On the Opportunities and Challenges of Transnational Crop Datasets

Maja Schneider
maja.schneider@tum.de

living planet symposium | BONN 23–27 May 2022
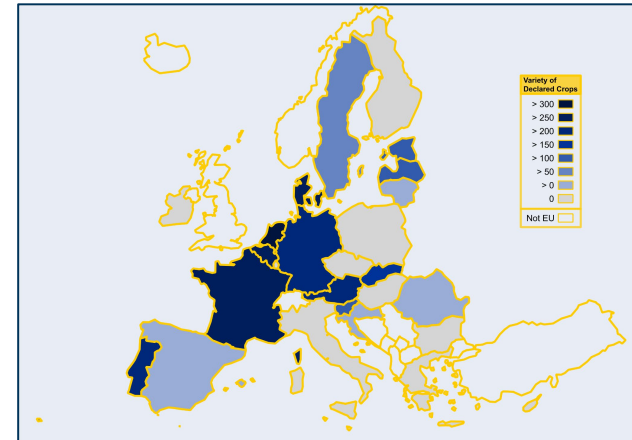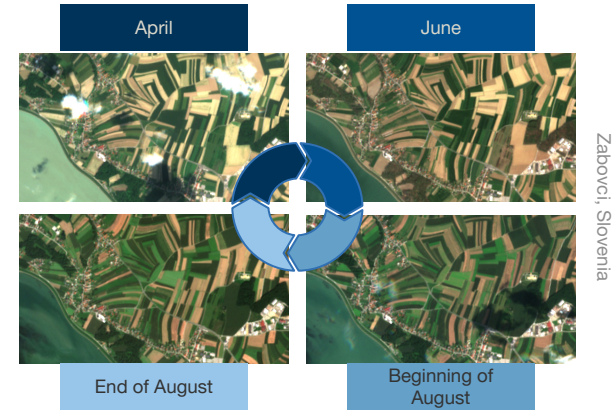
EuroCrops

Let's start with the data.

EuroCrops

- Pan-European dataset that combines and harmonises **georeferenced agricultural parcel data** with the corresponding **cultivated crop**
- Data is obtained by the countries **agricultural ministries** or **paying agencies**
- Contains the farmers' **self-declarations** which are collected within the subsidy control of CAP
- Currently including data from **16 countries**
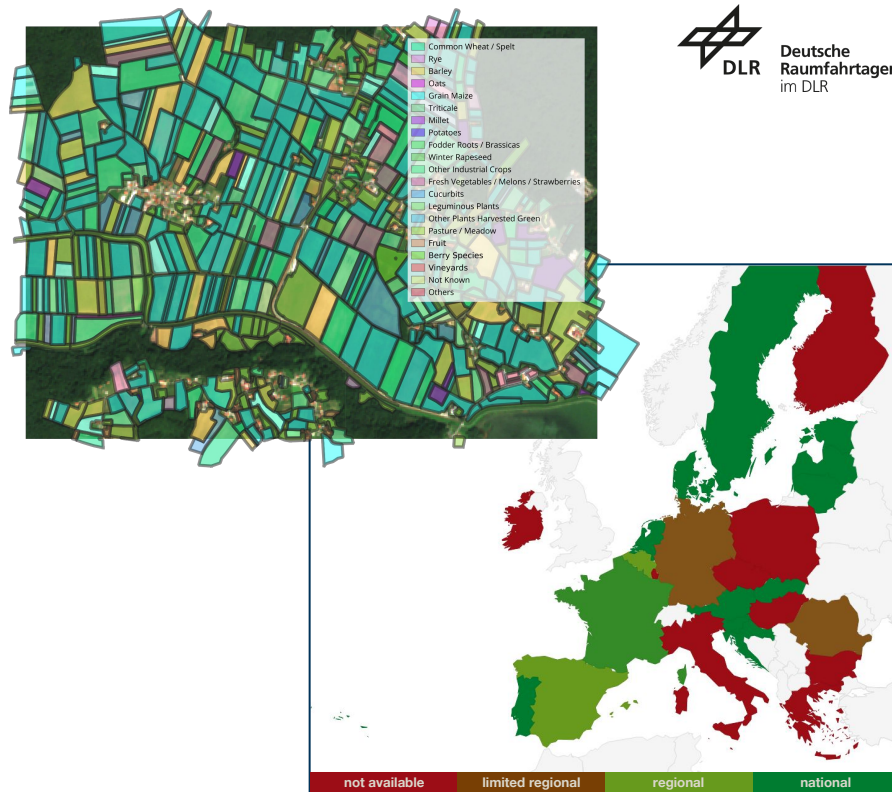
Motivation of the Project

- **Originally**:

  Use-Case for Copernicus data and

  Machine Learning

- **Now**:

  Broader impact into the directions of

  biodiversity and agriculture itself

April

June

End of August

Beginning of August

Zabovci, Slovenia

Variety of Declared Crops
> 300
> 250
> 200
> 150
> 100
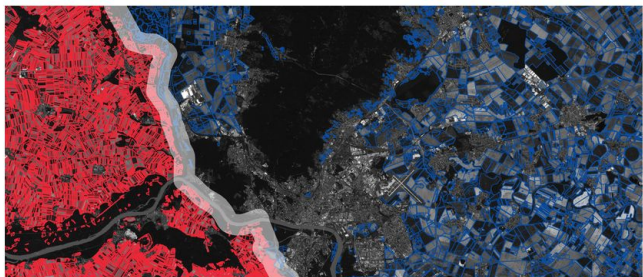> 50
> 0
0
Not EU

EuroCrops - Progress

- **Collection of data** over the past 1.5 years

- Development of new crop taxonomy due to
  - country dependent schemas
  - crop declarations in national languages

→ **Hierarchical Crop and Agriculture Taxonomy (HCAT)**
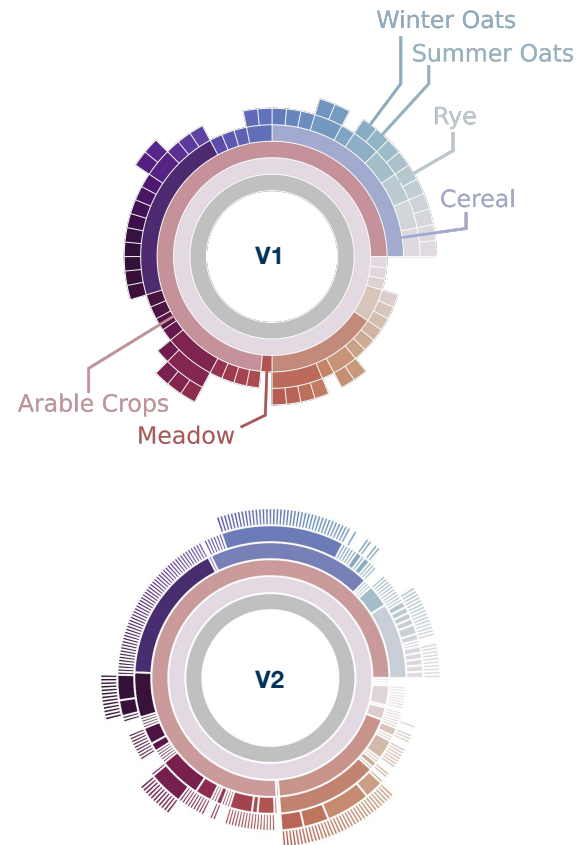
Publication of EuroCrops

- **Harmonised vector data**

- **Open access**

- Distribution link over **GitHub**

- Soon probably **GeoDB**

Border Region Austria - Slovakia around Bratislava

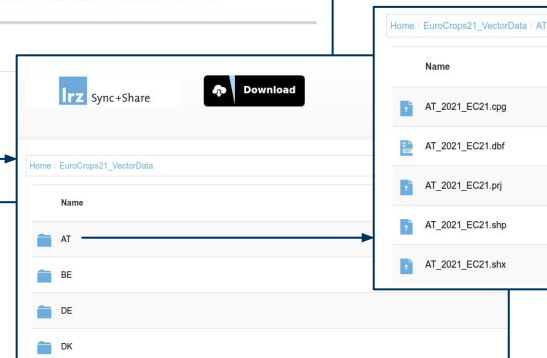**EuroCrops**

License CC BY 4.0

EuroCrops is a dataset collection combining all publicly available self-declared crop reporting datasets from countries of the European Union. The project is funded by the German Space Agency at DLR on behalf of the Federal Ministry for Economic Affairs and Climate Action (BMWK). This work is licensed under a Creative Commons Attribution 4.0 International License.

Right now EuroCrops only includes vector data, but stay tuned for a version that includes satellite imagery!

For any questions, please refer to our FAQs or use the Discussions/Issues to reach out to us.

Content

1. Background
2. Participating countries
3. GitHub project structure
4. Vector data download
5. Reference

Challenges

of transnational datasets

- Countries host their data on their **own platforms**
- Hosting responsibilities of certain data **differs** between countries → Different Authorities
- Not all platforms are fully available in **English**

Discoverability

TUM
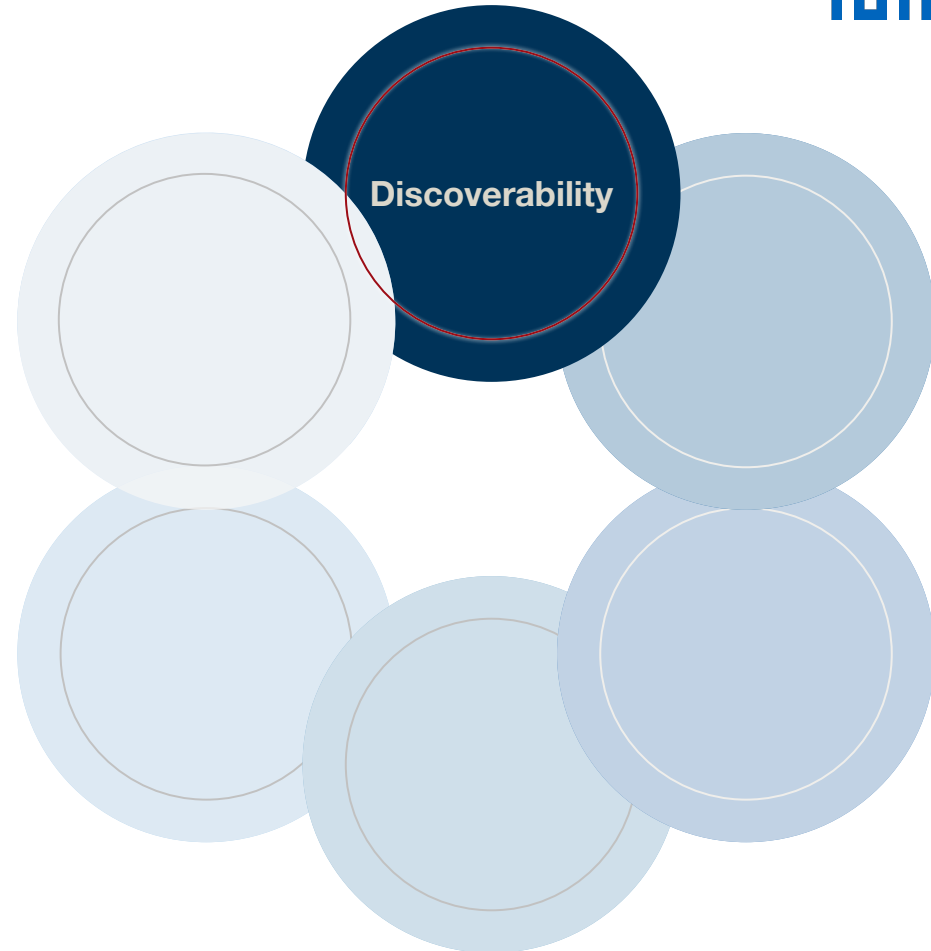
- Data is not always **freely**

  available due to

  - **GDPR**

  - Authorities don't see

    direct **benefit** from

    releasing data

  - **???**

- Data is **hidden** behind

  log-in/registration wall

Discoverability

Accessibility

- Countries release data in

   national-dependent

  - schemes

  - taxonomies

  - standards

  - language

- Huge variability in data
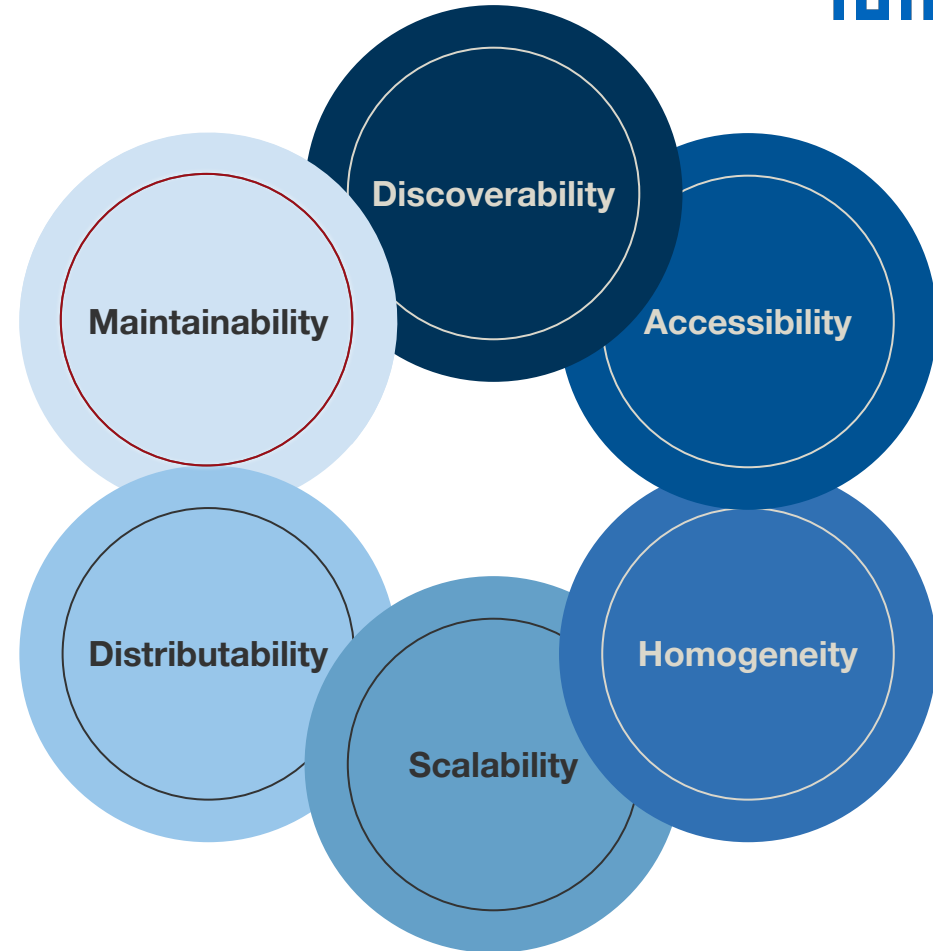
   formats (shp, gpkg, wms,...)

Discoverability

Accessibility

Homogeneity
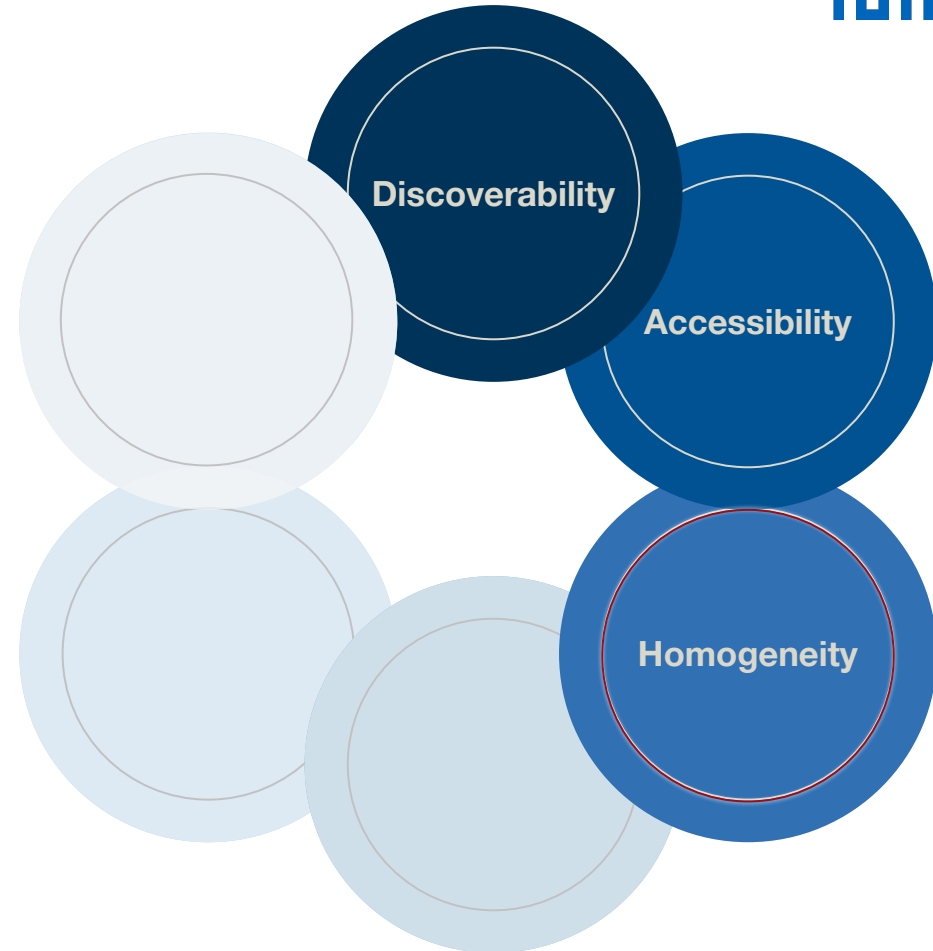
- **Extending** a dataset requires starting from the bottom again for each new partial dataset
- Countries **change**
  - website
  - taxonomy
  - format

  during years

Discoverability

Accessibility

Homogeneity

Scalability
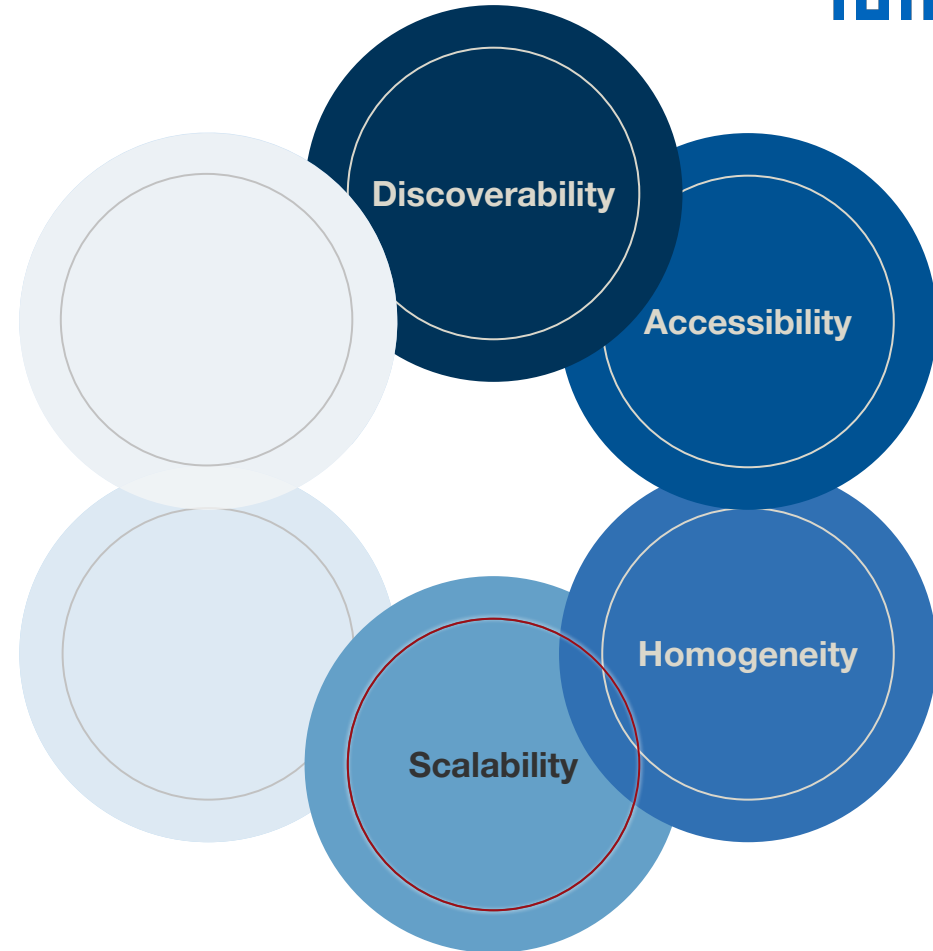
- Addressing all countries with dataset updates requires reaching out to each authority **individually**
- No established, **central point** of distribution
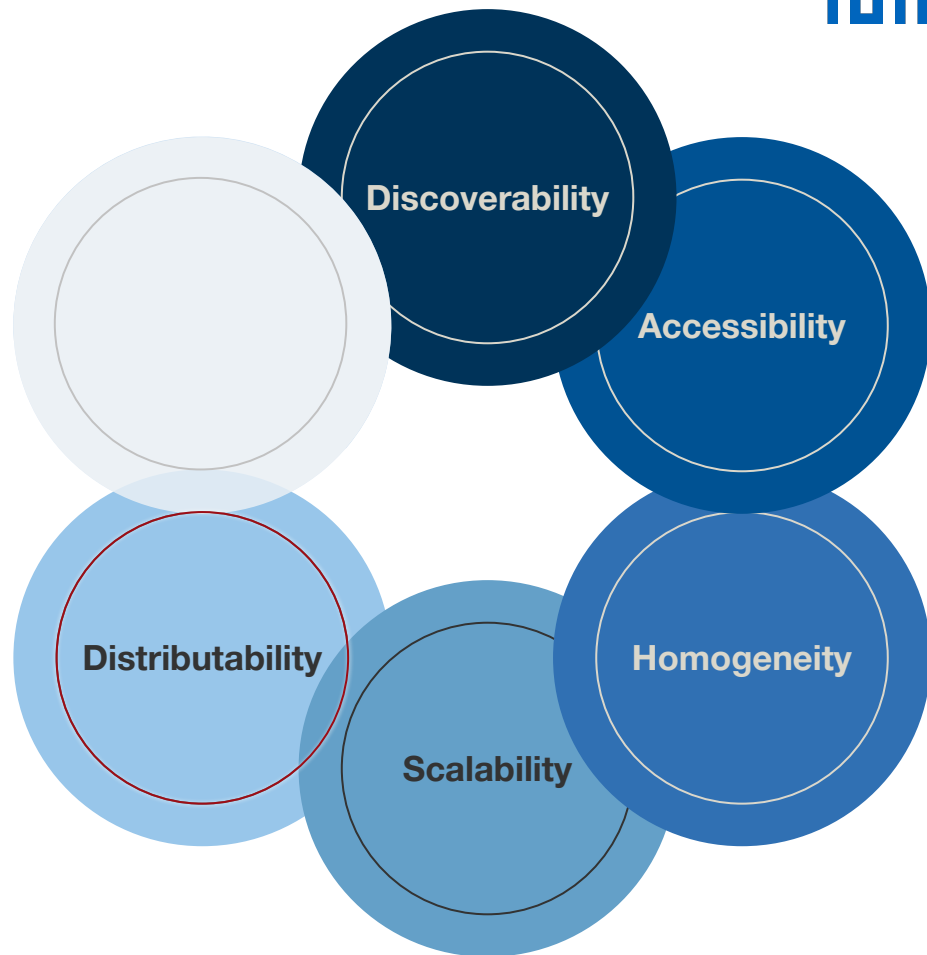
- Countries can **change**
  - policy
  - infrastructure
  - standards
- **Manual** checks required
- Impossible to pin down and correct **errors** in datasets
- Silent **updates**

Discoverability

Maintainability

Accessibility

Distributability

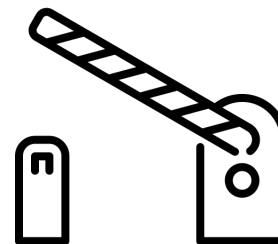Homogeneity

Scalability

Opportunities

of transnational datasets
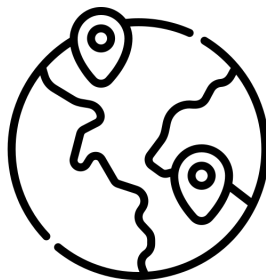
For the EO community

Opens the **domain**

EO data often **complicated**,
not easy to find as natural images and
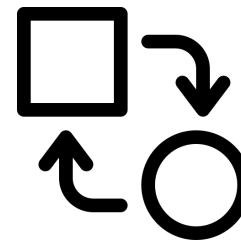**hidden** by multiple layers of abstraction

Lowers entry **barrier**

For Europe as a whole
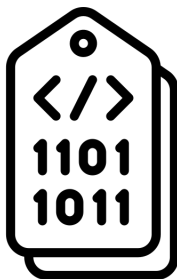
Inter-country **comparisons** and **aggregations**

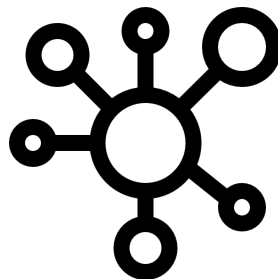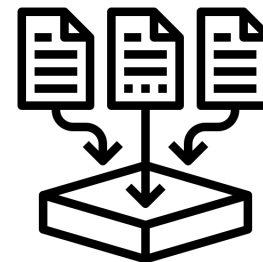Instead of encapsulated data and models,
everything works together

For ministries and authorities

All data together in one point,
version controlled,
with **meaningful metadata** instead of
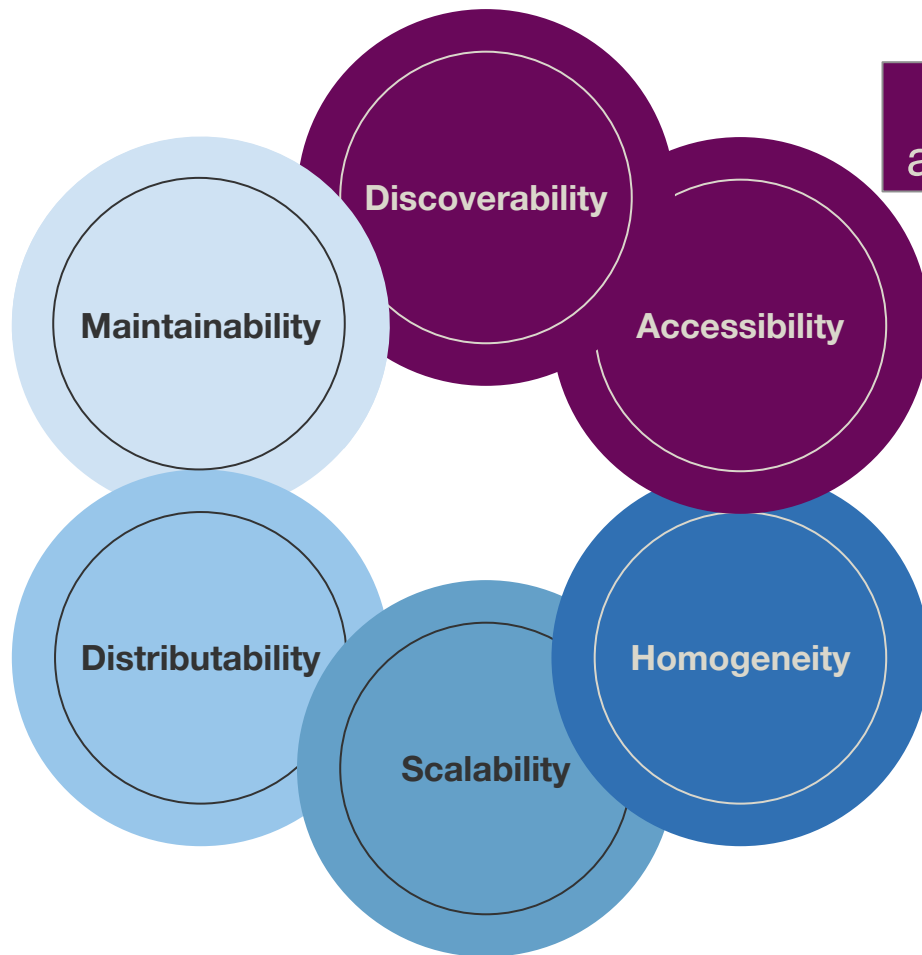multiple repos with silent updates

Easy to interact with the public
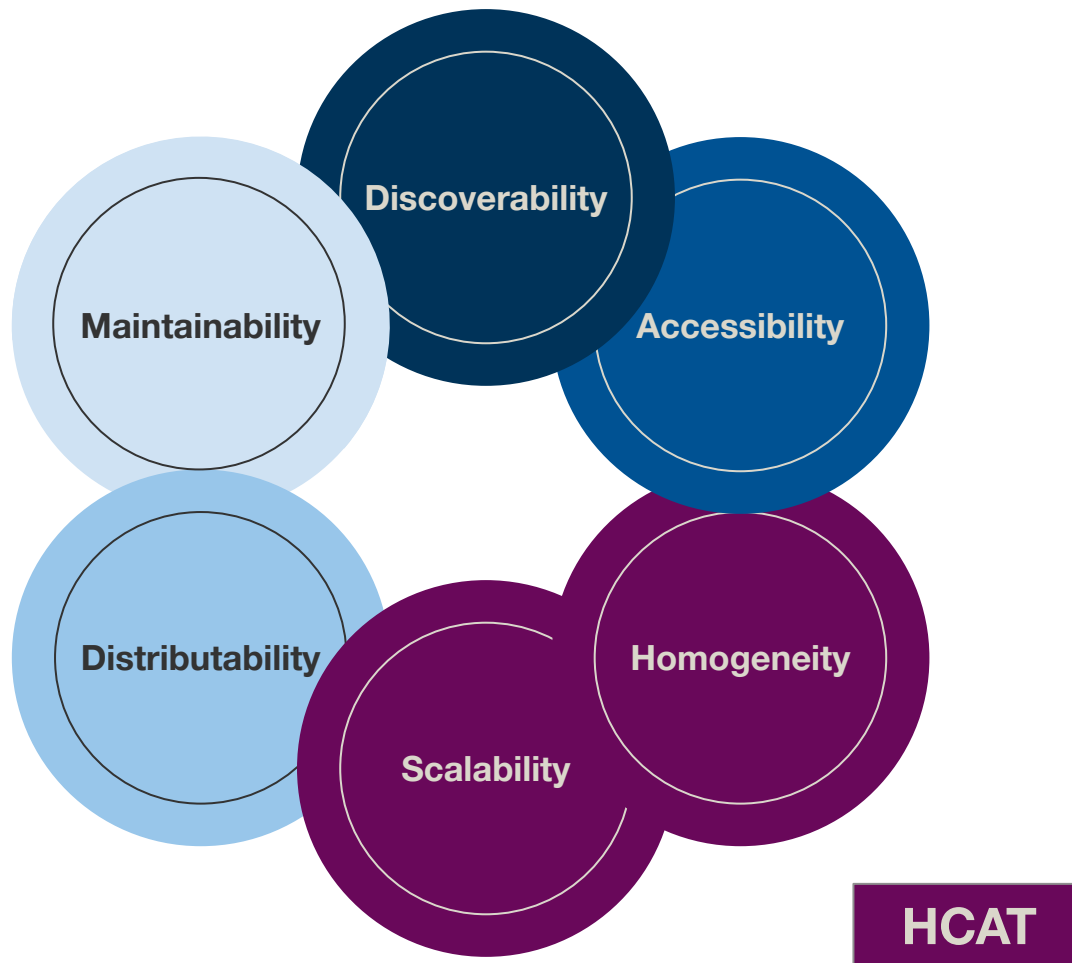because **one node** holds all
information

An established transnational
dataset builds a **baseline** for
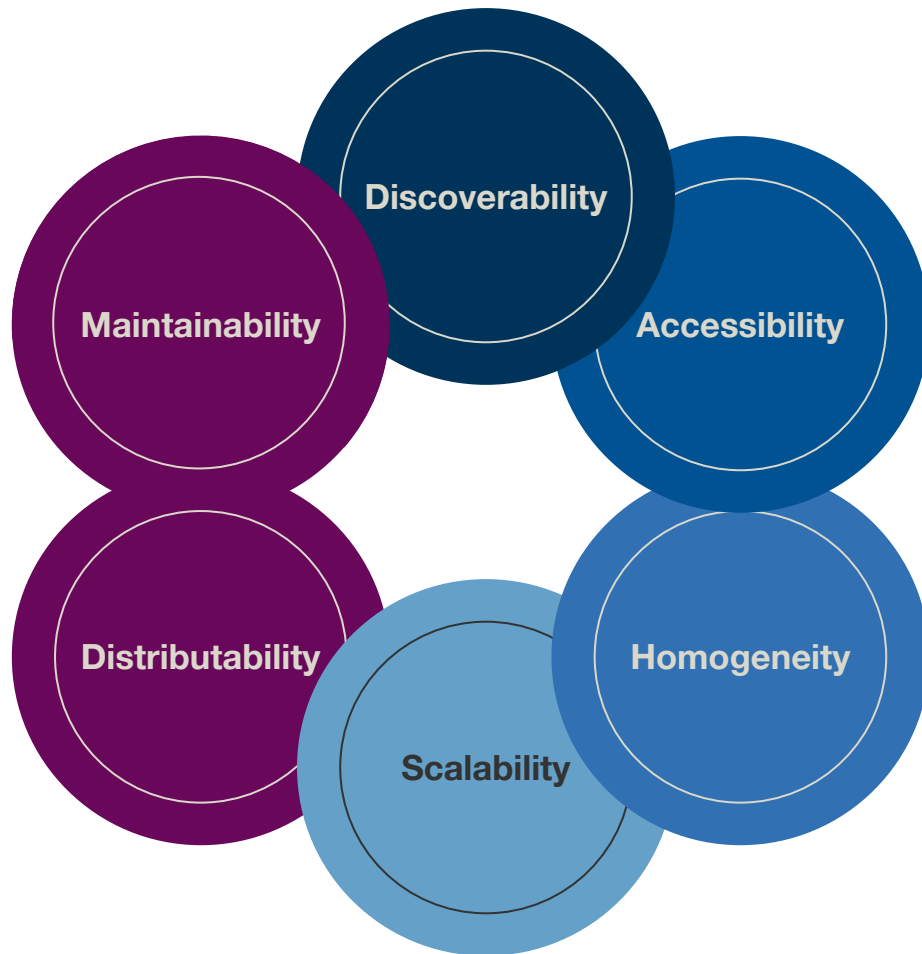other countries to be fitted in
easily and scaled up

Tackling the challenges with EuroCrops

One publicly available dataset

Discoverability

Accessibility

Maintainability

Distributability

Scalability

Homogeneity

HCAT

Building up a **community** and keeping **contact** with authorities

Discoverability

Maintainability

Accessibility

Distributability

Homogeneity

Scalability

Thank you!

WEB: www.eurocrops.tum.de
           https://github.com/maja601/EuroCrops
MAIL: maja.schneider@tum.de
TUM: www.asg.ed.tum.de/lmf/schneider/
LinkedIn: www.linkedin.com/in/maja-schneider/
Twitter: twitter.com/Maja4EO