

DISCUSSION DOCUMENT

A CENTRALISED FORMAT TO HARMONISE AGRI-ENVIRONMENTAL DATA



About WRAP

WRAP is a climate action NGO working around the globe to tackle the causes of the climate crisis and give the planet a sustainable future.

Our core purpose is to help you tackle climate change and protect our planet by changing the way things are produced, consumed, and disposed of.

Written by: Joseph Poore, Matthew Barker

Disclaimer: While we have taken reasonable steps to ensure this report is accurate, WRAP does not accept liability for any loss, damage, cost or expense incurred or arising from reliance on this report. Readers are responsible for assessing the accuracy and conclusions of the content of this report. Quotations and case studies have been drawn from the public domain, with permissions sought where practicable. This report does not represent endorsement of the examples used and has not been endorsed by the organisations and individuals featured within it. This material is subject to copyright. You can copy it free of charge and may use excerpts from it provided they are not used in a misleading context and you must identify the source of the material and acknowledge WRAP's copyright. You must not use this report or material from it to endorse or suggest WRAP has endorsed a commercial product or service.

For more details please see WRAP's terms and conditions on our website at www.wrap.org.uk

Executive summary

This document has been developed under the Food Data Transparency Partnership (FDTP). Through discussions in FDTP working groups and stakeholder meetings, data infrastructure and governance have been identified as fundamental in the delivery of accurate, accessible, and consistent environmental impact data to the food and drink sector.

Within the theme of 'data infrastructure', the FDTP has found that one of the main challenges is the lack of consistency in how agri-environmental data are represented. The development of a consistent data infrastructure to provide data interoperability and sharing has therefore been prioritised under workstream 2.3 of the FDTP for delivery by 2025. Whilst not representative of government policy, this note explores how such an idea could work in practice.

A data format can be used to represent data describing the inputs, products, practices, and environmental impacts of food and drink businesses. A single harmonised data format would mean these data are represented consistently across organisations. It would enable data systems to be developed independently which are interoperable, streamlining and enhancing data-driven approaches to solving food system challenges. A data format can also be used to set the required levels of data transparency for the entire agri-food sector or for parts of the sector.

There are multiple use-cases for, and benefits from, a harmonised data format. They range from supporting the harmonisation of farm carbon calculator results, to enabling environmental impact data to be exchanged in supply chains, to supporting a consistent approach for environmental impact labelling. Delivering a harmonised format also comes with certain risks (e.g., low international interoperability if developed for the UK), but the risks we identified can be mitigated.

It is our view that harmonisation can only occur through some degree of centralisation (i.e., approaches to drive widespread usage such as business collaboration, research collaboration, or through policy). It is also our view that the HESTIA format – which can represent both environmental impact and activity data, is focused on agri-food, is aligned to key standards, and drives a high level of transparency – provides the best starting point, although the format requires further testing to ensure it can deliver for key use cases and continuous development to be widely applicable.

Here we are seeking input from the Food Data Transparency Partnership (FDTP) on the following:

1. What are the risks and benefits of data format harmonisation, and how can the risks be mitigated? [Section 2]
2. What is the most appropriate existing data format for harmonisation, and how does this change with different use-cases? [Section 3]
3. What additional testing and development needs to be done to ensure the chosen data format is fit for purpose? [Section 4]

A CENTRALISED FORMAT TO HARMONISE AGRI-ENVIRONMENTAL DATA

4. What mechanisms can and should be used to drive uptake and usage of the data format, and how does this change with different use-cases? [Section 5]
5. How can members of the FDTP support the testing and development of a harmonised data format?

The rest of this document sets out necessary context for discussing data format centralisation for food systems, presents our views on the subject, and provides policy considerations for the next steps to create a fit-for-purpose data format and to ensure its widespread usage.

Contents

Glossary	1
1.0 Introduction	2
1.1 Which FDTP priorities does this document address?	3
1.2 Who is this document for?	3
2.0 Benefits and risks of harmonisation	4
2.1 What is a data format?	4
2.2 Features of a data format	4
2.3 Benefits of data format harmonisation	4
2.3.1 Transparency	4
2.3.2 Data porting	5
2.3.3 Data flow in supply chains	5
2.3.4 Database harmonisation	5
2.3.5 Primary data collection and usage	6
2.3.6 Application development	6
2.4 Risks of data format harmonisation and centralisation	6

A CENTRALISED FORMAT TO HARMONISE AGRI-ENVIRONMENTAL DATA

3.0	Comparison of formats	8
3.1	Overview	8
3.2	Comparison table	9
4.0	Work required	13
4.1	What parts of the food system should be covered by a harmonised and centralised format?	13
4.2	Who should the data format be tested with?	13
4.3	Work plan	14
5.0	Policy considerations	15
5.1	Which groups could be involved?	15
5.2	What might centralisation look like?	15
5.3	Who should own and maintain the format?	16

Glossary

Activity data: Data describing inputs, products, practices, and site measurements.

API: Application programming interface

LCA: Life Cycle Assessment

GHG: Greenhouse gas

FDTP: Food Data Transparency Partnership

Primary data: Environmental impact data that have been estimated using measurements that are specific to a company or their supply chain.

Secondary data: Environmental impact data that have been estimated using generic measurements that are not specific to a company or their supply chain. The data often come from databases that have been compiled by a third party.

1.0 Introduction

Today, data describing farms, food processors, and their environmental impacts are inconsistent, generally incomparable, often incomplete, hard to re-use, and often hard to access. This is hindering interventions to deliver a more environmentally sustainable and productive food system.

The Food Data Transparency Partnership (FDTP) is a partnership between the Department for Health and Social Care (DHSC), the Department for Environment, Food and Rural Affairs (Defra), the Food Standards Agency (FSA), industry, academia, and civil society. It aims to improve the availability, quality, and comparability of data to increase the production and sale of more environmentally sustainable and healthier food and drink products.

A fundamental issue identified by the FDTP was the lack of consistency in how agri-environmental data are represented: that today, virtually all organisations in this area use their own data structures, have their own terms for the same things, have their own minimum data requirements, and have their own ways to describe data quality and completeness. The development of consistent data infrastructure to provide data interoperability and sharing has therefore been prioritised under workstream 2.3 of the FDTP for delivery by 2025.

Here, we explore how a single, consistent, and prescriptive data format could be used to drive the desired levels of transparency and interoperability in agri-food supply chains. While this has a range of use cases, we focus on how it could:

- Increase the transparency of environmental impact data, including results generated by digital farm environmental impact calculators and other supply chain tools.
- Enable farmers to move their data more easily from one digital tool to another.
- Support the flow of consistent and transparent environmental data in food supply chains.
- Harmonise data and increase transparency in secondary environmental impact databases.
- Enable primary data to flow into secondary databases in a consistent way.
- Harmonise between different research databases (e.g., crop breeding datasets), thereby increasing knowledge exchange and scientific progress.
- Support the creation of web-based applications that use agri-environmental data.

We then identify the barriers to adoption of a single consistent data format. We explore the types of intervention that could drive adoption and lay out a plan for a pilot study to prove this out.

This document represents our best understanding of the opportunities and challenges of data format centralisation. Here, we are seeking views from FDTP members to complement and evaluate this briefing. At the end of each section, questions for readers are highlighted in a table.

1.1 Which FDTP priorities does this document address?

This document primarily focuses on the following FDTP priority for 2025:

- ▶ A pilot to trial a common, consistent, and unambiguous language or format for data storage and sharing of environmental impact data for agriculture and food products (Priority 2.3.1).

It also contributes to these priorities:

- ▶ Support for carbon tool providers to harmonise their underlying methodologies (Priority 2.1.2).
- ▶ More consistent and comparable secondary data for the environmental impact of agri-food products where data gaps are identified, made available on an open-access database (Priority 2.2.1).
- ▶ A transparent web based accessible database for 3rd party use, supplying appropriate secondary environmental impact data, to support industry scope 3 reporting and product level ecolabelling (Priority 2.2.3).
- ▶ Design for infrastructure to enable the collection of activity data to increase the supply of primary impact data, share this efficiently, and aggregate it to improve secondary data (Priority 2.3.2).

1.2 Who is this document for?

This document is for members of the FTDP, policy-focused civil servants at Defra, DHSC, and the FSA, and for organisations working with agri-environmental data.

2.0 Benefits and risks of harmonisation

2.1 What is a data format?

A data format provides a common framework and language to represent information. Data formats can be used to drive transparency by requiring a minimum level of information.

While a data format used by a single organisation can scale their own operations, its consistent use by many organisations can multiply benefits through supply chains and sectors: systems developed by different organisations can communicate and achieve interoperability.

2.2 Features of a data format

Different data formats provide different levels of functionality that are tailored to use cases. Some formats are well-tested and have many applications built upon them, whereas some are less well tested and built upon. Particularly useful features for food data transparency and reliability include:

- **An associated glossary of terms** (ontology), which reduces ambiguity and enables shared definitions. E.g., “wheat” could be with or without the husk, have a certain dry matter etc.
- **Minimum data requirements.** A data format can require that certain fields are recorded. E.g., that a quality score is assigned, or that any models used are specified.
- **Basic data validation.** A certain level of validation can be built into the format itself to prevent basic errors, for example, that a value for a certain data type must be greater than zero.
- **Software libraries built on top of the format.** These can perform tasks like sophisticated data validation, gap filling, visualisation, and calculation.
- **Data exchange APIs.** APIs make export and import of data in a certain format easier.

2.3 Benefits of data format harmonisation

2.3.1 Transparency

Data formats which require certain minimum fields drive transparency. While standards and guidance can achieve the same thing, there can remain ambiguity and omissions, whereas a data format is simply invalid if certain required data are not provided. In practice, data formats complement standards, e.g., by requiring the emissions disaggregation in the GHG Protocol.

While FDTP Priority 2.1.2 is to harmonise methods between farm carbon calculators, a data format could require that the models used are clearly documented through a minimum requirement that the model is specified for each emission. They can also require a minimum level of emission disaggregation too. And could even require that all activity data are reported, enabling remodelling with harmonized methods.

Further, if the FDTP requires a certain quality score to be assigned to data (FDTP Priority 2.2.2), this could be made a requirement in the data format itself.

2.3.2 Data porting

Farmers and other businesses can spend substantial time entering their data into a digital tool. But if they want to use a different tool (e.g., to benefit from different functionality or because their buyers ask them to) it can take time to enter their data again. This creates barriers to competition (data are “locked in”) and hinders farmers and businesses from maximising usage of their data.

A harmonised data format would allow farmers or other businesses to export their data from one tool and easily import it into another (called “data porting”). Data porting has been a major focus in other sectors (e.g., banking, cloud computing etc.) and should be a focus in agri-food too. Data that are easily ported is also more easily controlled by the data inputter, which could provide businesses with greater confidence to share data.

2.3.3 Data flow in supply chains

A single data format would allow data to flow from agricultural input manufacturer to farm to processor to retailer in a seamless way. It would enable data requests from one company to many suppliers or to many buyers to be handled, enabling sustainability information to flow. This is likely critical for Scope 3 accounting which is based on real supply chain data.

2.3.4 Database harmonisation

Secondary environmental impact datasets could be made more transparent and consistent (FDTP Priority 2.2.1) if they allowed export of their data in a consistent format. The minimum required fields by the format would also drive changes in database curation to reach the desired levels of transparency and consistency.

Further, the data behind many other agricultural datasets (e.g., crop breeding datasets, animal breeding datasets, and farm survey datasets) could be harmonised using a consistent data format. This is very important to allow the latest scientific research to be exchanged and used.

Of particular relevance, many large datasets of field measured emissions exist (e.g., datasets on soil carbon fluxes and N₂O) and harmonising these would support new model creation and provide more measured data to calibrate the accuracy of different farm carbon calculators.

2.3.5 Primary data collection and usage

FDTP Priority 2.3.2 is to increase the collection of primary data and enable it to flow into secondary databases. This would be supported by a common data format by: ensuring that primary data collection efforts were harmonised (although standards will be essential to complement this); and increasing the comparability of primary data to feed directly into secondary datasets.

2.3.6 Application development

Software applications, such as web-based applications to access environmental impact data (FDTP Priority 2.2.3) are directly supported by a common data format, particularly if the data format is in a widely used language (e.g., JSON).

2.4 Risks of data format harmonisation and centralisation

The following risks have been identified, with mitigation options suggested where possible:

- **Widespread usage of a data format requires legislation**, which could be a political barrier if there is not the will to pass such legislation. Even if legislation were passed, this is a lengthy process that could delay harmonisation.
- **The data format developed is not fit for purpose**. There are three migration measures that should be undertaken to minimise this risk:
 1. There should be an enhancement and pilot period for any proposed data format to ensure that it fits the needs of a government-backed format as far as possible.
 2. A core stable format should be provided, with regular subsequent releases on predictable timelines. Resourcing, funding, and responsibilities of partnering organisations should be defined to ensure delivery.
 3. For particularly intractable issues that may prevent widespread adoption, contingency plans could be enacted. This may involve altering the coverage or scope of the format, or other plans may be used as appropriate to ensure the effective harmonisation while being cognisant of practicalities involved.

- The data format is developed in a UK context, potentially resulting in a format that is well-suited to representing UK-based supply chains and food systems but unable to accurately represent other geographies. This risk can be minimised through inclusion of a testing phase that includes domestic and international users.
- The UK's approach misaligns with international partners. We believe the likelihood of this risk to be small, but the potential impact of misalignment is high because food supply chains often cover multiple jurisdictions. Therefore, due consideration of the emerging international landscape should be considered and opportunities to drive international alignment should be pursued (e.g., through OECD or UN networks, and through working with existing organisations active in this area such as WBCSD who provide the PACT format).
- A private data format may suit the needs of some businesses better than a centralised format and may be more responsive to user requirements. In order to minimise this risk, comprehensive testing should be undertaken for the proposed centralised format. Furthermore, a clear process should be developed to regularly incorporate user feedback into the centralised format and to provide updates that increasing the format's applicability.

Questions for the FDTP eco and data working

To what extent do you agree with the benefits, risks, and mitigations of data format harmonisation?

Are there any additional risks or mitigation measures related to data format harmonisation that Defra should consider?

To what extent do you agree that the benefits of data format harmonisation outweigh the mitigated risks?

3.0 Comparison of formats

3.1 Overview

We identified five existing data formats that may meet the goals above. All formats are well tested for certain use cases and have applications built upon them:

- Three main LCA formats exist (ILCD, ecoSpold, and openLCA). They store environmental impact data ("inventory data"). However, they do not have minimum requirements in key transparency areas (e.g., minimum levels of emission disaggregation), nor do not allow activity data to be stored consistently.
- The PACT format stores environmental impact data, but unlike the LCA data formats, it has more minimum data fields to increase transparency. However, it only stores GHG emissions data, and does not allow farm or supply chain activity data to be represented.
- The HESTIA format stores environmental impact data, has key minimum data requirements, and stores activity data (with a particular focus on agricultural activity data, although any activity data can be represented). Data in the HESTIA format can be exported into LCA formats, and work is currently being undertaken to export into PACT.

Given these formats exist, we do not explore the need to develop a new format here.

3.2 Comparison table

Features that are important for data transparency, interoperability, & reliability

Key use cases from the FDTP priorities

Data Format	Description	Features				Use Cases		
		Associated ontology?	Minimum data requirements?	Inbuilt data validation?	Software libraries?	Farm-level data harmonisation	Supply-chain impact and activity data exchange	LCA database harmonisation
ILCD	A format to share LCA data developed by the EU's JRC.	No – only recommendations for consistent naming are provided.	Limited – mandatory requirements are minimal and would not support transparency or data quality.	Limited – data validation tool available to check format usage and naming conventions.	Yes – import and export supported by most LCA software, with functionality in LCA software then available.	Inappropriate – lack of ontology, lack of data fields, and no agriculture-specific minimum data requirements.	Inappropriate – lack of minimum data requirements cannot support transparent data exchange.	Limited – substantial differences in the quantity, quality, and transparency of data are possible.
ecoSpold	The data format used byecoinvent for their lifecycle inventory.	Limited – only geographic data and locations are consistently defined.	<i>As above</i>	Limited – but further validation is available in software packages (e.g., ecoeditor).	<i>As above</i>	<i>As above</i>	<i>As above</i>	<i>As above</i>
openLCA	Software to support LCA	No	<i>As above</i>	Limited – but validation is	<i>As above</i>	<i>As above</i>	<i>As above</i>	<i>As above</i>

Data Format	Description	Features				Use Cases		
		Associated ontology?	Minimum data requirements?	Inbuilt data validation?	Software libraries?	Farm-level data harmonisation	Supply-chain impact and activity data exchange	LCA database harmonisation
	modelling developed by Green Delta.			available on import into some software.				
PACT	A data standard and network to support the exchange of consistent product carbon footprints through supply chains.	Limited – some external classification systems used (e.g., products under the CPC classification, but CPC doesn't enable comparison of farm/processing practices).	Yes – mandatory requirements that support transparency e.g. requirement to disclose the cross-sectoral standard used.	Partial – validation occurs through and API supplemented with conformance testing between organisations.	Limited – an API to import/export data.	Inappropriate – lack of ontology and food-specific minimum data requirements, or food-specific granularity requirements.	Limited – for GHG emissions it may be suitable (although it does not require a detailed emission breakdown), but it does not function for other environmental impact indicators nor support activity data representation.	Inappropriate – it does not store LCA inventory data beyond characterised GHG emissions.

		Features				Use Cases		
Data Format	Description	Associated ontology?	Minimum data requirements?	Inbuilt data validation?	Software libraries?	Farm-level data harmonisation	Supply-chain impact and activity data exchange	LCA database harmonisation
HESTIA (schema , glossary)	Free, openly available data format, consisting of a schema describing how data must be represented and what data are required, plus a glossary of terms (ontology).	Yes – a comprehensive glossary providing consistent definitions for agriculture and food systems including coverage of practices, products, inputs, processing, emissions, impacts, methods, models, regions, and properties.	Yes – mandatory requirements that drive transparency, for example requirements to disclose the completeness of the data related to an impact assessment.	Yes – substantial inbuilt validation (e.g., minimum and maximum values), with further sophisticated validation available in an open-source software package.	Yes – modelling tools are freely available to model impacts base on input data submitted in the HESTIA data format, as are data visualisation and analysis tools.	Appropriate – clear, comprehensive, food-specific ontology, granular emission breakdown required, description of data goes beyond emissions to multiple environmental impact and includes all activity data too.	Appropriate – some modifications may be needed for transparent multi-metric impact data exchange.	Appropriate – the format has represented ~750 individual LCA studies and data from existing LCA databases and drives high levels of transparency and granularity.

Given HESTIA's performance against the use-cases, we recommend that this data format as the basis for harmonisation efforts, and that a programme of work is started to make it fit for purpose.

Questions for the FDTP eco and data working

To what extent do you agree with the use-cases for data format harmonisation?

To what extent do you agree with our assessment of existing data formats performance against use-cases?

Are there any additional or different features of a harmonised data format that are important to support the use-cases identified?

Are there any additional or different use-cases for a harmonised data format that should be considered? If so, what are the important features of a data format to support the use-case?

4.0 Work required

4.1 What parts of the food system should be covered by a harmonised and centralised format?

The most complex and largest source of GHG emissions and environmental impacts for food and drink products occurs during primary production. Some degree of processing occurs to transform most commodities into ingredients and finished food products.

Our initial suggestion is to create a data format that can be used to describe the entire food production and processing system up to the point of products ready for consumption/use.

4.2 Who should the data format be tested with?

We recommend initial testing with more technologically focused users, and then expanding out to further groups:

- Digital farm tools (e.g., the Cool Farm Tool, Agrecalc, Farm Carbon Toolkit, Soil Association Exchange in the UK, and others internationally). This would prove the format's ability to import and export farmer data and explore its role in driving tool harmonisation.
- Digital supply chain tools (e.g., Mondra, Altruistiq, Foodsteps, and My Emissions in the UK, and others internationally), with a focus on proving that the format can be used to import farm and processor environmental impact data and flow it up the supply chain, and also to drive harmonisation between these tools.
- LCA datasets (e.g., CIEL, Dairy UK, and Seafish in the UK, and Bord Bia, ecoinvent, GFLI, Agribalyse, and others internationally), with a focus on proving the format can represent their data, support the flow of primary data into their datasets, and to explore how it could support harmonisation.
- Research organisations (e.g., Rothamsted and SRUC in the UK, and others internationally such as FiBL, CIMMYT, CIRAD, FAO), with a focus on proving the format can represent their data, support knowledge exchange, and enable their data to flow into secondary datasets.
- Government (e.g., Defra in the UK and other governmental organisation holding agri-environmental data internationally) with a focus on how the format could support National Inventory harmonisation, agri-environmental data harmonisation and re-use, and for their data to flow into secondary datasets.
- International standards setters (e.g., SBTi, GHG Protocol, WRI, ISO, UNEP, amongst others) to ensure that their requirements are reflected in the HESTIA format's requirements and key parts of standards compliance can be checked within the format.

4.3 Work plan

We recommend a 12-month timetable to develop and test the data format and deliver enhancements. This will include:

- A 6-month scoping period, during which key user groups will be identified, areas to focus on will be prioritised, and user needs and challenges assessed.
- A 4-month simultaneous development and consultation period will be used to enhance the format to address the problems identified. Longer-term work, if needed, will be identified during this period, and issues created and prioritised, and a resourcing plan identified.

After the 12-month testing and development period, we believe that the HESTIA data format would be in a position to fulfil the requirements of multiple groups identified above, and therefore of the FDTP’s ambitions.

We recognise this is an ongoing process, and after the initial period we will work with other interested organisations to support the harmonisation of their infrastructure with this format.

Questions for the FDTP eco and data working
To what extent do you agree with the proposed coverage of a centralised data format?
To what extent do you agree with our proposed approach to testing the data format and supporting harmonisation?
Is there additional testing that should be undertaken?

5.0 Policy considerations

Assuming data format centralisation has been accepted as an aim, and a data format suitable for centralisation has been agreed upon and tested, an approach to driving harmonisation and uptake through industry academia and government needs to be agreed.

In this section we suggest which groups could be involved in the process, and what policy mechanisms could be used to support harmonisation.

5.1 Which groups could be involved?

Given the focus on the entire food supply chain, the UK and international users, and the coverage of research, business, and policy, we expect this will require a multi-stakeholder coordinated effort, including at least:

- Farming industry bodies, including organisations like AHDB and NFU in the UK, as well as international organisations like Copa Cogeca.
- Commercial businesses that provide LCA data, including LCA calculators, supply chain data providers, LCA consultants, and LCA databases.
- Key businesses throughout the supply chain, from input producers to farms to food companies to retailers.
- Government, including Defra and international government agencies, as well as multi-national organisations like the OECD.
- Research organisations, covering both agriculturally focused and environmentally focused groups.

Ideally, we would work with the WBCSD to extend and enhance the work they have already done with PACT and accelerate it to cover a wider range of cases. In any case, we will ensure that the HESTIA format is PACT compliant at a minimum.

We propose to undertake a scoping exercise to identify, with the support of the FDTP, a list of organisations that could be involved, balancing the needs of Defra and FDTP with organisations' willingness to collaborate.

5.2 What might centralisation look like?

There are different ways to achieve centralisation, and we list some of them here. The approach taken will depend on discussion and engagement with stakeholders. We have organised the suggestions from least to most policy-driven.

- LCA or other databases providing their data in this format, which could result in more widespread usage.
- Businesses requesting data in this format. At a certain level, this would lead to widespread adoption, particularly if driven by the largest organisations.
- Research funding and consulting contract funding linked to providing results and data in this format.
- Defra (and other government organisations) providing public data in this format and requesting data (e.g., farm survey data) in this format.
- Public financial or non-financial support to re-work databases for format compliance or for training and usage.
- Recommending that certain organisations make user data available in the format.
- Mandating that certain organisations make user data available in the format.

5.3 Who should own and maintain the format?

Our proposed approach to a centralised data format is to utilise and build on the existing HESTIA format. The format is available on an open-source basis with a CC-BY licence and is hence available to reuse and modify by anyone for free. It is currently maintained by The University of Oxford. But maintenance requirements could be transferred or shared in the future.

Questions for the FDTP eco and data working

To what extent do you agree with the coverage of data businesses and potential user groups of a centralised data format?

To what extent do you agree with the centralisation option laid out in this section?

Which policy options for centralisation do you think are most appropriate and why?

Are there additional policy options that should be considered to support centralisation?

How can FDTP be best used to drive the momentum for data format harmonisation?



WRAP's vision is a thriving world in which
climate change is no longer a problem.

Our mission is to make the world a more
sustainable place. We bring people together,
we act on the facts, and we drive change.

Find out more at
www.wrap.org.uk