



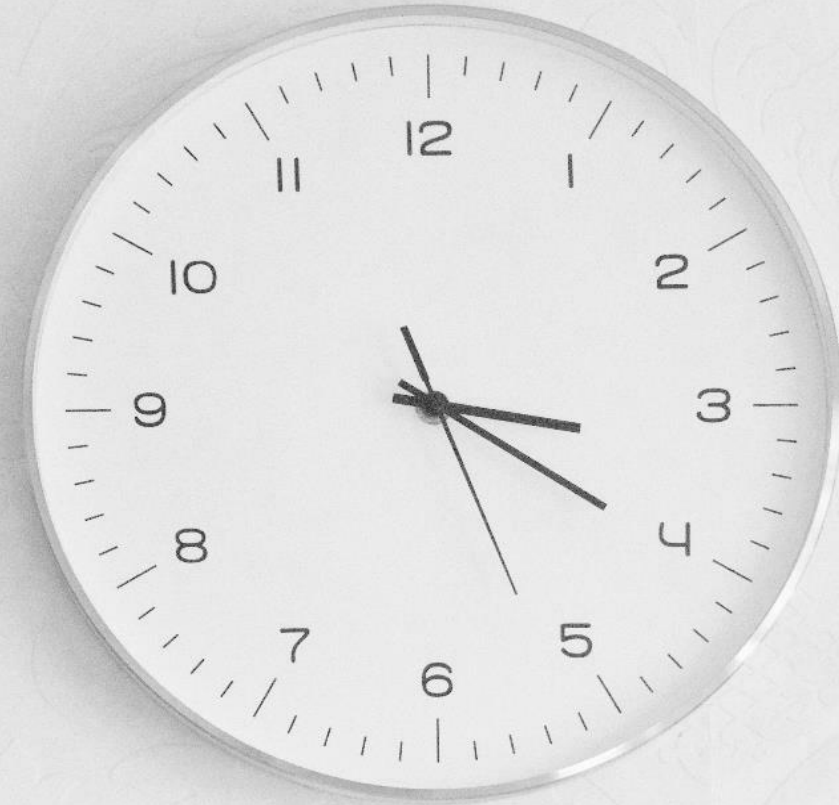
# **Introduction to the study on High Value Datasets**

**Impact assessment study on the list of High Value Datasets to be made available by the Member States under the PSI Directive**

February, the 11<sup>th</sup> 2020

# Agenda for this webinar

- Short introduction to **the study team**
- Objectives, scope and **understanding** of the assignment
- **Methodological** approach for the data collection and analysis tasks
- **Timeline** and planning of the activities and **where/when we will need your help**
- Q&A







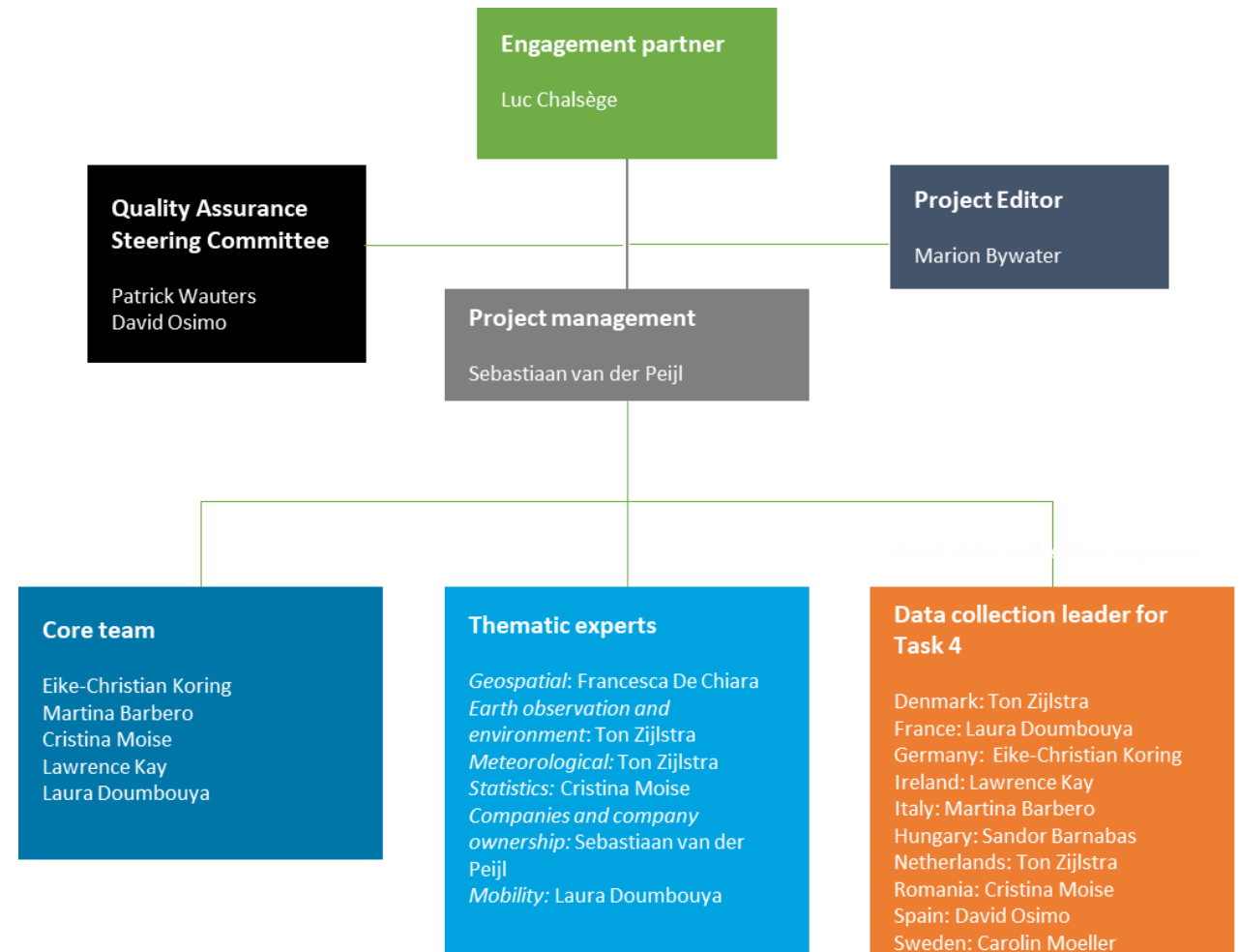
## **Short introduction to the study team**



# About our study team

The study team is composed by four different organisations and a number of thematic and country experts

- Four different organisations with complementary expertise:
  - **Deloitte**
  - **The Open Data Institute**
  - **The Green Land**
  - **The Lisbon Council**
- We have a **Project Manager and a core team** composed of different profiles.
- For **each thematic area** and we have one thematic expert or coordinator who will be in contact with national data providers and re-users. For geospatial → **Francesca De Chiara**.
- For **a limited set of data collection activities** in the 10 shortlisted Member States (only in the context of Task 4 – further discussed below) we have **lead data collection experts** already identified within the study team.







# Objectives, understanding and scope of the assignment

# Objectives and understanding of the assignment

## Objectives and scope 1/3

The overall objective of this assignment is to « **define concrete High Value Datasets that fall under the thematic categories included in the Annex of the revised Directive, based on an iterative process involving a number of cycles, by which an initial broad range of datasets will be narrowed down.** »

### Main content scope

- The categories of high value datasets identified in Annex I of the Directive (*geospatial, earth observation and environment, meteorological, statistics, companies and company ownership, transport*);
- The **other areas of legislation** (at the European level) which cover the provision of certain datasets by public sector authorities and must be taken into account for ensuring that the overall legislative framework on public sector information is coherent and effective.

### Additional aspects to take into account

- **Business models** of bodies who are currently charging for data and how these will be affected by the Directive;
- **Competition issues** for public undertakings opening up high value datasets whereas their private competitors do not have the same obligation;
- Importance of high value datasets for **SMEs and for building AI applications** and services.

Examples of datasets in scope	Examples of datasets out of scope
All datasets covered by the INSPIRE Directive.	Datasets covered by the Intelligence Transport System Directive and by the ITS Delegated Regulations
Datasets covered by the Directive on Public Access to Environmental information.	Datasets covered by the European Commission Delegated Regulation on multimodal travel information services
Datasets covered by EU company law legislation (i.e. Anti-Money Laundering Directive, Company Law Directive ...)	...
....	....

# Objectives and understanding of the assignment

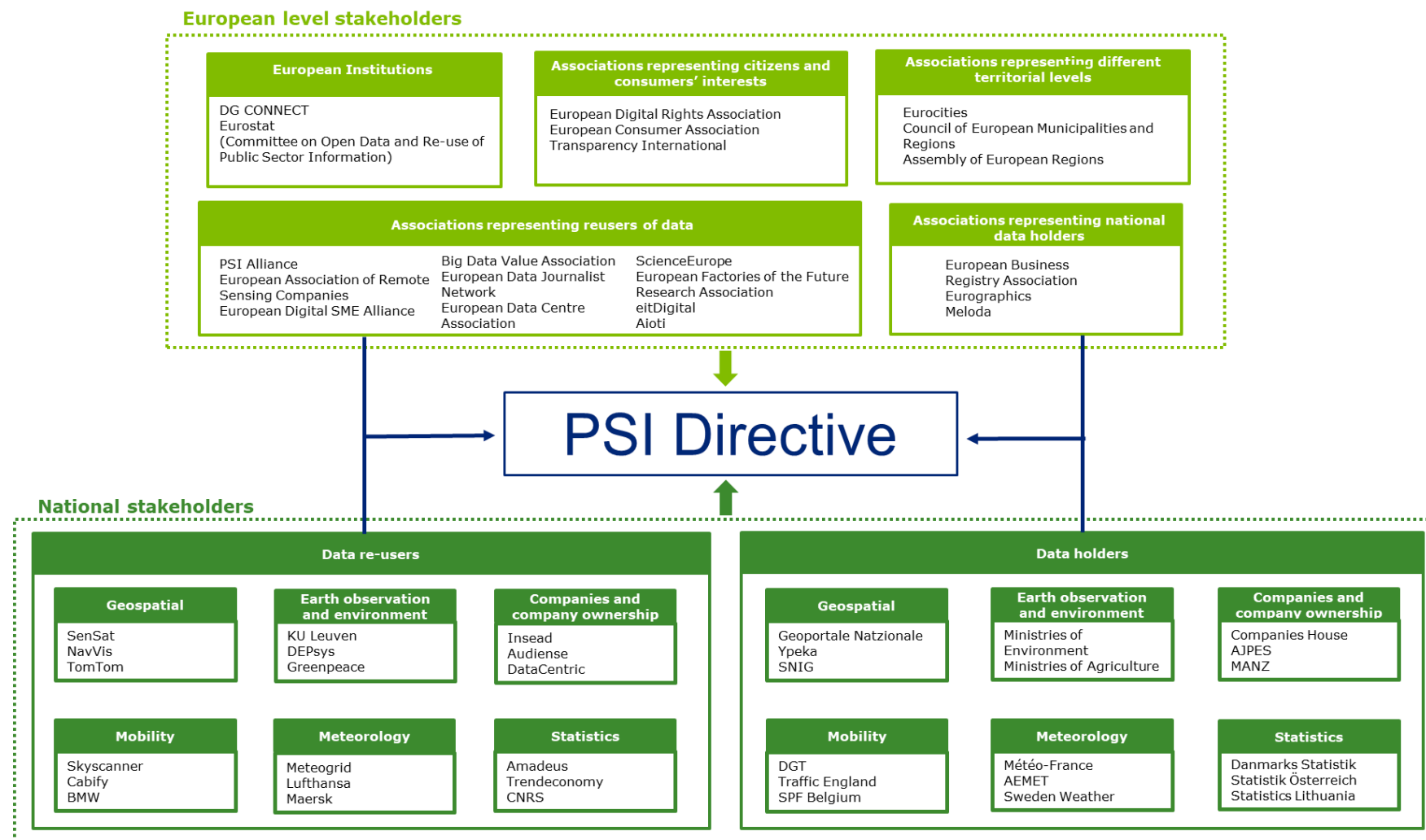
## Understanding and scope 2/3

### Stakeholders scope

We will carefully consider the two levels of stakeholders:

- **European level stakeholders:** European Institutions, EU level associations of different types.
- **National level stakeholders:** data reusers and data holders in particular.

Data collection activities will cover both levels and we will ask for your help for **national data collection activities**.





# Objectives and understanding of the assignment

## Understanding and scope 3/3

- **All European Member States are in scope of this assignment** and the general data collection tools will cover all of them.
- **10 Member States will be analysed more in-depth** for one task only (Task 4) and for developing and testing our economic impact model (which will nonetheless cover the entire EU).
- Our suggestion of “shortlisted countries” is based on the following criteria:
  - *Geographical position*
  - *Data economy value (in absolute term) and existence of an AI strategy*
  - *Maturity in terms of provision of public sector information*
  - *Charging and data provision practices*

### Shortlisted countries for Task 4 and data collection coordinators for this task

- **Denmark:** Ton Zijlstra
- **France:** Laura Doumbouya
- **Germany:** Eike-Christian Koring
- **Ireland:** Lawrence Kay
- **Italy:** Martina Barbero
- **Hungary:** Sandor Barnabas
- **Netherlands:** Ton Zijlstra
- **Romania:** Cristina Moise
- **Spain:** David Osimo
- **Sweden:** Carolin Moeller



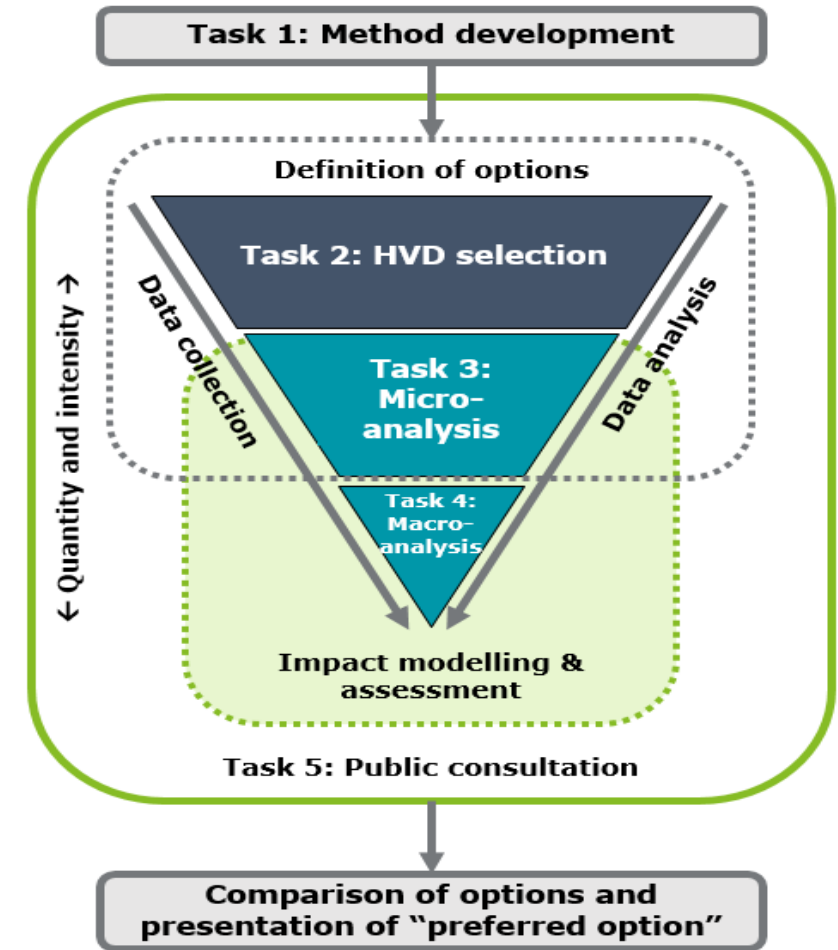


# **Methodological approach for the data collection and analysis**

# Methodology and state of play

## Our general approach and key challenges

- The project should not focus on “listing” High Value Datasets only but rather on **characterising them**
- There are 2 major challenges:
  - No *complete list* of datasets owned by governments;
  - Datasets have no *intrinsic value* (see Task 1)
- Our approach is therefore based on three pillars:
  - **Stakeholder involvement** (through interviews and focus groups) → stakeholders will help determining the value of datasets.
  - **Evidence based** → desk research will help establishing a good framework for identifying HVD and characterising them.
  - **Funnel approach** (for Task 2 and Task 3) → the iterative process supports DG CNECT in getting through the various phases of discussion.

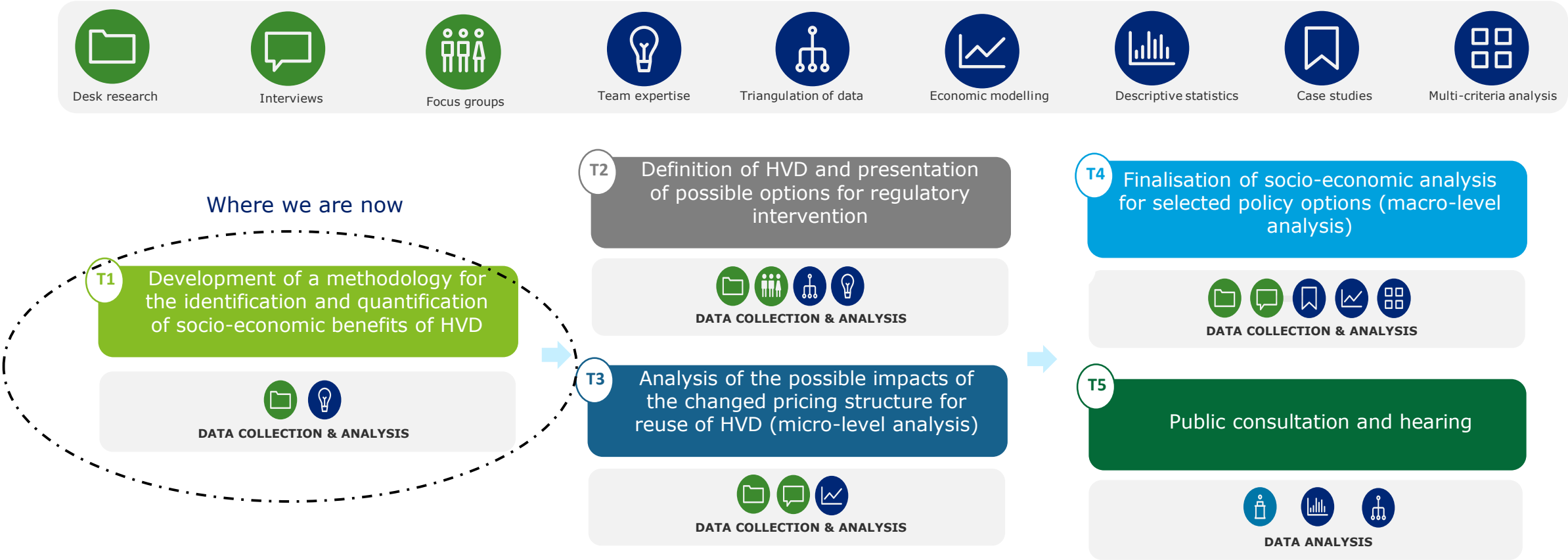


# Methodology

## Overall methodological data collection and analysis tools

### DATA COLLECTION TOOLS

### DATA ANALYSIS METHODS





# Methodology

## Task 1 – Provide a methodology to identify and quantify the socio-economic benefits of High Value Datasets

We focus the framework on the **6 macro characteristics of High Value Datasets** which are included in the PSI Directive definition:

- *Potential* to generate social benefits
- *Potential* to generate economic benefits
- *Potential* to generate environmental benefits
- *Potential* to generate innovation/AI
- High *potential* for reuse
- *Potential* to improve/strengthen/support public services and public authorities

For each of these aspects we identify specific value characteristics and for each characteristics possible quantitative and qualitative indicators as well as data sources.

Today we have a draft framework with **6 dimensions, 32 value characteristics and over 100 (qualitative and quantitative) possible indicators.**

**Based on desk research and stakeholders interviews we have identified a few datasets for each thematic area and we are testing them against this framework.**



# Methodology

## Task 2 – Define the specific High Value Datasets falling under the thematic categories –

- We will provide a detailed **description** of the structure of each of the datasets included in the list developed under Task 1 (see table)
- We will follow an **iterative approach and involve stakeholders** to a) validate the shortlist of datasets, b) define the datasets and the publications options, c) validate the final version of datasets and publications options,
- This task will be completed through four activities:
  - *Activity 2.1: Assessment of the datasets and identification of the high value ones;*
  - *Activity 2.2: First description of datasets and options;*
  - *Activity 2.3: Validation with stakeholders and intermediary release;*
  - *Activity 2.4: Final definition of datasets and options.*
- The main data collection tools for this task will be **a) desk research, b) interviews and c) focus groups.**

We will ask your support for this task!

Openness-data specification	License
	Format
	Machine-readability
Documentation	Metadata catalogue (adherence?)
	Metadata (dataset content description)
	Contact info
	Documentation
	Documentation (dataset creation and history)
	Shared vocabularies
Sustainability	Traceability
	Timeliness
	Primacy (Primary sources?)
Distribution vs. Release?	Coverage (regional/national/global)
	API / WebService
	automatic distribution in different formats

# Methodology

## Task 3 – Analysis of the possible impacts of the changed pricing structure

- Task 3 includes the analysis of **three aspects**:
  - **Cost benefit analysis** and impact on public bodies' budget (of making the list analysed in Task 2 High Value Datasets): will be analysed through a *selection of case studies covering individual data providers organisations at the national level*.
  - **Distortion of competition** (if the list analysed in Task 2 is made HVD): will be analysed through *interviews with relevant stakeholders at the national level* and legal analysis.
  - **Costs for public sector bodies and public undertakings** and impact on publication arrangements (if the list analysed in Task 2 is made HVD): will be analysed through *interviews with relevant stakeholders* at the national level.
- Task 3 will run in parallel to Task 2 and both will follow an iterative approach.
- Task 3 looks at the « micro-level » (by considering the **costs and benefits for PSBs and reusers**) while contributing to the macro-level analysis of the socio-economic benefits (for Task 2 and Task 4).

We will ask your support for this task!



# Methodology

## Task 4 – Financial and socio-economic analysis at the macro-level and selection of preferred option

### Interviews concerning impact modelling

- Purpose: Collect quantitative information for the financial and socio-economic analyses, and validate current approach and assumptions
- Step 1: Identify relevant stakeholders such as ministry officials, data holders, and data users in the Member States
- Step 2: Conduct 30 interviews in 10 Member States

### Final workshop

- Purpose: Present the final results of the study to a selected group of stakeholders for validation
- Step 1: Identify workshop participants and speakers and decide with DG CNECT
- Step 2: Provide background information for participants
- Step 3: Conduct the workshop and prepare a summary report

### Impact modelling

- Purpose: Identify the financial and socio-economic impacts of the policy options
- Step 1: Define the key economic, social, and environmental impacts to be considered
- Step 2: Conduct a financial analysis concerning costs and benefits for public bodies, businesses, citizens
- Step 3: Conduct an analysis concerning socio-economic benefits and government efficiency

### Multi-Criteria-Analysis

- Purpose: Rate the policy options vis-à-vis the baseline
- Step 1: Carry out a qualitative comparison of the different options (incl. the baseline scenario)
- Step 2: Fine-tune the assessment in an iterative manner, liaising closely with DG CNECT
- Step 3: Present the results of the analysis concerning the impacts of the policy options in a “user-friendly” format



# Methodology

## Task 5 – Public consultation and public hearing

The team will support DG CNECT in the organisation of **the public consultation and public hearing** and especially in the following activities:

- **Public consultation**

1. Analysis of the public consultation results

- **Public hearing**

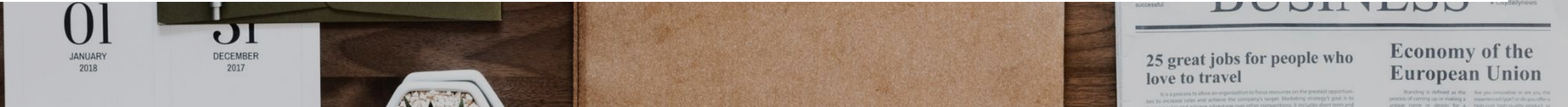
1. Definition of the scope and contents of the public hearing
2. Organisation and logistics
3. Analysis and summary of the outcomes





CALENDAR						
SUN	MON	TUE	WED	THU	FRI	SAT
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27

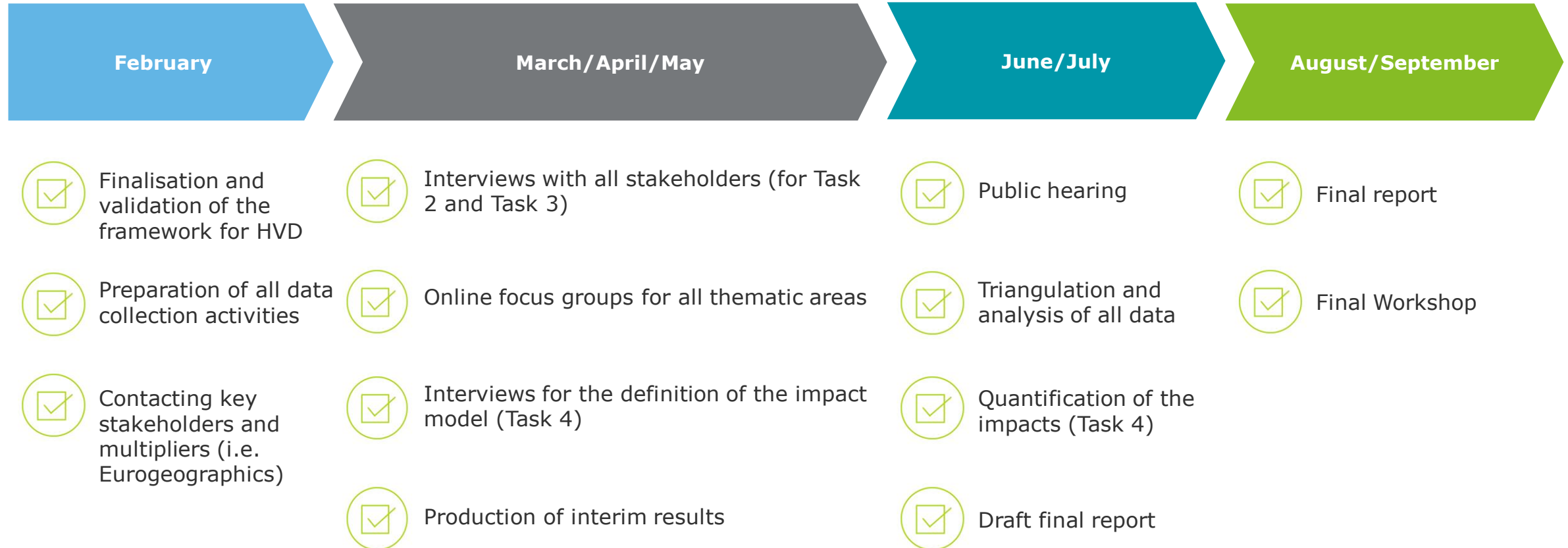
Timeline and planning and where/when we will need your help





# Timeline and planning

## Key milestones for the assignment



# Where/when we will need your help

## How we will interact with you

The **study team** will contact you all in a couple of weeks to ask about:

- **Whom should we speak to**, in your respective countries, when looking at the geospatial thematic area (including yourself of course);
- **Possible cases studies for costs and benefits analysis at organisations level** (i.e. if one of you has recently stopped charging for certain geospatial datasets we are particularly interested in hearing from them);
- **Other national stakeholders we should speak to** in order to estimate the impacts (i.e. national ministries, re-users associations at the national level).

*Francesca De Chiara and the study team will then organise the interviews with the relevant stakeholders.*

All along the study, the core team will remain at your entire disposal for any question.



### Main contact point of the core team:

- **Martina Barbero**  
(mabarbero@Deloitte.com)
- **Laura Doumbouya**
- (ldoumbouya@Deloitte.com)





**ADDITIONAL  
QUESTIONS?  
REMARKS?**

---



**Thank  
you!**



Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited ("DTTL"), its global network of member firms, and their related entities. DTTL (also referred to as "Deloitte Global") and each of its member firms are legally separate and independent entities. DTTL does not provide services to clients. Please see [www.deloitte.com/about](https://www.deloitte.com/about) to learn more.

Deloitte is a leading global provider of audit and assurance, consulting, financial advisory, risk advisory, tax and related services. Our network of member firms in more than 150 countries and territories serves four out of five Fortune Global 500® companies. Learn how Deloitte's approximately 286,000 people make an impact that matters at [www.deloitte.com](https://www.deloitte.com).

