

# **Defra Open Data Strategy**

**June 2012 – March 2014**

---

---

© Crown copyright 2012

You may re-use this information (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence. To view this licence, visit [www.nationalarchives.gov.uk/doc/open-government-licence/](http://www.nationalarchives.gov.uk/doc/open-government-licence/) or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or e-mail: [psi@nationalarchives.gsi.gov.uk](mailto:psi@nationalarchives.gsi.gov.uk)

Any enquiries regarding this document/publication should be sent to us at:

[defra.helpline@defra.gsi.gov.uk](mailto:defra.helpline@defra.gsi.gov.uk)

PB13785

# Contents

Summary .....	1
Introduction .....	1
Big Data .....	2
Datasets held by the Defra network .....	2
Publication of INSPIRE data .....	3
Widening access to science data .....	4
Publication of other data.....	5
Rationalising and improving efficiency of data collection.....	6
My Data .....	7
Customer Satisfaction and Experience Data .....	7
Creation of Dynamic Information Markets.....	9
Cross cutting initiatives.....	10
Continuous Improvement of the Quality of Data Published.....	11
What data quality means for Defra .....	11
Driving up data quality in Defra .....	12
Statistical data.....	13
Transparency governance and culture change .....	14
Annex: Data to be released as open data up to 31 March 2014	

# Summary

The Defra network has published nearly 400 datasets on data.gov.uk. Defra also leads the UK Location Programme which promotes the publication of location data across the UK in a format that is easily discovered and reusable by others. Over the coming two years the Defra network will publish many more environmental datasets, as required by the EU INSPIRE<sup>1</sup> Directive, as well as other environmental and agricultural data that will be of interest to citizens and business. In addition to publishing a range of corporate data, we will seek to identify and publish customer satisfaction and experience data to help the public and business to judge Defra's performance. We will raise awareness of the data we publish and encourage reuse under an Open Government Licence where appropriate. We will develop existing data quality systems, sharing best practice across the Defra network to drive up the quality of our data, making it easier for citizens and business to understand and reuse.

## Introduction

1. The Defra network collects, creates and exploits a wide range of data in the areas of the natural environment; sustainable development; food, farming and fisheries, animal health and welfare; environmental protection and pollution control; and rural community issues. The availability and use of this data by government, citizens, business and the not-for-profit sector has a vital part to play in achieving the department's policy objectives. The Defra network has a strong record of publishing data with nearly 400 datasets available on data.gov.uk. The GeoStore<sup>2</sup> service provides a wide range of datasets for download and the MAGIC<sup>3</sup> Web site provides visual access to datasets through interactive maps. Defra has published a range of information in response to the Environmental Information Regulations which require public bodies to proactively and progressively make available environmental information.
2. Between June 2012 and April 2014 the Defra network will publish more datasets relating to all aspects of its work. This will not only create potential for innovative reuse, but will seek to achieve positive changes to the environmental behaviour of citizens and businesses. Publishing more data also complements the government's drive to reduce regulatory burdens under the Red Tape Challenge by encouraging business to reduce the environmental impact of generating growth and prosperity without the need for undue regulation. The Defra network deals with a range of customers in relation to regulation of waste, pollution, wildlife and animal health and provides some online facilities for business customers. We will work to improve

---

<sup>1</sup> INSPIRE Directive 2007/2/EC

<sup>2</sup> <http://www.geostore.com/environment-agency/>

<sup>3</sup> <http://magic.defra.gov.uk/>

customer access to data and to publish customer experience and satisfaction data for Defra network services.

3. Defra leads on behalf of the UK government in promoting the publication and reuse of location-specific data through the UK Location Programme. The Programme ensures UK compliance with the EU INSPIRE Directive, and implements the Location Strategy which seeks to maximise sharing and reuse of data. Of the 900+ datasets which have already been published through the Location Programme, around 200 are mandated by INSPIRE, in line with the government's Location Strategy which seeks to exploit the full value of location data.
4. Defra's approach to Open Data is aligned with the *Public Sector Information Principles*<sup>4</sup> developed under the Government ICT Strategy<sup>5</sup>. Defra's Open Data Strategy as a whole is informed by two key principles that *Information is valued as an asset*, and that *Information is managed*. Defra has an ongoing programme of work in place to assure the confidentiality, integrity and availability of information, and to ensure that it is appropriately managed.
5. Defra recognises the importance of information and data to its business and is currently undertaking a review of its vision, strategy, services and structures for managing information and data. Plans are in hand to establish a Sector Board which we will use to drive identification and publication of datasets.
6. Defra has a large network of delivery bodies covered by this Open Data strategy. This includes Natural England (NE), the Joint Nature Conservation Committee (JNCC), the Environment Agency (EA), Forestry Commission England (FC), The Animal Health and Veterinary Laboratories Agency (AHVLA), Veterinary Medicines Directorate (VMD), the Centre for Environment, Fisheries and Aquaculture Science (CEFAS), the Rural Payments Agency (RPA), the Food and Agricultural Research Agency (FERA), the Marine Management Organisation (MMO), and the Royal Botanic Gardens Kew.

## Big Data

### Datasets held by the Defra network

7. The Defra network holds datasets on a wide range of topics for statutory regulation, operational delivery and as an evidence base to support policy making. Large quantities of data are routinely collected, shared and published by the Defra network either as part of everyday activities or through on-line services, such as What's in Your Back Yard<sup>6</sup> or MAGIC. Much of Defra's Big Data has for many years been shared by the Defra network for public and commercial reuse, either on-line or on request. The Defra network will continue to develop these services in line with the Open Data Principles<sup>7</sup> to deliver enhanced environmental

---

<sup>4</sup> Information Principles for the UK Public Sector. <http://www.cabinetoffice.gov.uk/resource-library/uk-government-ict-strategy-resources>

<sup>5</sup> Government ICT Strategy, March 2011. <http://www.cabinetoffice.gov.uk/content/government-ict-strategy>

<sup>6</sup> <http://www.environment-agency.gov.uk/homeandleisure/37793.aspx>

<sup>7</sup> <http://data.gov.uk/blog/new-public-sector-transparency-board-and-public-data-transparency-principles>

decision-making and more beneficial outcomes for our environment. Defra has published nearly 400 datasets on data.gov.uk and plans are being developed to publish more, as well as reducing the burden of data collection on business as outlined in paragraph 22 below. These plans align with principle 6 in the Information Principles for the Public Sector - *Public information is published* - as well as delivering across all of the 6 overarching benefits of Open Data:

- Accountability
- Choice
- Public service productivity
- Public service quality outcomes
- Social growth
- Economic growth

## **Publication of INSPIRE data**

8. Defra leads on behalf of the UK government in promoting the publication and reuse of location-specific data. The UK Location Strategy, launched in 2008, provides a framework that enables location information held by public authorities to be published, found, evaluated, combined with other data and reused. The Location Strategy is underpinned by the INSPIRE Directive which sets out 34 themes for location datasets to be published. The infrastructure that the UK Location Programme is delivering supports the publication of a range of big datasets including addressing, transport networks and land ownership. The implementation timetable for INSPIRE runs to 2020 during which different themed datasets will be published in a specific format with associated Web tools enabling exploitation and reuse of the data.
9. Defra is compliant with the INSPIRE requirements and has published 45 INSPIRE datasets using a shared service run on behalf of the Defra network by the Environment Agency. This number will increase during 2012. The next major publication deadline is December 2013 by which time Defra will identify and publish many more location datasets including those it holds on soil types, land use, utility and government services, environmental monitoring facilities, agricultural and aquaculture facilities, natural risk zones, atmospheric conditions, oceanographic geographic features, sea regions, bio-geographical regions, natural habitats, and species distribution. The detailed specification of which datasets need to be published is due to be finalised by the European Commission around April 2012. In the light of this Defra will identify which datasets it must publish and put in place the necessary systems and procedures to do this. We anticipate that several hundred new INSPIRE datasets will be published by December 2013. Plans are in place to develop the publication system to meet the INSPIRE requirements during 2012, and to publish more datasets on a phased basis

during 2013. It is not possible at this stage to specify exactly which datasets will be published at what date.

## Widening access to science data

10. Defra recognises the importance of making scientific and research information and data as readily and easily available as possible. Our *Evidence Investment Strategy: 2010-2013 and beyond, 2011 update*<sup>8</sup> highlights our commitment to this where feasible. Widening access will help gain best value from investments, encouraging the reuse of data, and improving knowledge exchange, innovation and economic growth.
11. Defra has a wide range of evidence-gathering activities<sup>9</sup>, including natural, social and veterinary scientific research and development (R&D), and non-R&D science such as monitoring and surveillance. Core Defra currently has over 600 active R&D and non-R&D science projects. Costs range from £3k for discrete short-term work to over £12m for long-term or large-scale projects, often involving interdisciplinary and multidisciplinary consortia. Science projects increasingly involve co-funding partners. In widening access to science data, Defra will take account of the complexity of project architectures and the needs of our partners.
12. Information on Defra science projects and R&D reports are published on Defra's website; we also encourage publication of outputs as peer reviewed articles. Research reports and publications do not generally include the full datasets on which the work is based. We aim to extend open access to a much wider range of our research datasets. We will pay particular attention to stress testing for sensitivities, particularly around social science data, to ensure we comply with the Data Protection Act and avoid releasing data that would undermine national security or policies such as protecting rare species. In some cases open access may not be possible.
13. Some key monitoring and surveillance data and information are already available on [data.gov.uk](http://data.gov.uk), linking where appropriate to substantial datasets giving a long-term publicly available archive. We aim to release further monitoring and surveillance data including, for example for animal diseases, horizon scanning, threat assessment and other qualitative risk assessments. In many cases Defra data can best be made available with meaningful context by publishing it through research Web sites and data centres. We will explore the potential to archive data in co-operation with Research Council data centres and similar archives, and to establish links via multiple portals to promote easy discovery and interrogation.
14. To stimulate value-added services, we will build data management for open access into

---

<sup>8</sup> <http://www.defra.gov.uk/publications/2011/04/27/pb13471-evidence-investment-strategy/>

<sup>9</sup> By 'evidence' we mean reliable and accurate information that Defra can use to inform sound decisions in developing and implementing policy. It includes economics, statistics, natural and veterinary scientific information, social research, operational research, engineering, analysis, advice, monitoring and surveillance.

the design of major new research projects. This is already underway for the large-scale Demonstration Test Catchment R&D platform (joint Defra, Environment Agency and Welsh Assembly Government) and the Greenhouse Gas R&D platform, which also include capture, collation and archiving of some relevant pre-existing data streams alongside new monitoring data. Within core Defra our aim is to extend access to science data using Open Data Licences where possible, concentrating first on large new datasets seen to have the highest value-added potential. We aim to use open data standards where possible, taking account of the needs of end-users, for instance in terms of data resolution and granularity. We will release data as soon as possible, mindful of tensions between ongoing research processes and making data available in a timely way.

## Publication of other data

15. Natural England is committed to making its evidence as widely accessible as possible to provide transparency about the decisions it makes and to engage others with its work. Existing programmes such as Customer Service Excellence, Evidence, Geographic Literacy Strategy and Standards will indirectly contribute towards Open Data objectives or are directly aligned with the Information Principles where appropriate. Examples of the actions being taken include: measuring progress in accessibility of evidence through a key performance indicator; publishing geographical datasets and evidence publications; and adopting the Open Government Licence by March 2012 which will facilitate reuse of data.
16. The Centre for Environment, Fisheries and Aquaculture Science (CEFAS) holds scientific datasets which are collected as part of ongoing research work underpinning government objectives, and in support of specific projects for a wide client base including public and private sector organisations. There are a number of large datasets which have been collected within the publicly funded project portfolio. CEFAS already provides public access to many of these datasets through public Web sites, national portals such as MEDIN (Marine Environment Data and Information Network), or international databases such as International Council for the Exploration of the Sea database<sup>10</sup>. CEFAS is undertaking a programme to improve data management which includes provision of Web-based services to the public to search its repository of open data, and to view and download data in line with the requirements of the INSPIRE Directive.
17. The Environment Agency has compiled an Information for Reuse Register<sup>11</sup> (IfRR) which lists datasets that have been approved for reuse under licence. Datasets listed in the IfRR have undergone policy and legal checks to ensure that the data can be released. The presumption is that datasets will be made available under licence, unless there are reasons such as national security or 3rd party intellectual property

---

<sup>10</sup> <http://www.ices.dk/datacentre/Submissions/index.aspx>

<sup>11</sup> <http://www.environment-agency.gov.uk/aboutus/work/38759.aspx>.



rights that mean they cannot be published. A major dataset that the Environment Agency is assessing for possible publication is a database of fly-tipping incidents dealt with by Local Authorities and enforcement action taken.

18. Forestry Commission England already publishes and updates a wide range of datasets. It has plans to publish further data from the National Forest Inventory and the Farm Woodland Scheme.
19. The Joint Nature Conservancy Council will continue to develop access to a range of data through three national portals: the National Biodiversity Network (NBN) which provides species and habitat details from locally collected data; the Marine Environment Data and Information Network (MEDIN) which will be exploring long term data storage in Data Archiving Centres; and the Biodiversity Action Reporting System (BARS) which provides data on conservation initiatives.
20. The Marine Management Organisation (MMO) will identify and publish datasets in line with INSPIRE Directive requirements. The data will cover a range of MMO's key business activities such as Fishing Activity and Fisheries Management, Marine Licensing and Marine Planning. The MMO also holds datasets which are collected as part of specific research and evidence related work underpinning MMO objectives, and in support of specific projects. The presumption is that all these datasets will be made available under licence, unless there are reasons such as commercial or personal sensitivity, national security or 3<sup>rd</sup> party intellectual property rights that mean they cannot be published. Data to be published is likely to include fishing activity for UK vessels, closure of fishing areas and marine licences.
21. In all data management and publishing we will ensure that we comply with the Data Protection Act. We will seek to avoid publishing data that allows inadvertent or indirect disclosure of personal information in breach of the legislation.

## **Rationalising and improving efficiency of data collection**

22. During 2011 as part of the Government Data Review coordinated by the Treasury, Defra identified around 400 datasets where it collects data from external bodies or individuals. 60% – 70% of Defra's collection is mandatory to comply with EU or national legislation. However, within the remaining data collection activities the review identified potential to reduce data collection or to improve collection processes, reducing the burden on business and delivering savings. The Environment Agency is running a pilot to assess the feasibility of providing businesses with a secure portal where they can provide the agency with all the regulatory data that is required in one place rather than supplying it in response to a number of different requests. In response to the recent Farming Regulation Task Force Report<sup>12</sup> a data sharing project has been established,

---

<sup>12</sup> Report of the Independent Farming Regulation Task Force May 2011

<http://www.defra.gov.uk/publications/2011/05/17/pb13527farming-reg-report/>

which aims to reduce the burden of information requests that Defra, its agencies and delivery partners place on farmers by sharing data more fully and effectively within the Defra network. It is hoped that this project will begin to deliver benefits from mid 2012.

## My Data

23. Defra provides a wide range of information and data for citizens but it has comparatively few services that require the collection of data about individuals. The Environment Agency issues permits for a range of regulated activities such as fishing and is currently running a pilot to explore the costs and benefits of enabling permit holders to have online access to their data. This has the potential benefits of reducing the administrative costs of regulation as well as providing users with facilities to update their details or submit new applications online. This approach aligns with principle 7 - *Citizens can access information about themselves* - in the *Public Sector Information Principles*.
24. The Defra Farm Surveys and Assessments service which is accessed through Business Link provides online transactional and self assessment tools to farmers, allowing them to submit and review farm practice data. Defra intends to develop this facility in the context of the new Government Digital Service (GDS) to increase efficiency and improve the experience of farmers in their interaction with the Defra network.
25. Defra provides citizens with access to all the data it holds about them by responding to Subject Access Requests under the Data Protection Act.

## Customer Satisfaction and Experience Data

26. The Defra network has a broad and varied customer base including citizens, businesses of all sizes, other parts of the public sector, and the not-for profit sector. These customers engage with Defra for many reasons including complying with regulations; understanding and responding to environmental problems such as flooding; receiving payments under the Common Market Agricultural Policy; and obtaining information about the natural environment and leisure opportunities it affords. The Defra network regularly collects and analyses customer satisfaction and experience data to ensure that its services are meeting customer needs. It uses survey and complaint data to identify where problems have arisen so that steps can be taken to address them. Customer satisfaction and experience data also help improve and develop services and shape policy. Our approach is aligned with the *Choice* principle in the overarching Open Data benefits, and is also linked to the *Public service quality* principle through the creation of visible feedback mechanisms to drive improvement in the quality of the services delivered by the Defra network.
27. Core Defra's Customer Contact Unit responds to letters and email requests received by the department. It undertakes customer satisfaction surveys on a regular basis and the results of these will be published in the future. Defra has an outsourced telephone

helpline providing responses to around 8,000 enquiries a month. During the first half of 2011 a number of customer satisfaction surveys were undertaken as part of the contract management process which revealed high levels of overall customer satisfaction. These proved costly in relation to the benefits derived and there are currently no specific plans to repeat them, but feedback and complaints about the Helpline are followed up in order to improve the service. Defra has undertaken regular surveys to obtain customer satisfaction feedback in relation to its work to promote sustainable development. Some of this information has been published in themed articles, and we aim to publish more detailed results of future surveys.

28. Forestry Commission England provides a range of services to land owners, businesses, civil society and community organisations aimed at improving and expanding woodland areas. It has established a variety of online and other mechanisms for gathering customer feedback including an innovative online customer rating system which allows customers to review online ratings and user comments. The rating system for Cannock Forest<sup>13</sup> is a good example and provides Forestry Commission England with 12,000 – 15,000 customer ratings a year. These provide valuable feedback on which local managers can take action. We will consider whether this type of approach could be adopted more widely across the Defra network to provide online customer assessments at minimal administrative cost.
29. The Veterinary Medicines Directorate uses a market research agency to undertake regular customer satisfaction surveys with the pharmaceutical industry using a Web based questionnaire. This is followed up by selected in-depth discussions on particular issues. The VMD will assess the scope to publish the results of these ongoing surveys along with plans to address issues they identify. At CEFAS plans are in place as part of the Data Management Programme to provide mechanisms for online customer feedback to establish levels of satisfaction and the quality of the user experience.
30. The Rural Payments Agency undertakes a regular customer satisfaction survey for farmers and traders. In the latest annual survey it met the Ministerial target of 75% satisfaction, and the agency has seen improvements across all of its key performance indicators. Late payments are cited as the most common reason for dissatisfaction. The agency will assess the scope to publish customer satisfaction data on an on-going basis to demonstrate its responsiveness to customer concerns, improve the reputation of the agency and develop confidence in its ability to administer the Common Market Agricultural policy efficiently in support of business.
31. The UK Location Programme will continue to encourage the open publication of location datasets as an ever increasing range of datasets comes within scope of the INSPIRE Directive. The Programme will complete work to deliver a truly interoperable information infrastructure for location data, and work with other partners to ensure the benefits of location data are fully realised.

---

<sup>13</sup> <http://www.forestry.gov.uk/cannockforest>

# Creation of Dynamic Information Markets

32. Within the Defra network individual organisations have initiatives and plans for increasing awareness of the data they publish, engaging with stakeholders to encourage reuse of data, and enabling the commercial sector to use this data to create value added products and services. In addition there are a number of centrally managed cross cutting initiatives with similar objectives. The cross cutting and network initiatives outlined below align with Principles 4 and 5 in the *Information Principles for the Public Sector*, ensuring that data is standardised and linkable in a way that supports reuse.
33. Natural England publishes a range of geographical information datasets under an Open Government Licence to enable its evidence to be readily accessed and reused for broader purposes including commercial reuse. Plans are being developed to publish data in different formats to enable reuse in specific services such as Google Earth that are increasingly used by wide audiences. Natural England carries out a regular Monitor of Engagement exercise in the Natural Environment survey. The publication of this data supports social growth by providing data about enjoyment of the natural environment, and enables the development of opportunities in areas that are accessible to people. Natural England also supports the development of infrastructure to enable the market in biodiversity data including creation of a shared resource of information collected by volunteers, government and non-government organisations. Examples of this include the National Biodiversity Network, Biodiversity Action Recording System and support for citizen science.
34. The Environment Agency's aim is to look after the environment and make it a better place for current and future generations. The data it collects, creates and publishes has an important role in driving social and economic growth. The Environment Agency has plans to develop its DataShare download service which provides almost 800 commercial and public sector partner organisations with access to more than 40 key datasets. It will continue to develop its dissemination of data through commercial channels using value added resellers, of which it currently has 31 including property search companies and the insurance industry. This ensures wide use of data and encourages development of new products by the commercial sector. To promote growth in the use of data the Environment Agency has recently introduced a 3 month free product development licence as a way of achieving Open Data requirements within the current charging framework. The Environment Agency is accredited under the Information Fair Trader Scheme which encourages fairness and transparency in reuse of public sector information.
35. Forestry Commission England is developing plans to provide further datasets on its spatial Data Download Web site.<sup>14</sup> New data releases with potential to drive social and economic growth include data on soil types, recreation, conservation, heritage, and economic assets such as bridges, quarries and reservoirs. The Joint Nature

---

<sup>14</sup> [www.forestry.gov.uk/datadownload](http://www.forestry.gov.uk/datadownload).

Conservancy Council is increasingly focusing on localisation in the context of the Big Society and has a number of initiatives to promote access to data at local levels. These have potential to generate social and economic growth driven by local communities. The Centre for Environment, Fisheries and Aquaculture Science (CEFAS) Data Management Programme will promote discovery of its open data through portals such as MEDIN<sup>15</sup>. It will also publish a range of unique and highly specialised data with potential to support social or economic growth.

## Cross cutting initiatives

36. The 2011 Natural Environment White Paper<sup>16</sup> made a commitment to improve access to environmental information by developing a Web Portal called *My Environment* which will bring together environmental information and provide access to raw data. Work is underway on detailed scoping of *My Environment* and an initial service is expected during 2012. *My Environment* will provide a visible, user-friendly shop window highlighting data published by the Defra network. This can be used to increase awareness of data and stimulate innovative re-use. One option being considered is to support the creation of online communities around specific local environmental initiatives. These would bring together stakeholders from business and local groups and provide access to all the data and information relevant to the community within a single online environment. We are also exploring the possibility of presenting data in customer friendly ways through online interactive maps, allowing citizens to explore and engage more effectively with environmental issues and opportunities in the local and national context.
37. The UK Location Programme has developed an engagement strategy covering INSPIRE data providers and users, focusing on how to publish more and better quality data via data.gov.uk to meet demand and deliver benefits across all sectors. The Programme's benefits strategy will include raising skills and awareness and will look to address barriers to re-use of data. The strategy links to Defra and other departments' strategies, for example the Natural Value Programme and the National Infrastructure Plan. Defra's delivery bodies will prioritise the publication of datasets under INSPIRE using potential for re-use as a key criterion.
38. We recognise that while free and open data is our preferred approach it is not always possible for various good reasons. Defra and its network will work towards ensuring its data is as easily available as possible by using the Open Government Licence where possible or by using simple licences, common approaches and standards, and when necessary, clear and simple charging regimes.

---

<sup>15</sup> <http://www.oceannet.org/>

<sup>16</sup> The Natural Choice: Securing the Value of Nature. Cm 8082, 7<sup>th</sup> June 2011.  
<http://www.defra.gov.uk/environment/natural/whitepaper/>

# Continuous Improvement of the Quality of Data Published

## What data quality means for Defra

36. In the Defra context data quality means that data is fit for the purpose for which it is collected, used or published. This aligns with Principle 3 of the *Principles for Public Sector Information - Information is fit for purpose*. The approach to data quality will vary between different organisations in the Defra network and between different datasets but will include:

- Accuracy of data.
- Timeliness of data publication.
- Compliance with recognised standards such as INSPIRE, the UK Statistics Authority and standards mandated by the Open Standards Board.
- Provision of adequate metadata so that users can match data to their needs in terms of coverage and content.
- Provision of metadata to support searching and discovery of data on the Web.
- Use of appropriate formal or *de facto* standards in controlled vocabularies and code lists.
- Ease of access to data including publication in formats enabling it to be easily understood and re-used by citizens and business, with minimum licensing restrictions.
- Provision of appropriate contextual information helping users to understand it, but ensuring that the integrity of scientific data is not compromised through subjective interpretation.
- Comparability and interoperability of data with other relevant data sets and of the same data over time.

37. We will define appropriate data quality standards and criteria using subject matter experts within Defra and externally where necessary. Our approach to data quality will recognise the trade-off between achieving high data quality and the cost of doing so. We will take this into account when setting standards so that Defra does not waste resources achieving unnecessarily high levels of data quality. The metadata associated with a dataset will include information about the quality of data such as collection and validation methods, so that users can determine the quality of the data.



## Driving up data quality in Defra

38. Defra will drive up the quality of its data to appropriate levels by adopting a process of continuous improvement. We will do this in a number of ways:

- Many Defra network bodies already have data quality policies, standards and initiatives in place. The Environment Agency has developed an Integrated Data Mapping, Modelling and Information Action Plan which covers data quality. Natural England is developing a set of Knowledge, Information and Data strategic principles which support a drive to improve the quality of published data. Specific activities include a programme of work to collate consistent metadata; positional accuracy improvements in core datasets to align with the latest Ordnance Survey maps; creation of operational guidance to support data management; and seeking feedback from users to make improvements in data quality. The Food and Environment Research Agency (FERA) has a set of policies and procedures for data validation. Other delivery bodies have similar initiatives. We will ensure that best practice is shared across the Defra network to achieve consistently high standards.
- We will progressively improve the timeliness and accuracy of the staff related data that we publish to meet transparency requirements for operational data. We will do this by improving our systems and procedures for managing personnel data in a way that will support both internal operations and external transparency requirements.
- Defra already publishes its Quarterly Data Summary which provides a snapshot of how each part of the department is spending its budget, deploying its workforce and the results it has achieved. Plans are in place to improve the quality, timeliness and coverage of this data to support improvements in operations and to reduce costs through public scrutiny and by sharing best practice.
- A large proportion of Defra's published data complies with level 3\* of Sir Tim Berners-Lee's 5 star linked Open Data Scheme so that data is published in a structured way on the Web in non-proprietary formats. There are a few cases where we have gone beyond that level. We will look to expand provision of linked data through [data.gov.uk](http://data.gov.uk). This will include the identification of reference datasets that support wider innovative use of environmental data provided through [environment.data.gov.uk](http://environment.data.gov.uk) and [location.data.gov.uk](http://location.data.gov.uk).
- We will actively look for opportunities to provide more data at 4\* and 5\* using the likely use of this to users to guide our efforts.
- We will use the public scrutiny that the Transparency agenda brings to motivate staff in all parts of Defra to achieve improved data quality.
- We will continue to seek feedback from data users and use this to make improvements to the data we publish.

- Where appropriate we will work with agricultural and environmental stakeholders in government, industry, the not-for-profit sector and the academic community to identify and implement appropriate standards.
- The UK Location Programme reports annually to the European Commission on the quality and extent of INSPIRE compliant data available. The programme facilitates work by groups of data providers to coordinate, monitor and assure data quality. This work will continue and expand over the next two years as INSPIRE data publishing requirements mean further themed data sets are published. The Programme is coordinating the transformation of data as required by the INSPIRE Directive to make it interoperable and encourage re-use.
- Where data is audited under UK or EU legislation we will use the results of audits to assure and plan improvements to data quality systems.
- We will work with The National Archives and other relevant bodies to ensure that we have appropriate retention policies and procedures in place for our data, managing the risks to digital continuity that evolving technology presents. This will ensure that data is available over time not only to meet internal Defra requirements, but also the requirements of the scientific and research community, the not-for-profit and commercial sectors, and the general public.
- As well as driving up data quality within the Defra's network we will work with Defra funded initiatives, such as the National Biodiversity Network (NBN), the Marine and Environmental Data and Information Network (MEDIN) portal, and the Biodiversity Action Reporting System (BARS) to improve data quality.
- We will work across different groups of experts – statisticians, data architects, evidence professionals, geographical information specialists, and IT professionals – drawing on their skills and expertise to achieve an integrated approach to improving data quality. We will support these groups in maintaining and developing the professional expertise needed to contribute fully to this integrated approach.

## Statistical data

39. The Defra statisticians are working very closely with the department's data management team on improving data availability and enabling its effective use. Key drivers for the statistical work are fully in line with the Open Data Strategy:

- Increasing the relevance of the information available developing new sources and analysis as appropriate informed by internal and external user consultation
- Ensuring costs effectiveness collecting the right data efficiently
- Reducing compliance costs making maximum use of administrative sources
- Ensuring compliance with the quality standards required in national and official statistics.
- Statistics are relevant , trustworthy and easily understandable



40. In developing our approach to statistical data our key priorities are:

- Better understanding the needs of users and how the data is being used to improve our impact and contribution
- Developing more inventive ways to bring the issues and data alive with easier access via the websites and data sharing benefitting from the INSPIRE programme
- Improving the data and our understanding of the drivers of change working in partnership with other specialists.
- Continuing to develop closer links on the Defra statistical network , the Government Statistical Service and internationally
- Considering how we can meet local data needs with reducing resources

41. A key statistical initiative is the Farm Business Benchmarking programme which provides a Web-based service<sup>17</sup> allowing farmers to compare the financial performance of their business with their peers. The programme uses data from the Defra-commissioned Farm Business Survey for comparison by farm type, farm size, and performance level. It supports benchmarking using trustworthy statistics in an easily understandable visual way. The initiative provides better understanding of the needs of farmers and uses data to improve Defra's impact in helping the farming industry.

## **Transparency governance and culture change**

42. Defra has a Board level Transparency Champion responsible for overseeing Defra's response to the government's transparency agenda. The Chief Information Officers Group across the Defra network also plays an important role. During the first half of 2012 Defra is undergoing significant organisational change and re-structuring. We will ensure that the new structure provides an effective means of responding to and pro-actively developing a transparent approach to data publication. As part of this we will establish new governance arrangements for transparency aligned to the new organisational structure.

43. We are in discussions with the Cabinet Office Transparency Team about the creation of a Sector Board that will play a key role in driving data quality and transparency across the Defra network and the wider environment sector. The Sector Board will also help to achieve the culture change needed to embed Transparency in policy development and operational delivery throughout the Defra network, so that it becomes an integral part of the systems we use and the way we work.

---

<sup>17</sup> [www.farmbusinesssurvey.co.uk/benchmarking/](http://www.farmbusinesssurvey.co.uk/benchmarking/).

## Annex: Data to be released as open data up to 31 March 2014

Dataset Name	Description of data	To be first published on	Frequenc y of update	Issued under the Open Government Licence (OGL) Yes or No <sup>*18</sup>
INSPIRE Directive Annex III datasets	<p>Datasets published under the EC INSPIRE Directive under the following themes:</p> <ul style="list-style-type: none"> <li>– Utility and governmental services</li> <li>– Environmental monitoring facilities</li> <li>– Production and industrial facilities</li> <li>– Area</li> </ul>	Phased publication up to December 2013	Varied	Published under OGL where possible. A range of licensing schemes are in place where this is not possible.

---

<sup>18</sup> The OGL allows the public and business to, for free: copy, publish, distribute and transmit the Information; adapt the Information; exploit the Information commercially for example, by combining it with other Information, or by including it in their own product or application. The full licence can be found at <http://www.nationalarchives.gov.uk/doc/open-government-licence>

	<p>management / restriction / regulation zones and reporting units</p> <ul style="list-style-type: none"> <li>– Natural risk zones</li> <li>– Meteorological geographical features</li> <li>– Oceanographic geographical features</li> <li>– Species distribution</li> <li>– Energy resources</li> </ul>			
UK Marine Strategy Area Sub-Regions	Formal boundaries of the UK Marine Strategy Area and the MSFD sub-regions in UK waters for the purposes of implementation of the EU Marine Strategy Framework Directive	October 2012	As needed	Yes

Charting Progress II Assessment Areas	Biogeographically defined assessment areas used in the UK Marine Monitoring and Assessment Strategy Charting Progress 2 assessment of the state of UK seas. These areas have been used on an informal basis in the UK implementation of the EU Marine Strategy framework Directive.	October 2012	As needed	Yes
<b>Environment Agency datasets</b>  Catchment Abstraction Management Strategies (CAMS) Reference Boundaries  Extractive Materials Management Statement Summaries - Corporate Entities Only  Historic General Quality Assessment of Headline Indicators of Water Courses – Biology  Historic General Quality Assessment of Headline Indicators of Water Courses – Chemistry  Historic General Quality Assessment of Headline Indicators of	Various datasets which have been approved for release as part of a continued data sharing programme.	Progressively from July 2012	Varied	Yes

Water Courses – Nutrients Historic River Quality Objectives Marine Benthic Invertebrate Species Monitoring of Pesticides and Trace Organics in Water (1992-2008) Referrals of Red List Discharges to Sewers - Corporate Entities Only Saltmarsh Extents Saltmarsh Species Seagrass Taxa and Abundance Sensitive Area Maps Summary Shellfish Directive Assessments Waste Management Licence Current Exemptions WFD Lake Waterbodies WFD Monitoring Network WFD Rocky Shore Macroalgal Species				
Environmental statistics	Multiple environmental statistics datasets including air quality, water quality and biodiversity.	2012 – 2013	Varied	Yes

Agricultural statistics	Multiple datasets including farm surveys, payments and production.	2012 – 2013	Varied	Yes
Operational transparency	Administrative data including procurement expenditure and staffing details.	2012 – 2013	Varied	Yes