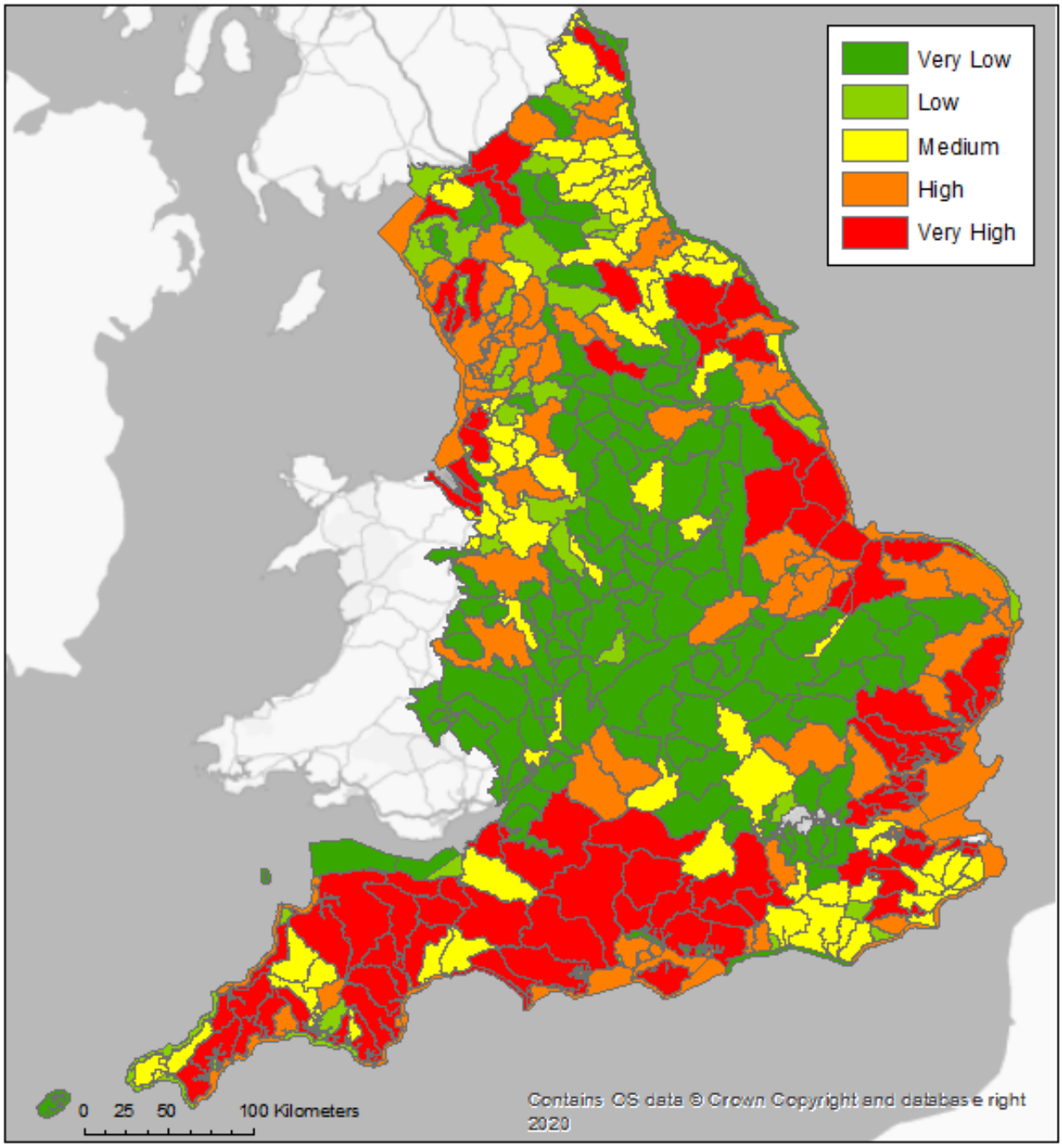


Figure 10 - Catchment sensitivity map

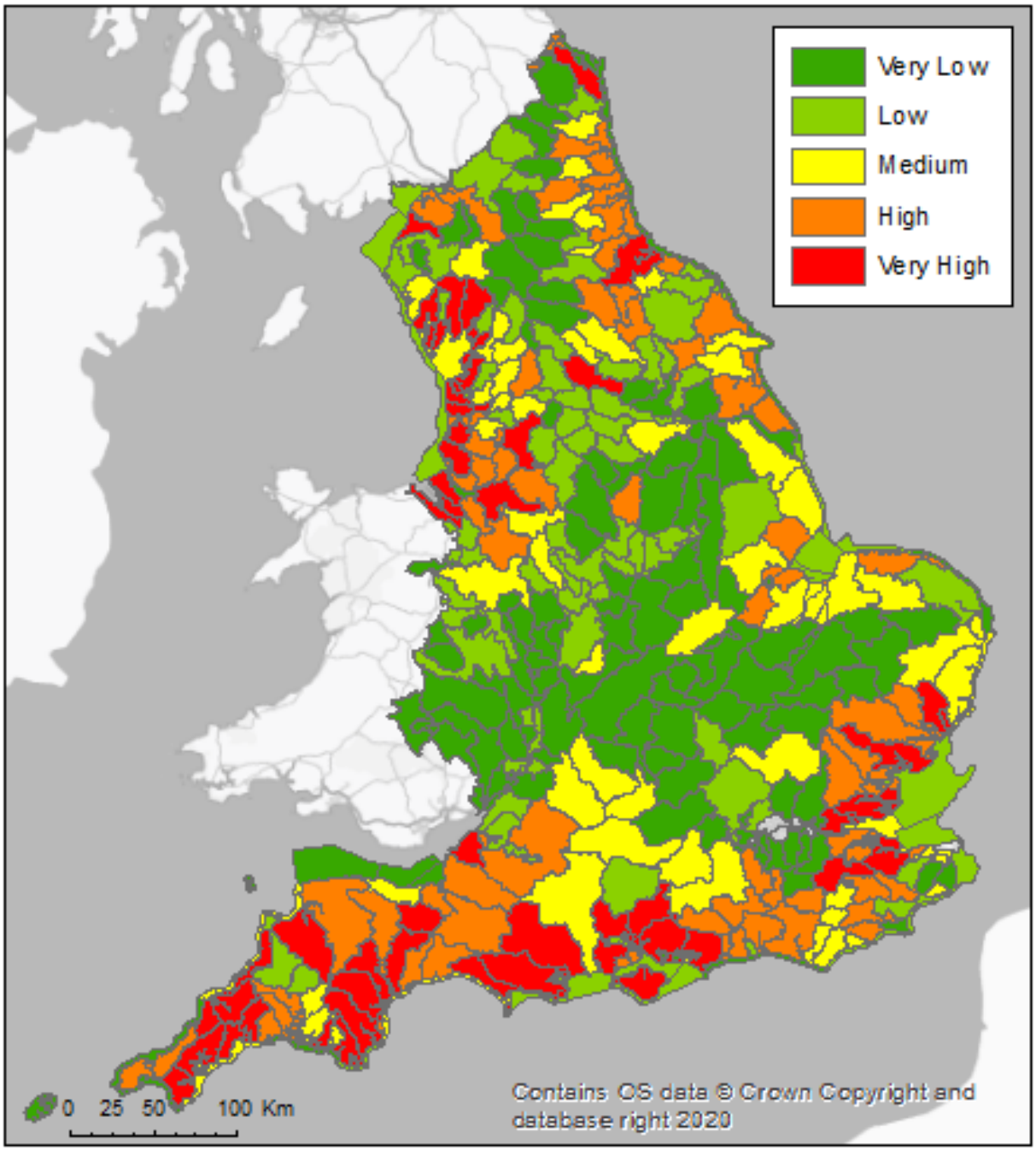


Figure 11 - Potential AMR exposure map

Table 11 - Exposure score and ratings used in AMR exposure map

Exposure rating	Exposure score
Very High	17 - 25
High	11 - 16
Medium	7 - 10
Low	4 - 6
Very Low	1 - 3

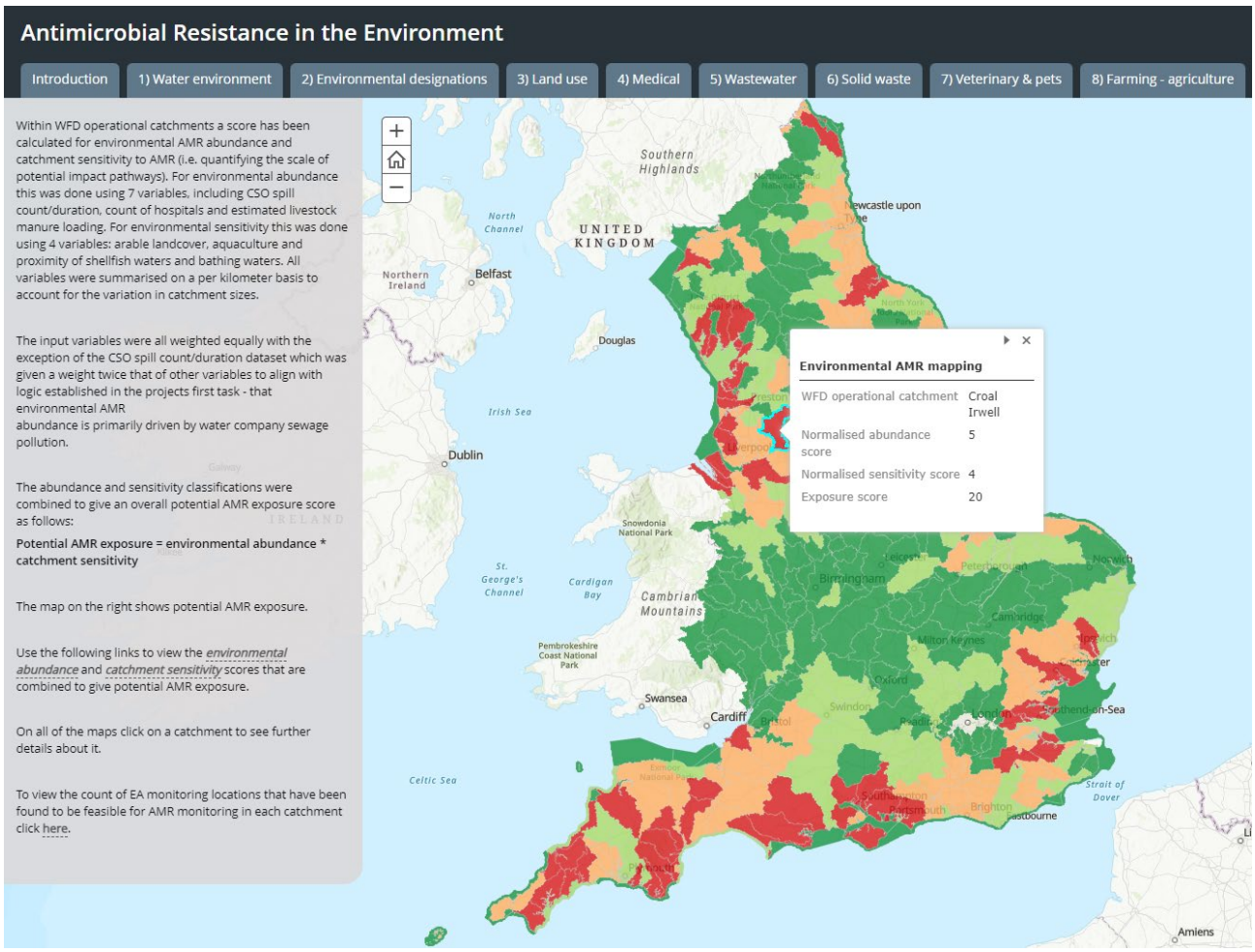


Figure 12 - Potential AMR exposure shown on projects interactive webmap

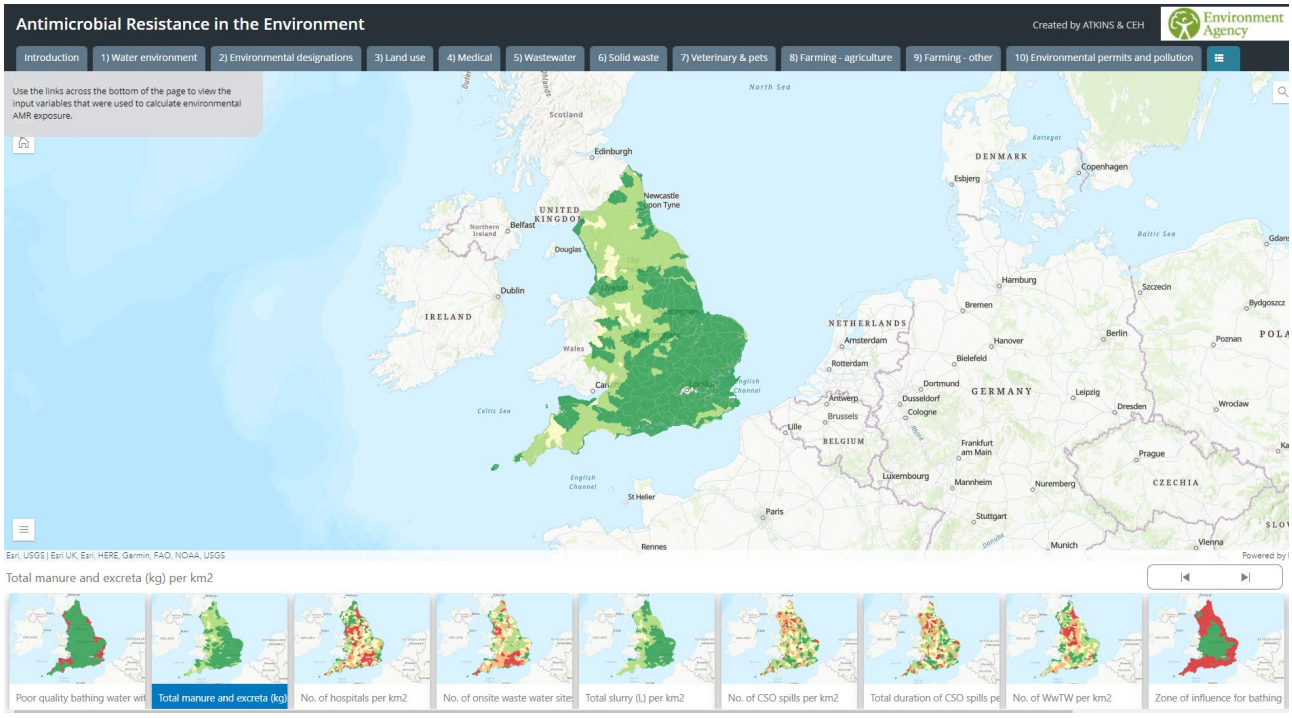


Figure 13 - Exposure variables shown on projects interactive webmap

Observations from a review of the results are provided below:

- AMR abundance is highest along the south coast, in the south west and in the north west (around Cheshire, Staffordshire, Derbyshire and Greater Manchester):
  - High abundance in the south is driven by the large number of on-site private wastewater treatment discharges (from Environment Agency active consented discharges dataset) in these areas as well as the large number of hospitals.
  - In the south west high slurry application is the key driver.
  - High abundance in the north west is driven partly by the CSO dataset and the hospital dataset but primarily by the count of active water company Environment Agency consented WwTWs discharges.
  - It is noted that the hospital dataset shows a high density of hospitals in London which does not come through in the abundance map, with many other variables showing lower scores in that area.
  - The regional trends discussed above are not affected by the locations of failing bathing waters which occur all around the UK with each only affecting a single coastal catchment.
- AMR sensitivity is highest in the south and south west as well as along the east coast:
  - In these areas arable land cover is high, many catchments could affect adjacent bathing waters and shellfish waters.
  - Aquaculture is a key driver for high sensitivity in the south.
- AMR exposure (combining abundance and sensitivity) therefore is highest in the south west with less significant hot spots in the north west and on the eastern coastline.

Limitations

Recognised limitations of the approach are set out below:

- The methodology applied for Task 4 represents one potential approach to AMR mapping. The selected input variables could be combined in different ways to those applied in this report to provide a provisional exposure map. The programme has not allowed for sensitivity testing of the weightings applied. It is suggested this is undertaken in future phases of work.
- Bathing water data are currently used in the calculation of abundance (if a catchment is in proximity to a failing bathing water) as well as for sensitivity (if a catchment is in an ‘area affecting a bathing water’). This, therefore, introduces some bias toward bathing water catchments in the exposure calculation. An alternative approach would be to weight the two bathing water variables by 0.5 so that collectively they had no more control on the output than most other variables. However, given that failing bathing waters are usually a proxy for sewage loading issues and to align with the the logic established in Task 1, that sewage pollution is the primary driver for AMR abundance, it is suggested that if this exercise was redone only the areas affecting a bathing water variable had it’s weighting reduced.

- Due to the variables selected to define sensitivity the maps currently focus on AMR exposure to humans. Additional variables or, perhaps more appropriately, a separate exposure map could be produced to examine broader environmental impacts.
- Further work could be undertaken to refine how the data for each variable is used, for example it is noted that not all types of aquaculture have the same AMR sensitivity. With additional time it may be possible with further data processing to separate out those areas that are most sensitive within the dataset and then only use them within the sensitivity calculation.
- Inclusion of additional variables or more detailed data could improve the results. For example, if crop map data was added as an abundance variable then specific crops associated with antimicrobial pesticides that are suspected to enhance antimicrobial resistant could be identified.
- It is noted that two of the four sensitivity variables are coastal datasets (bathing waters and shellfish waters). The sensitivity maps therefore reflect a strong coastal trend, introducing a bias in the output that should be examined in future work.
- It would be possible to use the same, or a refined method, to produce more granular results, e.g. at a sub-catchment level.

Review of site selection

A review of the site selection from Task 1 was undertaken with consideration of the exposure map developed in Task 4. Catchments with the highest AMR exposure scores in Task 4 were investigated in the same way as in Task 1.

First, the operational catchments identified with the highest AMR exposure scores were ranked based on the number of feasible Environment Agency sampling sites within them (data provided by Environment Agency based on sites practicality for AMR sampling). If catchments had under five potential sampling sites, they were removed from the site selection. The outcome is shown in Table 12.

Table 12 – Catchments with Very High AMR exposure scores and >5 feasible Environment Agency monitoring sites

WFD Operational Catchment	Count of feasible Environment Agency monitoring sites in catchment
East Hampshire Rivers	44
Isle of Wight Rivers	33
Itchen	33
Camel	21
Ellen and West Coast	16
Stour Dorset	15
Poole Harbour Rivers	13
Fal	11
Deben	10
Colne Essex	9
Fowey	9
Strat Neet and North Coast Streams	9
Torridge	9
Creedy and West Exe	8
Tone	8
Crouch and Roach	7
West Dorset Rivers	7
Alt	6
Cober and Lizard	5

There were 19 catchments with high AMR exposure and five or more feasible Environment Agency sampling sites. These 19 catchments were investigated to determine how many of these feasible Environment Agency sampling sites were less than 1 km downstream of a high risk CSO site (>100 spills per year). Three catchments satisfied this criterion: Ellen and West Coast (also identified in Task 1); Fal; and Strat Need and North Coast Streams. Ellen and West Coast had three sites that satisfied this criteria; Fal, Strat Need and North Coast Streams catchments had one.

Recommendation

A summary of the findings from Task 1 and Task 4 with regard to catchment selection for AMR monitoring is presented in Table 13.

As the Ellen and West Coast catchment was identified as being suitable for AMR monitoring in both Task 1 and Task 4 it is recommended that it is taken forward by the Environment Agency. It contains 16 feasible Environment Agency monitoring sites, three of which are <1 km downstream of a high risk CSO site (>100 spills per year). It also has very high AMR abundance, sensitivity, and exposure scores.

For undertaking surveillance in an additional catchment, Table 13 provides five options. Discretion of the Environment Agency should be used to choose another suitable catchment for AMR surveillance from these. Based on the finding of this report it is recommended that the Croal Irwell catchment is selected. It contains more feasible Environment Agency monitoring locations than any of the other short listed catchments in Table 13, four of which are <1 km downstream of a high risk CSO site (>100 spills per year) – again more than any of the other short listed catchments in Table 13. Additionally, it was identified as having a very high AMR exposure score. It is noted that its AMR sensitivity (high) is not as large as some of the other catchments (very high). This is because it is not in proximity to a shellfish water and has a lower than average agricultural landcover. It does however feed into a bathing water and contain more than the average number of aquaculture discharges.

There are four alternatives to selecting the Croal Irwell as the second monitoring catchment, listed in Table 13. Given that the Goyt Etherow Tame and Nottingham Urban catchment have low exposure scores and very low sensitivity it is not recommended that these are considered. This leaves the Fal catchment and the Strat Neet and North Coast Streams catchment. These both have very high exposure scores, however both catchments only have one feasible Environment Agency monitoring site within them that is <1 km downstream of a high risk CSO. If one is considered sufficient then these may be considered as alternatives to the Croal Irwell.

Table 13 – Review of catchment selection for AMR monitoring.

Catchment	**Count of feasible Environment Agency monitoring sites in catchment**	**Count of feasible Environment Agency sampling sites <1km downstream of a high risk CSO site (>100 spills per year)**	**Identified by Task 1**	**Identified by Task 4**	**Abundance** score (normalised score from Task 4)	**Sensitivity** score (normalised score from Task 4)	**Exposure** score (normalised score from Task 4)
Ellen and West Coast	16	3	yes	yes	5 - Very High	5 - Very High	25 - Very High
Croal Irwell	17	4	yes	no	5 - Very High	4 - High	20 - Very High
Goyt Etherow Tame	15	2	yes	no	4 - High	1 - Very Low	4 - Low
Nottingham Urban	13	2	yes	no	4 - High	1 - Very Low	4 - Low
Fal	11	1	no	yes	5 - Very High	5 - Very High	25 - Very High
Strat Neet and North Coast Streams	9	1	no	yes	5 - Very High	5 - Very High	25 - Very High

Summary & conclusion

This project has built on an earlier study (Environment Agency, 2021), which identified and collated a library of datasets that could be useful in understanding AMR in the environment.

This project has completed four key tasks in support of a future Environment Agency AMR surveillance pilot study. Key outcomes from each of these are outlined below:

- Task 1: This project has developed preliminary selection criteria to quickly identify suitable river catchments for environmental AMR surveillance.
- Task 2: This project has extended the existing AMR database in accordance with the project scope.
- Task 3: This project has examined the AMR database for data gaps and suggested how to address them.
- Task 4: This project has shown how the geodatabase can be used for mapping of AMR sensitivity, abundance and exposure. It has also shown how this mapping could be used to support decision making, such as the task 1 selection of river catchments for environmental AMR surveillance. Outputs of the mapping have been provided within this report as well as within the projects online data viewer.

References

CEFAS. (n.d.). Public register of Aquaculture Production Businesses in England and Wales. Available at: https://www.cefas.co.uk/eu-register/ (Accessed 8th December 2021).

Data Carpentry. (n.d.). Introduction to Raster Data. Available at: https://datacarpentry.org/organization-geospatial/01-intro-raster-data/ (Accessed 6th December 2021).

Defra and CEFAS. (2021). Fish, shellfish or crustacean farm authorisation. Available at: https://www.gov.uk/guidance/fish-shellfish-or-crustacean-farm-authorisation#records-you-must-keep (Accessed 1st December 2021).

Defra and Environment Agency. (2021a). Water Quality Archive. Available at: https://environment.data.gov.uk/water-quality/view/explore?search=&area=&samplingPointType.group=&loc=529763%2C164813&_limit=500# (Accessed 8th December 2021).

Defra and Environment Agency. (2021b). Catchment Data Explorer. Available at: https://environment.data.gov.uk/catchment-planning?fbclid=IwAR0jAtZfxIGJmgSNpVEROhj_gHx7Zci8SEPzRPN1Fxc-38yNaGrSjLrugX8 (Accessed 8th December 2021).

Environment Agency. (2020). Framework for understanding environmental antimicrobial resistance in England. Research Contractor: Centre for Ecology and Hydrology. Framework for understanding environmental antimicrobial resistance in England - GOV.UK (www.gov.uk)

Environment Agency. (2021). Data exploration of antimicrobial resistance (AMR) in environmental settings. Contractors: Atkins (lead contractor) and CEH.

Environment Agency. (2022). Consented Discharges to Controlled Waters with Conditions. Available at: https://data.gov.uk/dataset/55b8eaa8-60df-48a8-929a-060891b7a109/consented-discharges-to-controlled-waters-with-conditions#licence-info (Accessed 8th December 2021).

Eramo A, Delos Reyes H, Fahrenfeld NL. Partitioning of Antibiotic Resistance Genes and Fecal Indicators Varies Intra and Inter-Storm during Combined Sewer Overflows. Front Microbiol 2017; 8: 2024

Hammond, P., et al., (2021) Detection of untreated sewage discharges to watercourses using machine learning. npj Clean Water, 2021. 4(18).

Health and Safety Executive. (2010). Bioaerosol emissions from waste composting and the potential for workers' exposure. Available at: https://www.hse.gov.uk/research/rrpdf/rr786.pdf (Accessed 5th December 2021)

Honda R, Tachi C, Yasuda K, et al. Estimated discharge of antibiotic-resistant bacteria from combined sewer overflows of urban sewage system. npj Clean Water 2020; 3: 15.

House of Commons Environmental Audit Committee. (2022). Water quality in rivers. Fourth Report of Session 2021–22. Water Quality in Rivers (parliament.uk) (accessed 28/01/2022).

Open Geospatial Consortium. (n.d.). Getting Started With GeoPackage. Available at: http://www.geopackage.org/guidance/getting-started.html (Accessed 13th December 2021).

Phillips PJ, Chalmers AT, Gray JL, Kolpin DW, Foreman WT, Wall GR. Combined sewer overflows: an environmental source of hormones and wastewater micropollutants. Environ Sci Technol 2012; 46: 5336–43.

Singer, A.C., et al., (2021) ERAMMP Report-55: Evidence Review on the Entry and Spread of Antimicrobial Resistance (AMR) in the Rural Water Environment in Wales.

The Guardian. (2020). Wild swimming site removes online map to ease overcrowding. Available at: https://www.theguardian.com/travel/2020/jun/02/wild-swimming-site-removes-online-map-to-ease-overcrowding (Accessed 26th November 2021).

UKCEH. (n.d.). UK Lakes Portal. Available at: https://eip.ceh.ac.uk/apps/lakes/ (Accessed 1st December 2021).

UK Grid Reference Finder. (n.d.). Batch Convert Tool. Available at: https://gridreferencefinder.com/batchConvert/batchConvert.php (Accessed 6th December 2021).

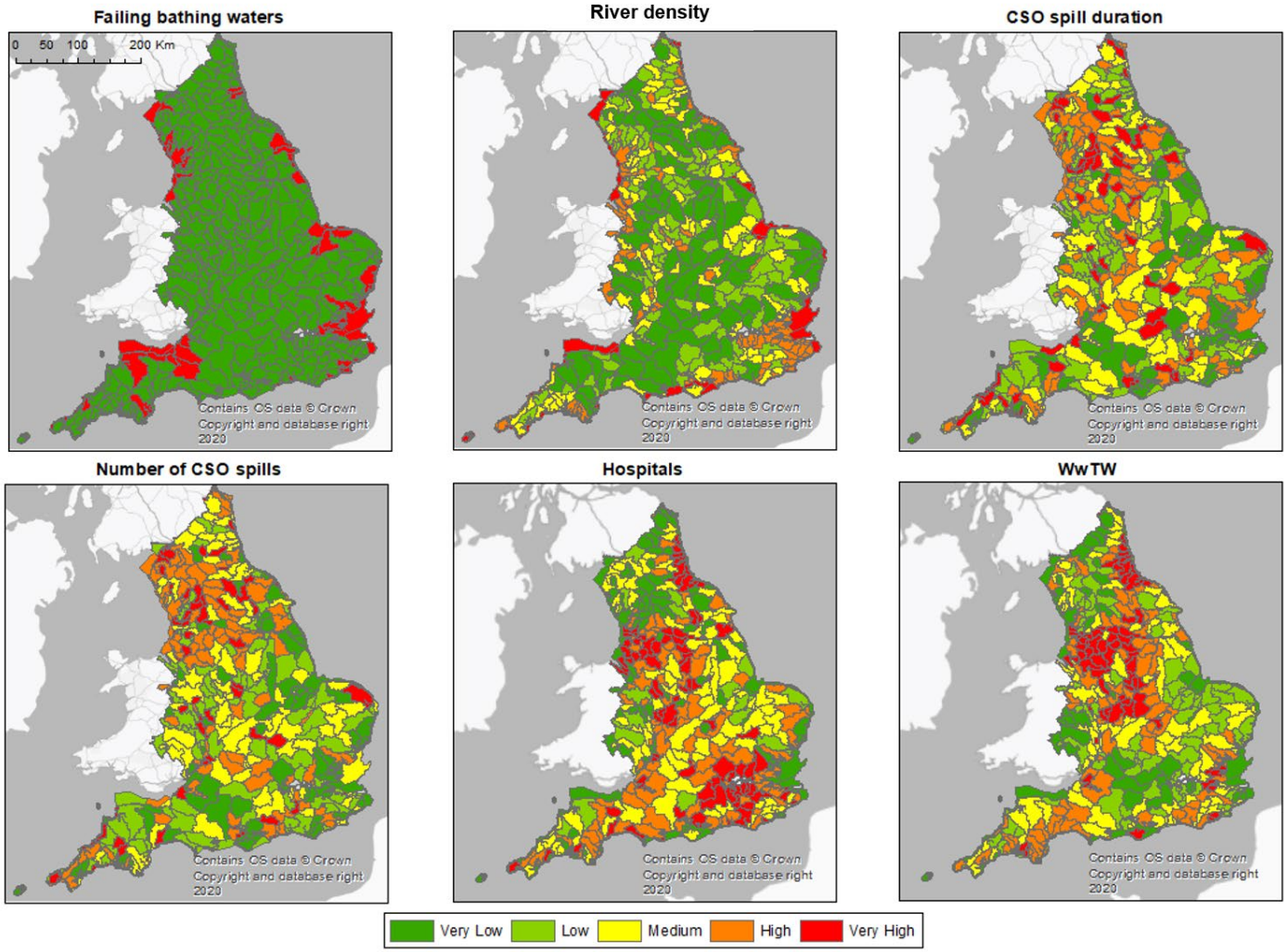
UK-VARSS. (2021). Veterinary Antibiotic Resistance and Sales Surveillance Report (UK-VARSS 2020). Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1033825/VARSS_Main_Report__Final_Accessible_version.pdf (Accessed 9th December 2021).

Wild Swimming. (n.d.). Wild Swim Map UK. Available at: http://www.wildswimming.co.uk/wild-swim-map-uk/?multi_region=wild-swim-map-uk (Accessed 26th November 2021).

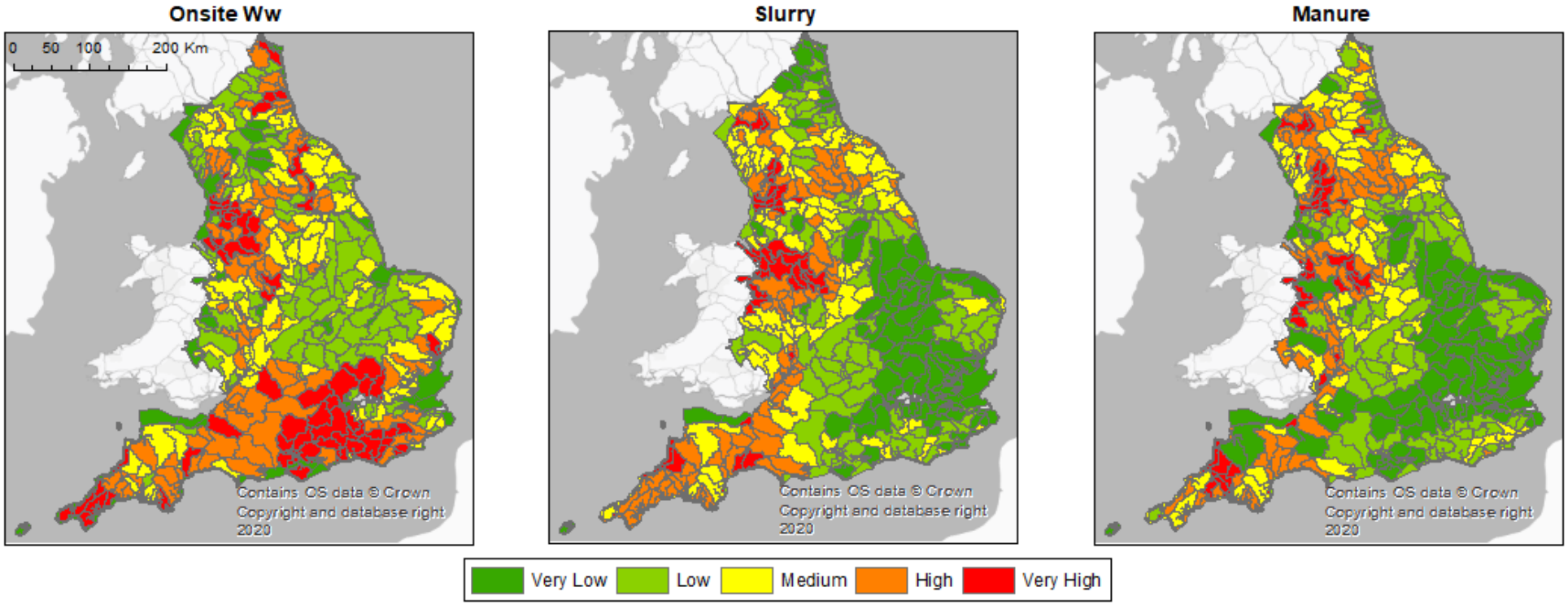
Williams, B., Douglas, P., Barcelo, A. R., Hansell, A. L., & Hayes, E. (2019). Estimating Aspergillus fumigatus exposure from outdoor composting activities in England between 2005 and 14. Waste Management, 84, 235-244.

Appendix A: AMR exposure mapping variables

A1 - Abundance variables

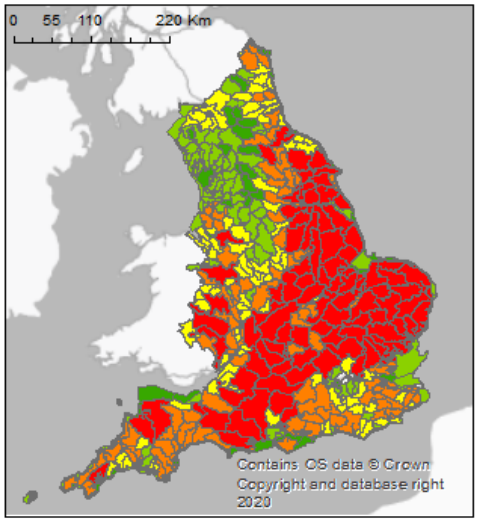


A1 - Abundance variables (continued)

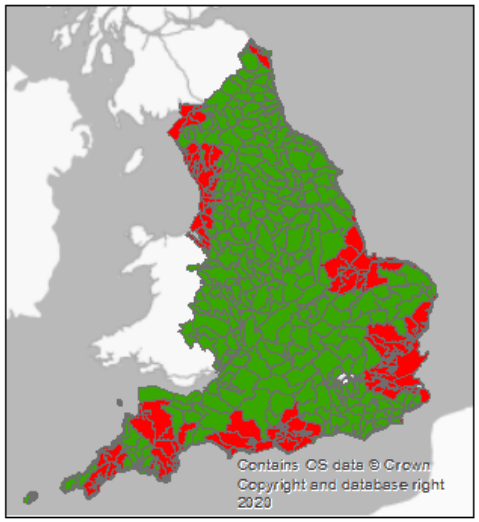


A2 - Sensitivity variables

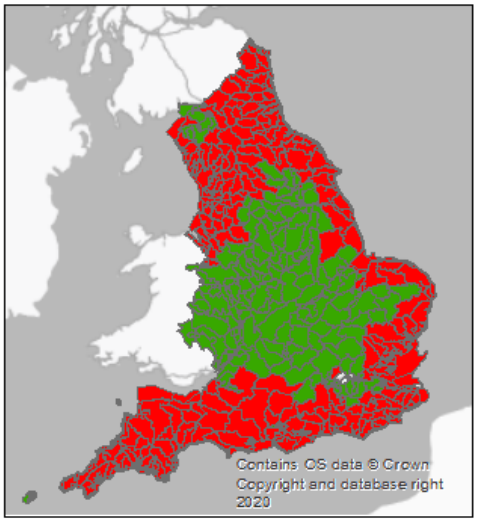
Arable land cover



Shellfish waters

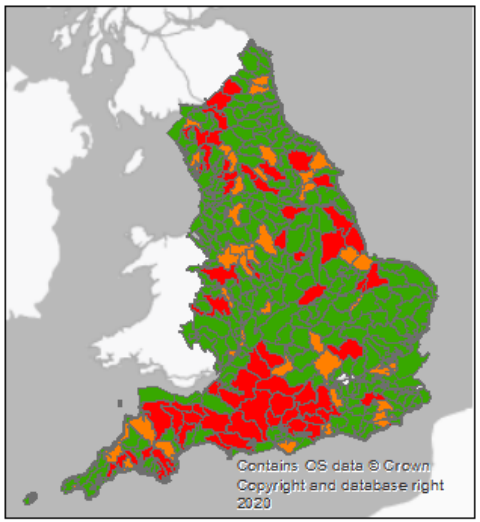


Bathing water ZOI



Note ZOI: zone of influence

Aquaculture



Would you like to find out more about us or your environment?

Then call us on

03708 506 506 (Monday to Friday, 8am to 6pm)

Email: enquiries@environment-agency.gov.uk

Or visit our website

www.gov.uk/environment-agency

incident hotline

0800 807060 **(24 hours)**

floodline

0345 988 1188 **(24 hours)**

Find out about call charges (https://www.gov.uk/call-charges)

Environment first

Are you viewing this onscreen? Please consider the environment and only print if absolutely necessary. If you are reading a paper copy, please don't forget to reuse and recycle.