



The journey of data for AI – and what the EU is doing to permit more travel

Malte Beyer-Katzenberger – European
Commission

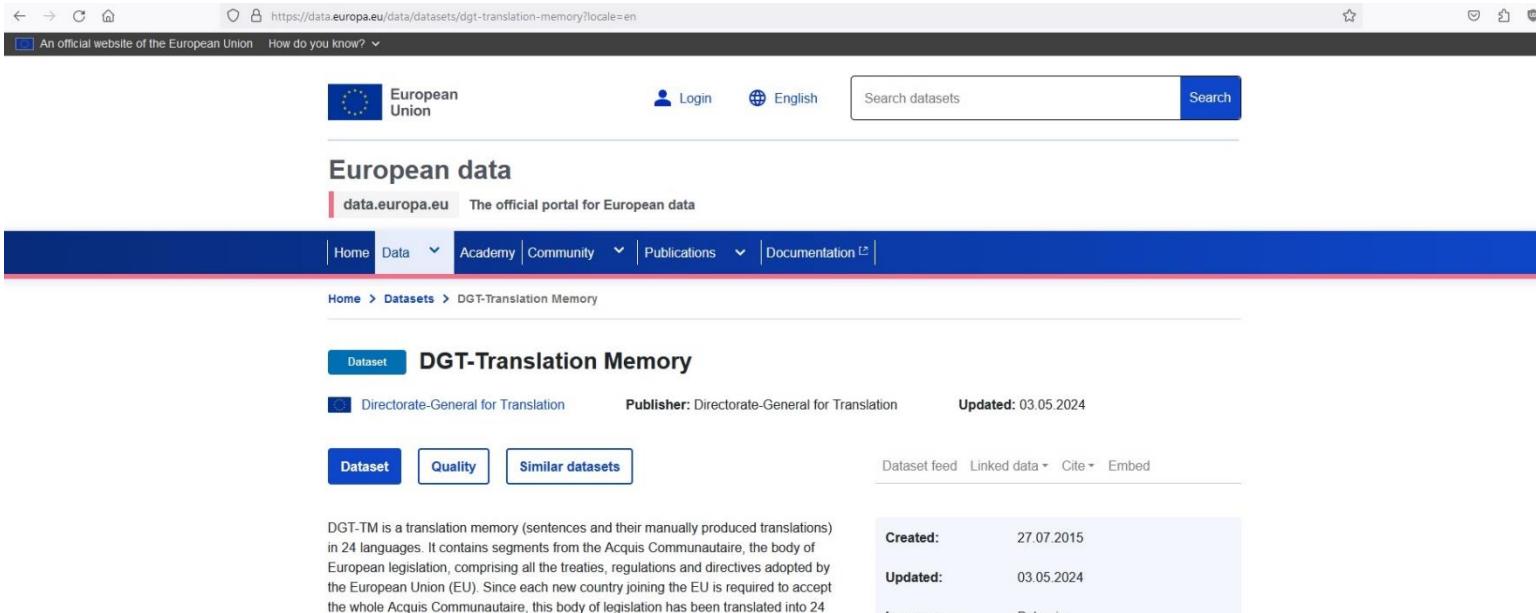
«AI firms will soon exhaust most of the internet's data»



European
Commission

What has the EU done to open more data sources?

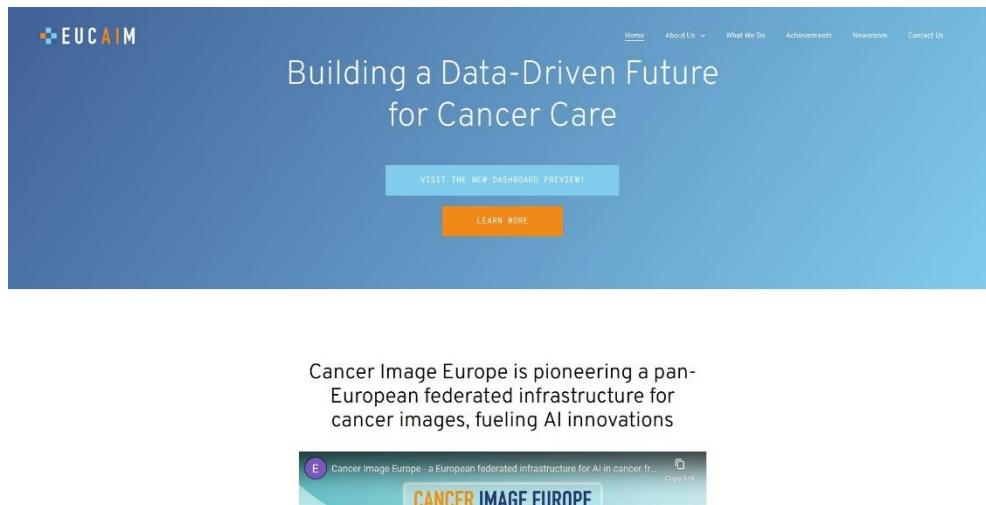
1) Providing certain data as „open data“



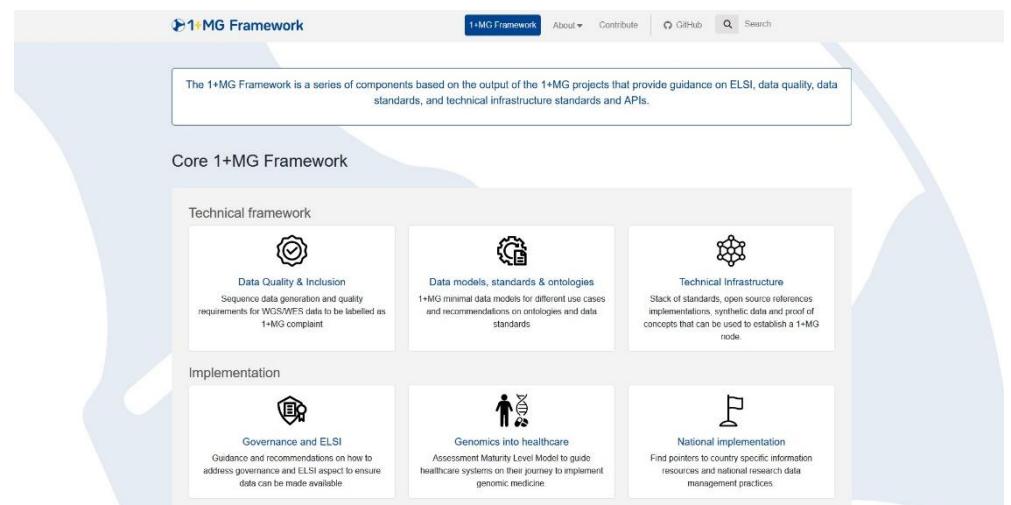
The screenshot shows a web browser displaying the data.europa.eu website. The URL in the address bar is <https://data.europa.eu/data/datasets/dgt-translation-memory?locale=en>. The page title is "European data" and the sub-page title is "data.europa.eu | The official portal for European data". The main content area shows a dataset titled "DGT-Translation Memory". The dataset is published by the "Directorate-General for Translation" and was last updated on "03.05.2024". The dataset is described as a "translation memory (sentences and their manually produced translations) in 24 languages. It contains segments from the Acquis Communautaire, the body of European legislation, comprising all the treaties, regulations and directives adopted by the European Union (EU). Since each new country joining the EU is required to accept the whole Acquis Communautaire, this body of legislation has been translated into 24 languages". There are tabs for "Dataset", "Quality", and "Similar datasets". Below the dataset description, there are buttons for "Dataset feed", "Linked data", "Cite", and "Embed". A sidebar on the right shows the "Created" date as "27.07.2015" and the "Updated" date as "03.05.2024".

What has the EU done to open more data sources?

2) Actively federate selected data repositories



The screenshot shows the EUCAIM website. The header features the EUCAIM logo and a navigation menu with links to Home, About Us, What We Do, Achievements, Newsroom, and Contact Us. The main title is "Building a Data-Driven Future for Cancer Care". Below the title are two buttons: "VISIT THE NEW DASHBOARD PREVIEW!" and "LEARN MORE". A sub-section below the main title reads: "Cancer Image Europe is pioneering a pan-European federated infrastructure for cancer images, fueling AI innovations". At the bottom, there is a footer with the EUCAIM logo and a "CANCER IMAGE EUROPE" button.



The screenshot shows the 1+MG Framework website. The header includes the 1+MG Framework logo and links for Home, About, Contribute, GitHub, and Search. A main text box states: "The 1+MG Framework is a series of components based on the output of the 1+MG projects that provide guidance on ELSI, data quality, data standards, and technical infrastructure standards and APIs." Below this is a section titled "Core 1+MG Framework" with two main categories: "Technical framework" and "Implementation". The "Technical framework" section contains three items: "Data Quality & Inclusion" (with a gear icon), "Data models, standards & ontologies" (with a gear icon), and "Technical Infrastructure" (with a brain icon). The "Implementation" section contains three items: "Governance and ELSI" (with a gear icon), "Genomics into healthcare" (with a DNA helix icon), and "National implementation" (with a flag icon).

What has the EU done to open more data sources?

Working on overcoming the costs of collection where there is in principle the possibility to use certain data:

3) Create obligation to have data access and permit bodies in all EU countries for a range of personal health data (and create a right to use): European Health Data Space legislation

PRESS RELEASE | 15 March 2024 | Brussels | 1 min read

Commission welcomes political agreement on European Health Data Space

Page contents

[Top](#)

[Quote\(s\)](#)

[Print](#)

The Commission welcomes the political agreement reached today between the European Parliament and the Council of the EU on the European Health Data Space (EHDS) - one of the **central building blocks of a strong European Health Union**. The rules [proposed](#) by the Commission in May 2022 have two main aims:

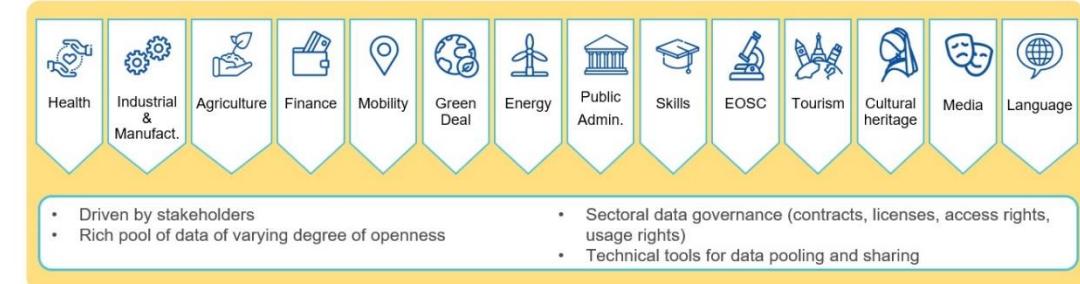
What has the EU done to open more data sources?

Working on overcoming the costs of collection where there is in principle the possibility to use certain data:

4) Support voluntary data sharing by building data sharing ecosystems (data spaces)

Common European Data Spaces

Digital Europe funding focuses on the creation of 14 data spaces, each linked to a strategic topic:



A Common European Data space brings together data infrastructures and governance frameworks

What has the EU done to open more data sources?

But who is gonna do me a „data-as-a-service“ inside such data spaces?

Do we need data fiduciaries/ data trusts

Can we push the legal frontiers on copyright and data protection?

**Should we?
Have we already?**

Can we knit the link between data spaces and AI factories closer?

Who would pay for data-pooling-as-a-service?

Summing up: Fishing for data



Crawl the web



Know your fish dealer
Become a processing
partner



Go to a fish market
(data space)

Thank you



© European Union 2024

Re-use of this presentation is authorised under the [CC BY 4.0](#) license with the exception of the images on slide 2, 4 and 10. The image on slide 2 is © by Mike Hadad for The Economist. Images on slide 4 are screenshots of the websites to which links are provided by clicking on the image.

