Experiences of working with Open Data

Miruna Bădescu, Eau de Web
IT Consultant to the EEA
Kyiv, 5 March 2019

Advices from an IT perspective

1. Start with what you have, improve the quality of datasets over time
2. Don’t do this alone
   - Interinstitutional exchange of experiences and resources
   - Involve the public and potential beneficiaries of the datasets
   - Multilingual documentation and best practices are already available
3. The cost of opening up the data is usually not high, reusable solutions exist
4. Keep providing series of data, the end users will rely on it
   - Establish a frequency for each dataset and stick to it
5. The power of Open Data can be found in combining datasets from multiple domains
   - Use a national Open Data portal to list and categorise the datasets
6. Reward good results rather than punish lack of results

Metadata and classification

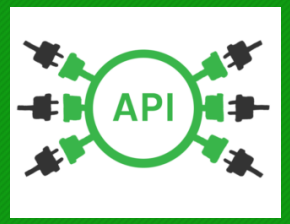
- The value of datasets not described by metadata is heavily decreased
- Several metadata standards are available, use:
  - Generic one to describe each resource (e.g. Dublin Core elements set) *and*
  - Specific standards for the specific domain
- Use multilingual glossaries to tag data
  - While the tagging is done in a national language, the datasets become widely discoverable with no extra-effort
- Classification of data from multiple perspectives
  - Using multilingual glossaries makes the tagging and finding the data easier
  - Extra-tags from the national strategy or international strategies (e.g. SDGs, Aichi Targets)
- Standardised classification allows data harmonisation and comparison
- Provide the codelists and vocabularies separately from the data

Formats - rule of thumbs

Machine-readable

Data format that can be automatically read and processed by a computer

Use open formats



Data needs to be manually extracted and processed

Don’t make it necessary to buy a license to read the data

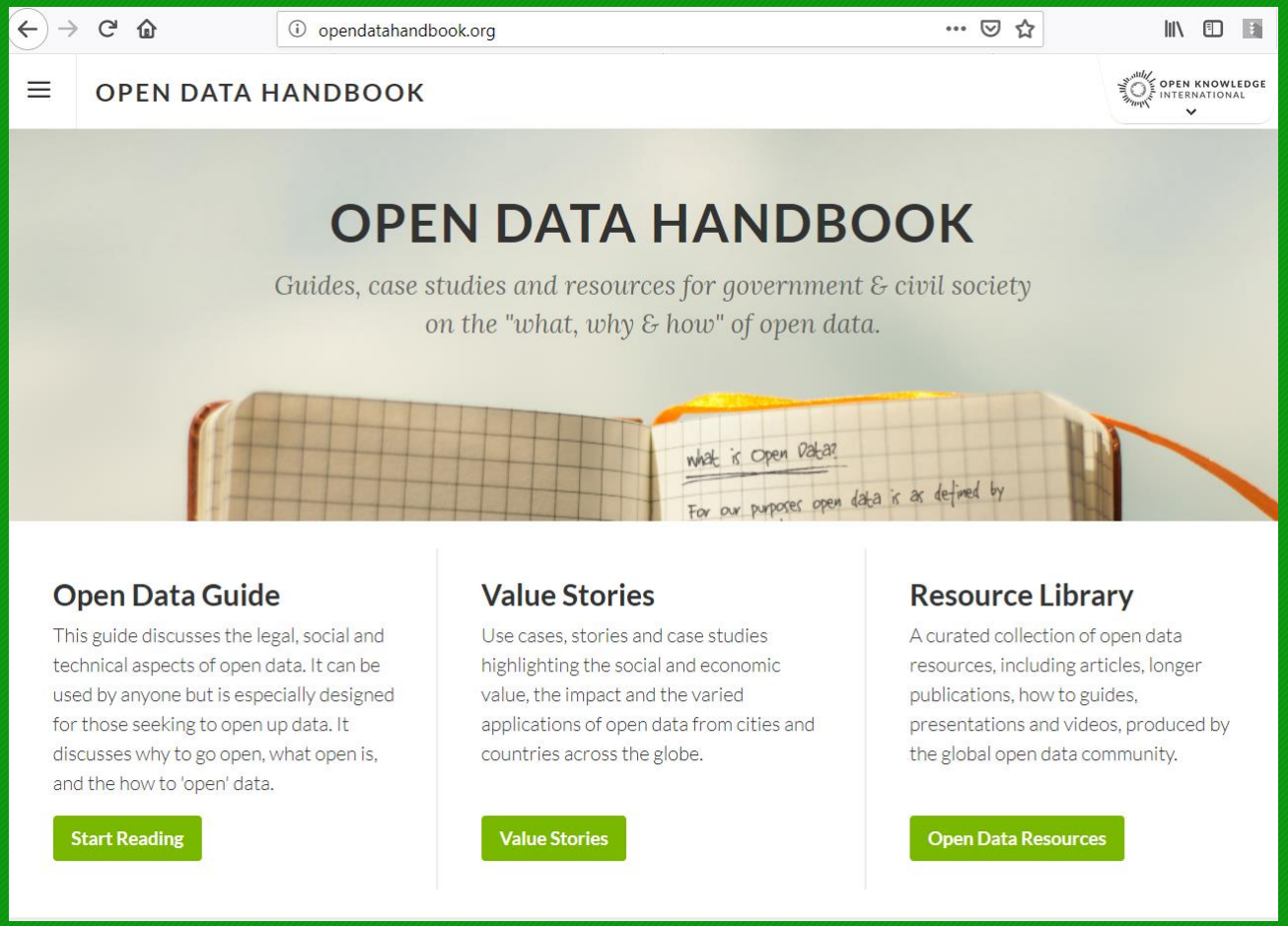
Scanned PDFs

Data restrictions – when to publish?

- Open by default non-sensitive data
- Don’t get stuck in data restrictions
  - Think in terms of what’s reasonable and change legislation to follow
  - When a dataset contains sensitive data, try to see if you can anonimise or cut that data, rather than holding back the entire dataset
- Look into licenses to facilitate reuse http://opendefinition.org/licenses/, https://opendatacommons.org/guide/

Get help

- http://opendatahandbook.org



Руководство по открытым данным

Languages: de el en es fr he hr id is it ja ko lt lv my ne nl_BE pt_BR ro ru zh_CN zh_TW

В настоящем руководстве обсуждают правовые, общественные и технические особенности открытых данных. Пользоваться им может каждый, но главным образом оно рассчитано на тех, кто хочет **открыть** данные. Оно отвечает на вопросы **зачем, как и что** — зачем и как «открывать» данные, и что значит открытость.

Вы можете начать со Введения. Справа есть панель содержания, с её помощью можно попасть в нужное место.

Мы будем рады любым замечаниям и в будущем учтём ваши пожелания. Мы также ждём дополнений и предложений дополнительных разделов и областей изучения.

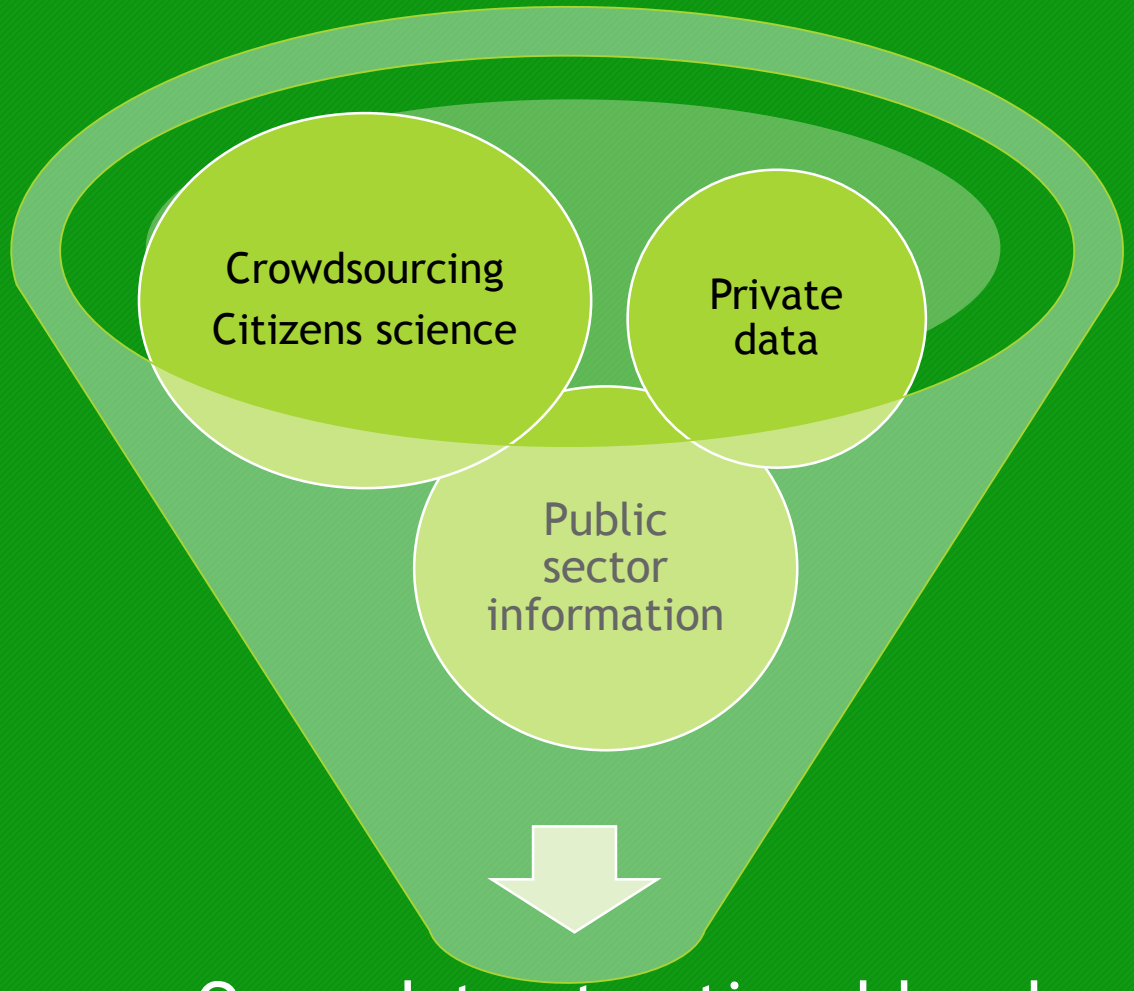
Оглавление

- Введение
- Почему Открытые данные?
- Что такое Открытые данные?
- Как открывать данные
- Итак, я открыл данные, что дальше?
- Глоссарий
- Приложения

Proving data and information upon request vs publishing it openly

- Don’t assume people know what to ask for
- The public might be reluctant to ask public institutions
- Asking for data is time consuming for both sides
- It is more costly for an institution on a long run to extract and provide customised data than to provide all non-sensitive data openly, on regular basis
- Involve citizens through research projects, involve specialised NGOs, encourage citizens science projects

Open Data Portal – fashion or need?



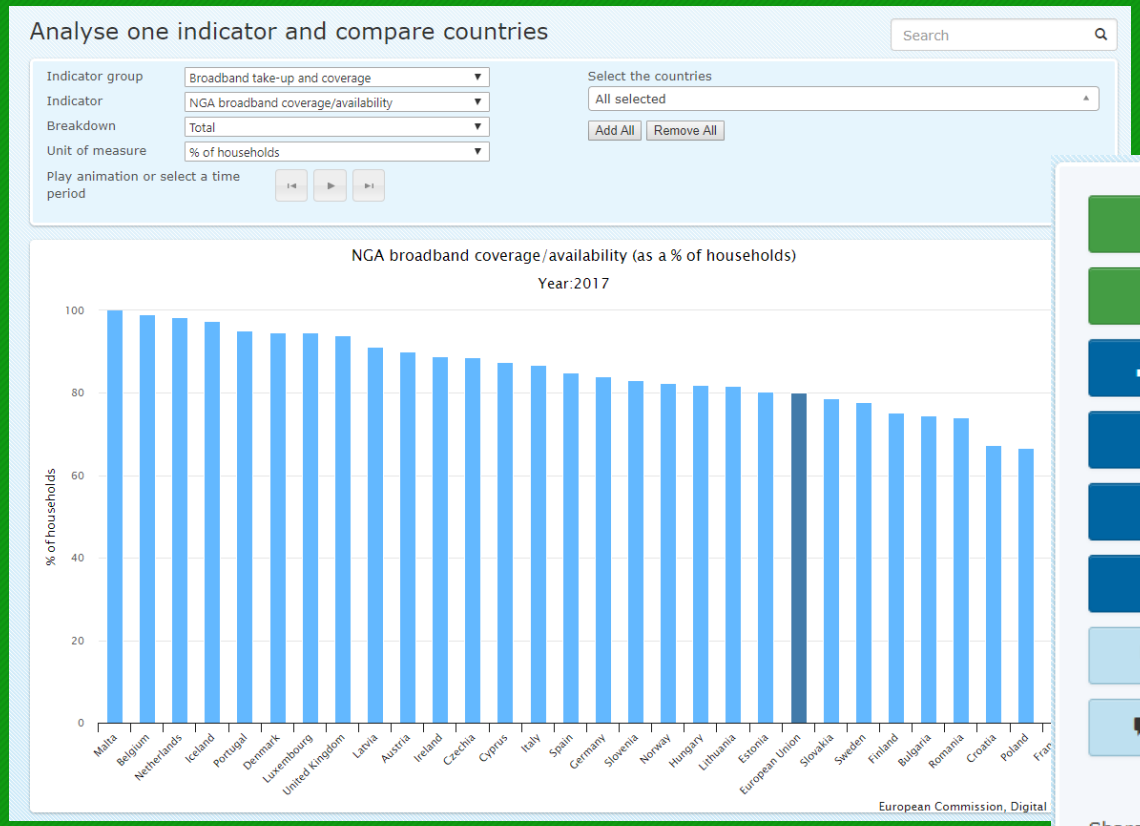
Open data at national level

- Central place to find open data from all domains in a country
- Different types of data sources can go into an open data portal
- Public sector information is just one
- Reuse existing open source software and make publish LinkedData

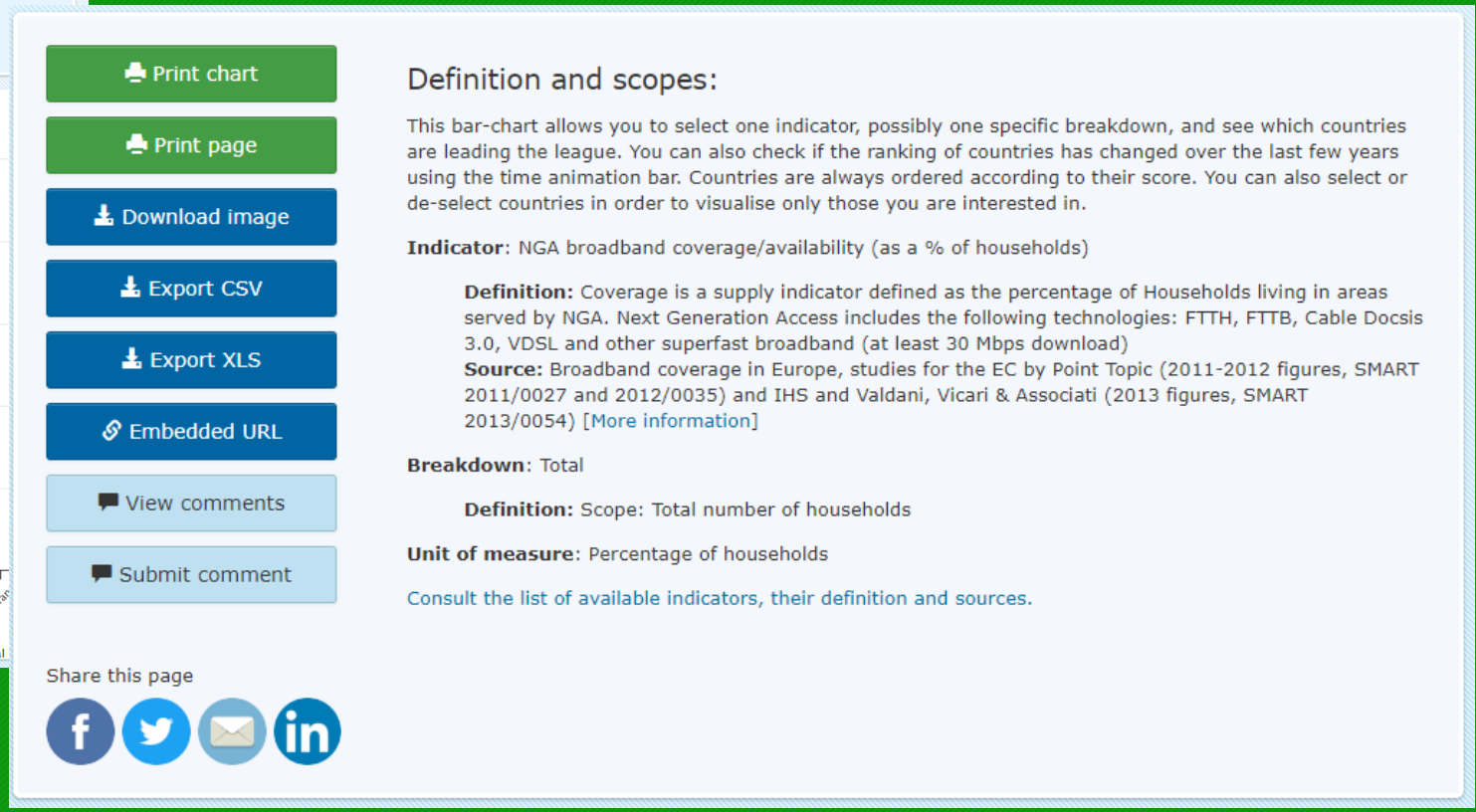
Good practice examples

Digital Agenda Scoreboard
https://digital-agenda-data.eu

Based on open statistical data about the European information society - meaningful interactive visualizations



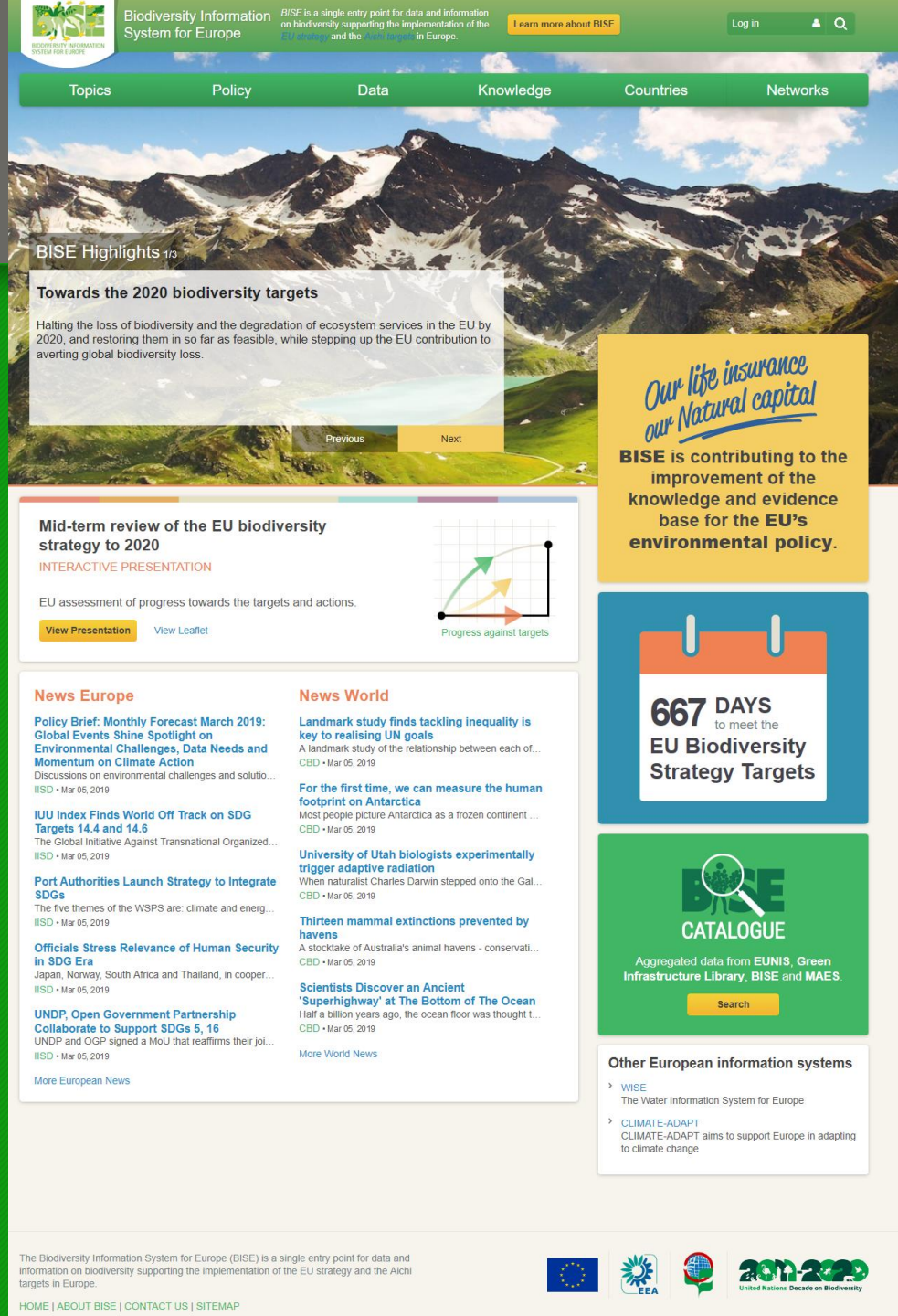
Data is offered for download, code lists downloadable separately, visualizations can be exported and reused



Data collected from Eurostat's Data portal

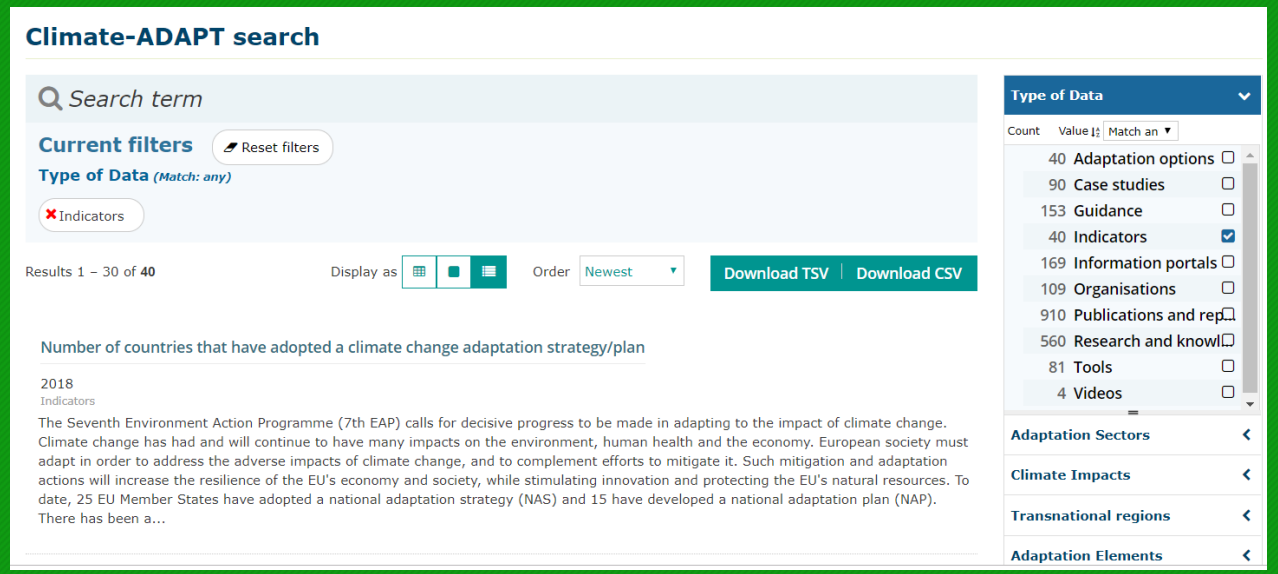
Biodiversity Information System
https://biodiversity.europa.eu

- DG Environment thematic portal on biodiversity
- Aside from locally produced data, harvests and collects data from the EEA's semantic database (LinkedData)
  - Displays indicators and graphs by topic (e.g. Green Infrastructure, Ecosystem services)
- Locally-produced content is harvested by the EEA semantic database

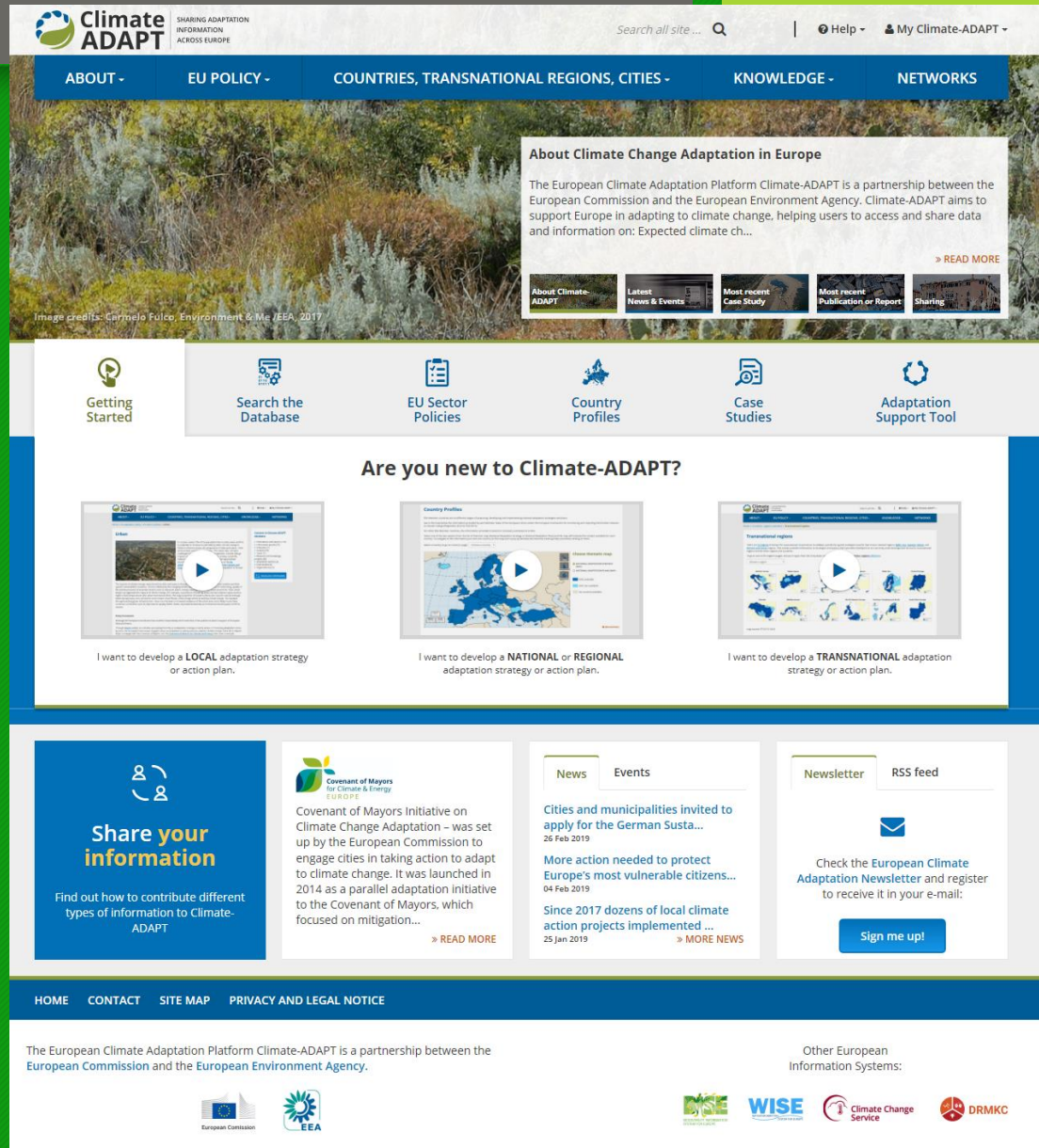


Climate-ADAPT portal
https://climate-adapt.eea.europa.eu

- DG CLIMA's platform dealing with Climate Adaptation issues
- Faceted search from multiple sources

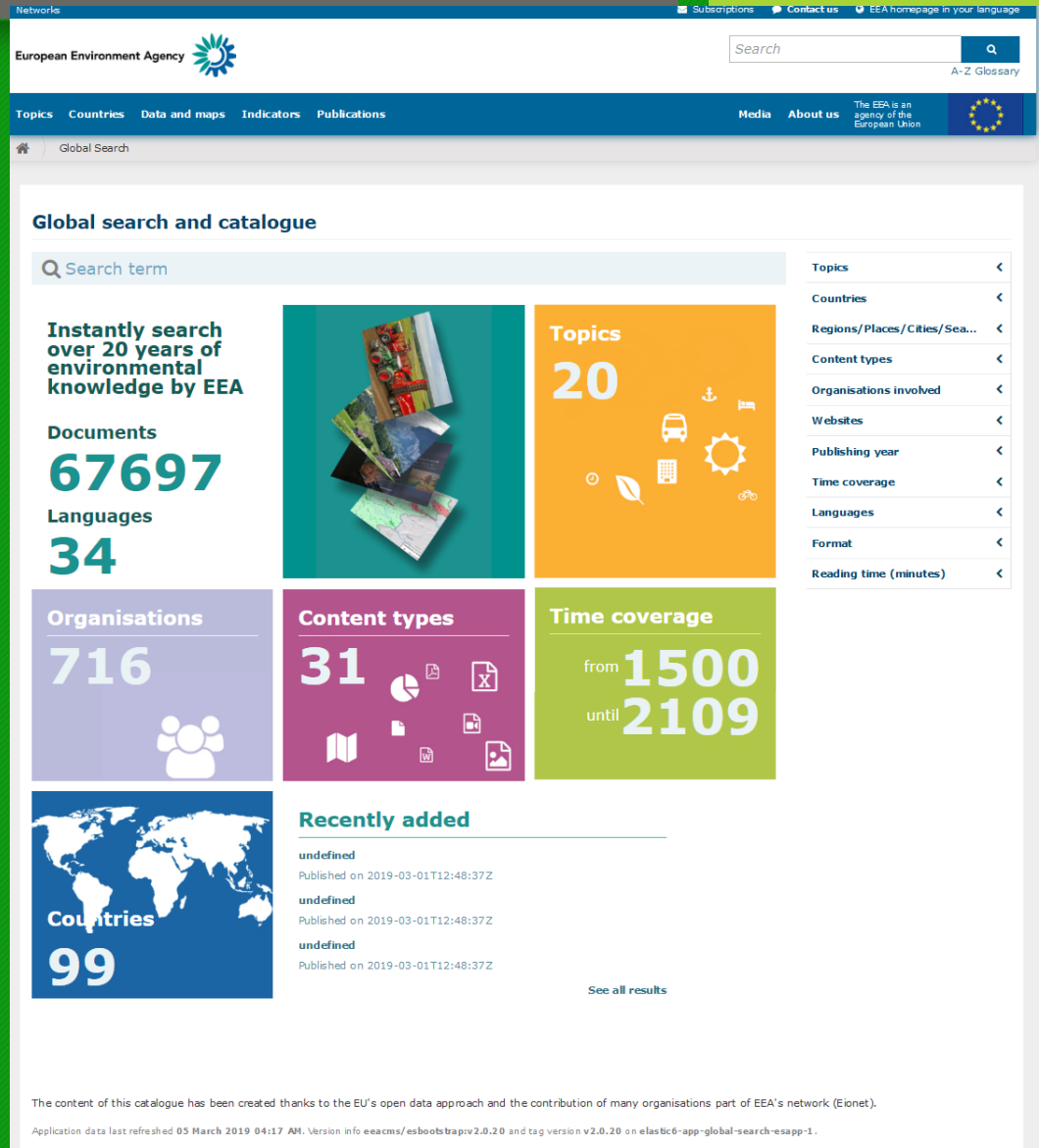


- Locally-produced content is harvested by the EEA semantic database

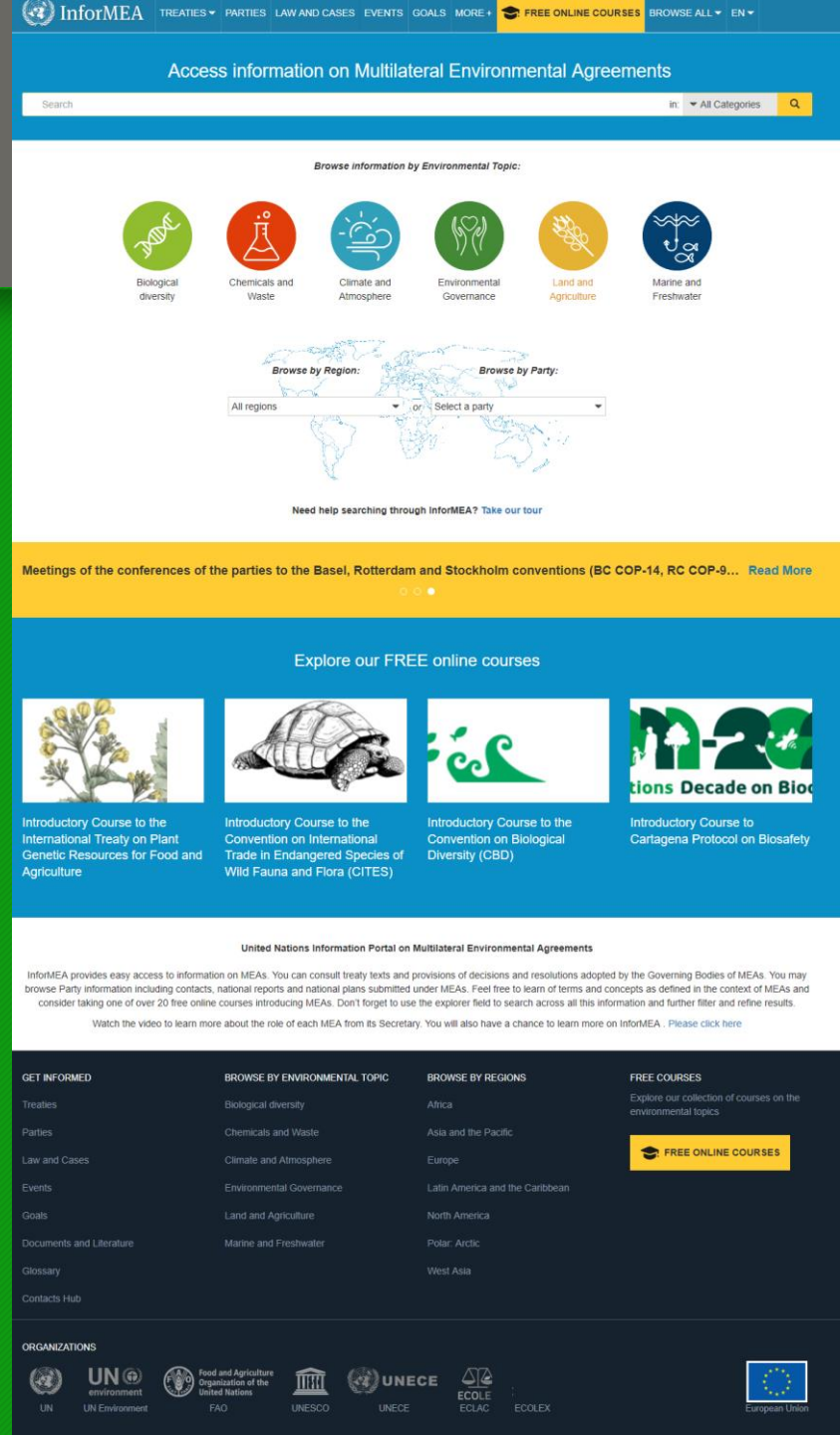


European Environmental Agency Search
http://search.apps.eea.europa.eu

- Searches in data created thanks to the EU's open data approach and the contribution of many organisations part of EEA's network
- All datasets are harvested from their sources and stored in a semantic database (LinkedData)
- Because the data was described by metadata, possible to filter by:
  - Topics
  - Countries /Regions/Places/Cities/Seas...
  - Content types
  - Organisations involved
  - Websites
  - Publishing year
  - Time coverage
  - Languages
  - Format
  - Reading time (minutes)



InforMEA

https://www.informea.org

- UNEP's platform that brings together legal data from different Multilateral Environmental Agreements
- Harvests data on Decisions, Resolutions, Documents, events, etc. from the MEAs via APIs
- Shows data classified by topic, Goals and Principles (SDGs, Aichi Targets, GEGs), region, period, type, country, etc.
- Faceted search with the same filters