A Deep Dive into Metadata Quality: Open Datasets Across Europe

Barbara Šlibar

Teaching Assistant, University of Zagreb, Faculty of Organization and Informatics, Croatia

EU OPEN DATA DAYS 2025, 19-20 MARCH 2025

Although the Open Data Movement Emphasises the Availability of Resources, this Does Not Imply That Those on the Demand Side will Successfully **Search**, **Discover**, or **Use** Available Resources!

EU OPEN DATA DAYS

METADATA
are data about the data.

DATA
are all content formats in digital form (text, image, audio, video, etc.) or physical form.

WHAT IS METADATA QUALITY, AND HOW IT CAN BE MEASURED?

Quality is a **multidimensional concept**, defined by Juran (1951, 2010) as "**fitness for use**" and later as "**fitness for purpose**".

Quality **cannot be captured** by a **single indicator/aspect**!

It is important to **include all relevant aspects** in order to determine **whether** and **to what extent** something **is of quality**.

foi

What
Has Been Done So Far, And Where Are The Gaps?

- To explore existing work and identify gaps.

What
Has Been Done So Far, And Where Are The Gaps?

- To explore existing work and identify gaps.

Why
Do These Gaps Need To Be Addressed?

- To improve metadata quality assessment for open datasets.

What
Has Been Done So Far, And Where Are The Gaps?

- To explore existing work and identify gaps.

Why
Do These Gaps Need To Be Addressed?

- To improve metadata quality assessment for open datasets.

How
Can These Gaps Be Addressed?

- By developing a composite indicator of metadata quality for open datasets.

PRACTITIONERS AND ACADEMICS HAVE DIFFERENT VIEWS ON THE ASSESSMENT OF THE QUALITY OF METADATA IN OPEN DATASETS!

CRITERIA FOR DEVELOPING THE COMPOSITE INDICATOR:

Based on automatically retrieved metadata.

Facilitate comparisons within and between portals, and over time.

Contain indicators that do not depend on data from a specific portal.

Contain indicators that produce the same score when applied to the same data.

Findability	Completeness	5 indicators
	Conformance	1 indicator
Retrievability	Completeness	2 indicators
	Conformance	1 indicator
	Retrievability	1 indicator
Interoperability	Accuracy	1 indicator
	Completeness	2 indicators
	Conformance	1 indicator
	Openness	2 indicators
Reusability	Completeness	5 indicators
	Conformance	1 indicator
	Openness	1 indicator
Contextuality	Completeness	7 indicators
	Conformance	1 indicator
	Timeliness	1 indicator

Benchmarking Countries Using a Composite Indicator of Metadata Quality for Open Datasets

1. Germany
2. Czechia
3. France
4. Spain
5. Austria
6. Italy
7. United Kingdom
8. Ukraine
9. Netherlands
10. Ireland
11. Poland
12. Sweden
13. Belgium
14. Switzerland
15. Bulgaria
16. Greece

17. Portugal
18. Norway
19. Slovenia
20. Slovakia
21. Finland
22. Denmark
23. Romania
24. Croatia
25. Serbia
26. Luxembourg
27. Hungary
28. Lithuania
29. Latvia
30. Cyprus
31. Moldova

31 countries meet the criteria of having **over 1,000** datasets on data.europa.eu.

Countries **ordered by the number of datasets** available on October 11, 2024, from highest to lowest.

Developed composite indicator applied to datasets, scores summarized by country!

EU OPEN DATA DAYS

Findability

is the extent to which humans and machines can easily discover (meta)data through unique and unambiguous identification, as well as information about the temporal and geographic area(s) covered by the data.

Findability scores per country

Legend:
- ○ : Avg. Score for a Country
- — : +/- SD per Country
- | : Overall Avg. Score Across All Countries
- ✕ : Lowest Avg. Score
- ✕ : Highest Avg. Score

Y-axis: Country (two-letter country codes)
X-axis: Findability scores

Countries (top to bottom): FI, BG, LT, CY, CZ, HR, GR, FR, RS, PL, NL, SE, LV, SI, GB, LU, MD, AT, RO, ES, PT, IE, UA, HU, NO, CH, BE, SK, DE, IT, DK

Retrievability
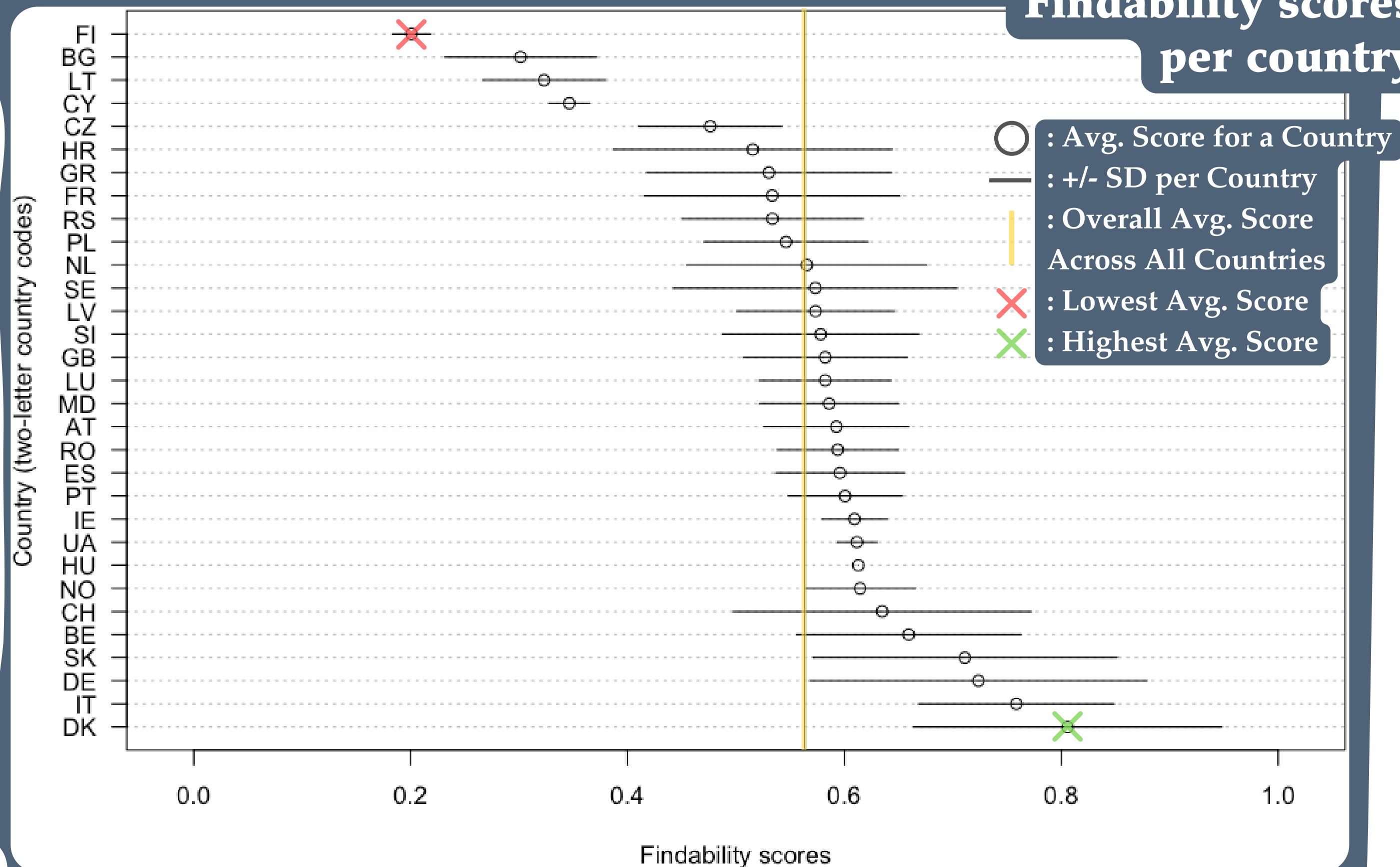
Retrievability scores per country

is the extent to which humans and machines can successfully fetch (meta)data.

Legend:
- ○ : Avg. Score for a Country
- — : +/- SD per Country
- │ : Overall Avg. Score Across All Countries
- ✕ : Lowest Avg. Score
- ✕ : Highest Avg. Score

Country (two-letter country codes): FI, CY, CZ, BG, NL, SE, NO, BE, SI, LV, HR, PL, SK, GB, PT, LT, MD, LU, DK, IT, RO, FR, RS, UA, DE, GR, ES, IE, HU, CH, AT

Retrievability scores
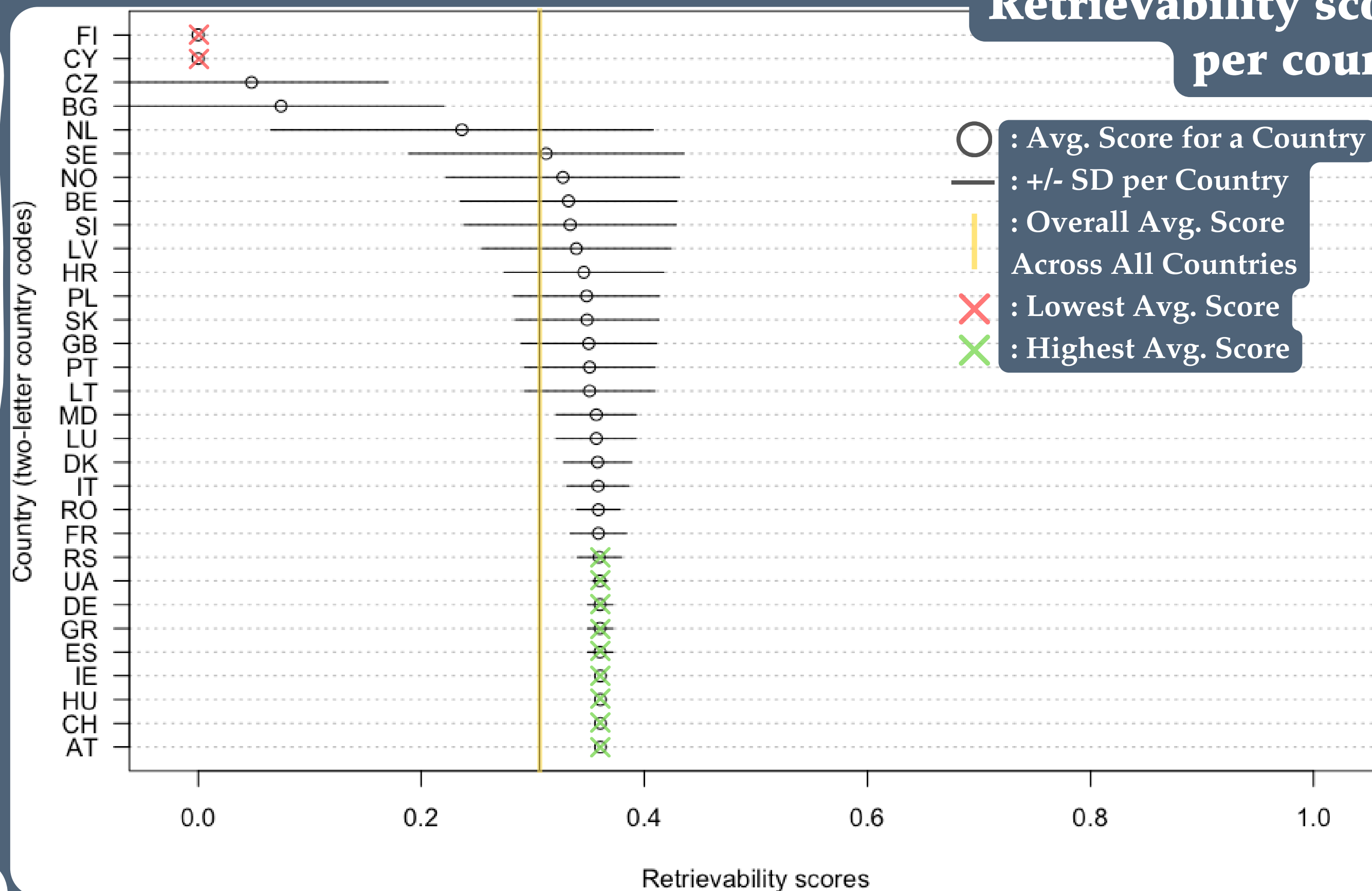
Interoperability

is the extent to which different applications and systems can successfully communicate and exchange data with unambiguous, shared meaning. It includes both syntactic (compatible formats and protocols) and semantic (uniform data codification) aspects.
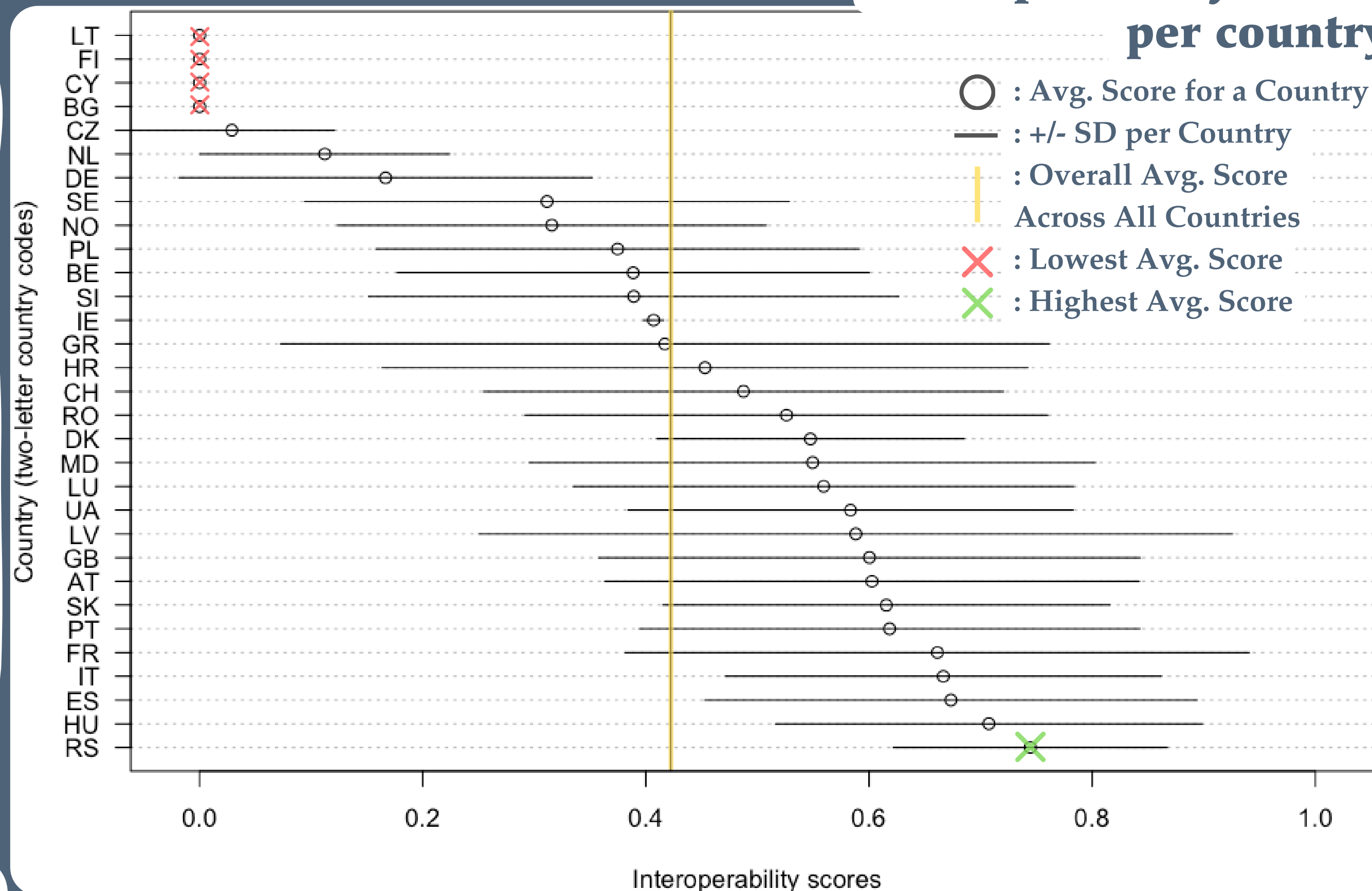

Interoperability scores per country

○ : Avg. Score for a Country
— : +/- SD per Country
| : Overall Avg. Score Across All Countries
✕ : Lowest Avg. Score
✕ : Highest Avg. Score

Reusability

is the extent to which (meta)data are well-described, enabling replication by different teams within different setups.
It includes terms and conditions for access and reuse, provenance information, and contact details for further inquiries.
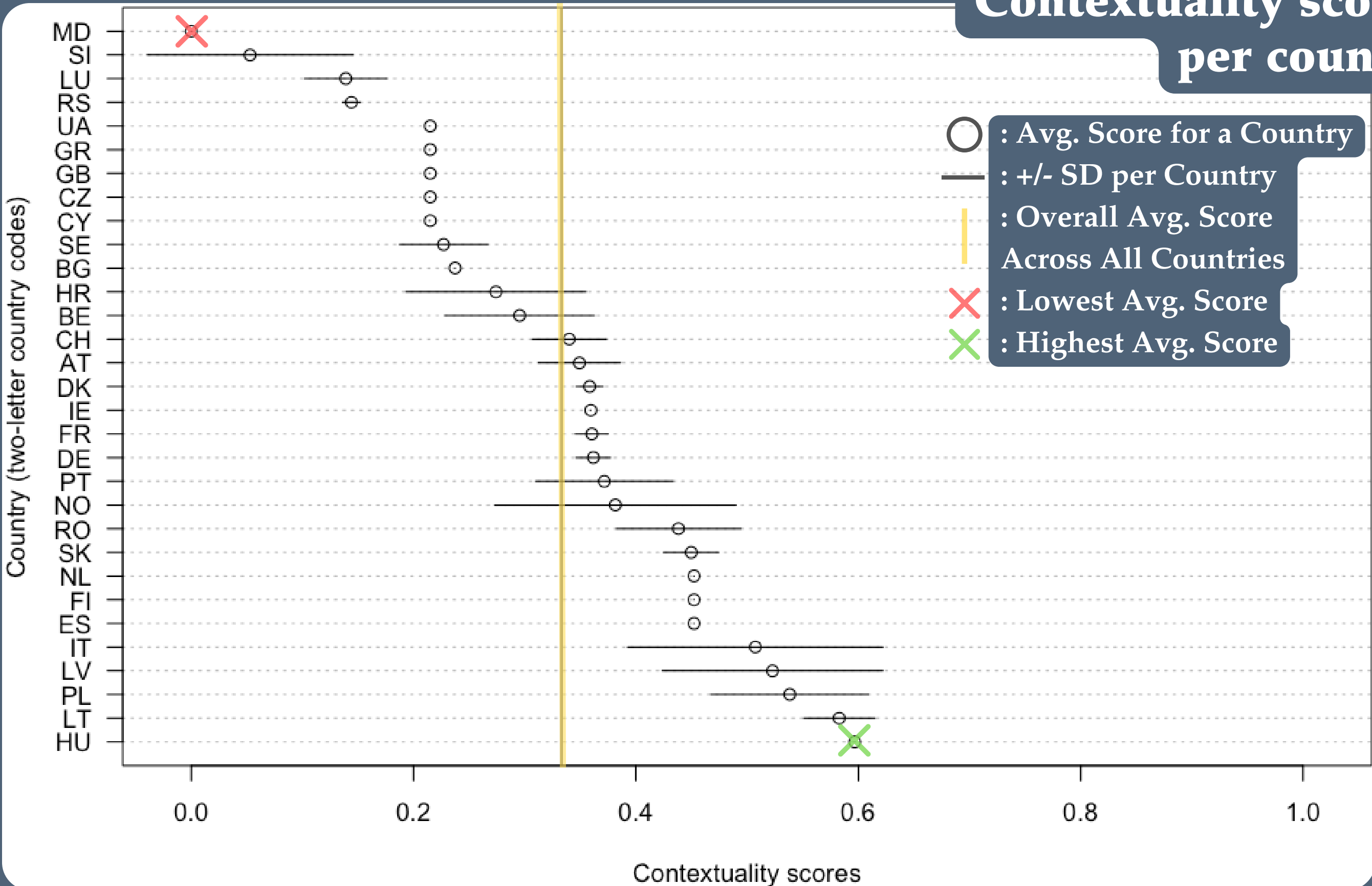
Reusability scores per country

Legend:
- ○ : Avg. Score for a Country
- — : +/- SD per Country
- | : Overall Avg. Score Across All Countries
- ✗ : Lowest Avg. Score
- ✗ : Highest Avg. Score
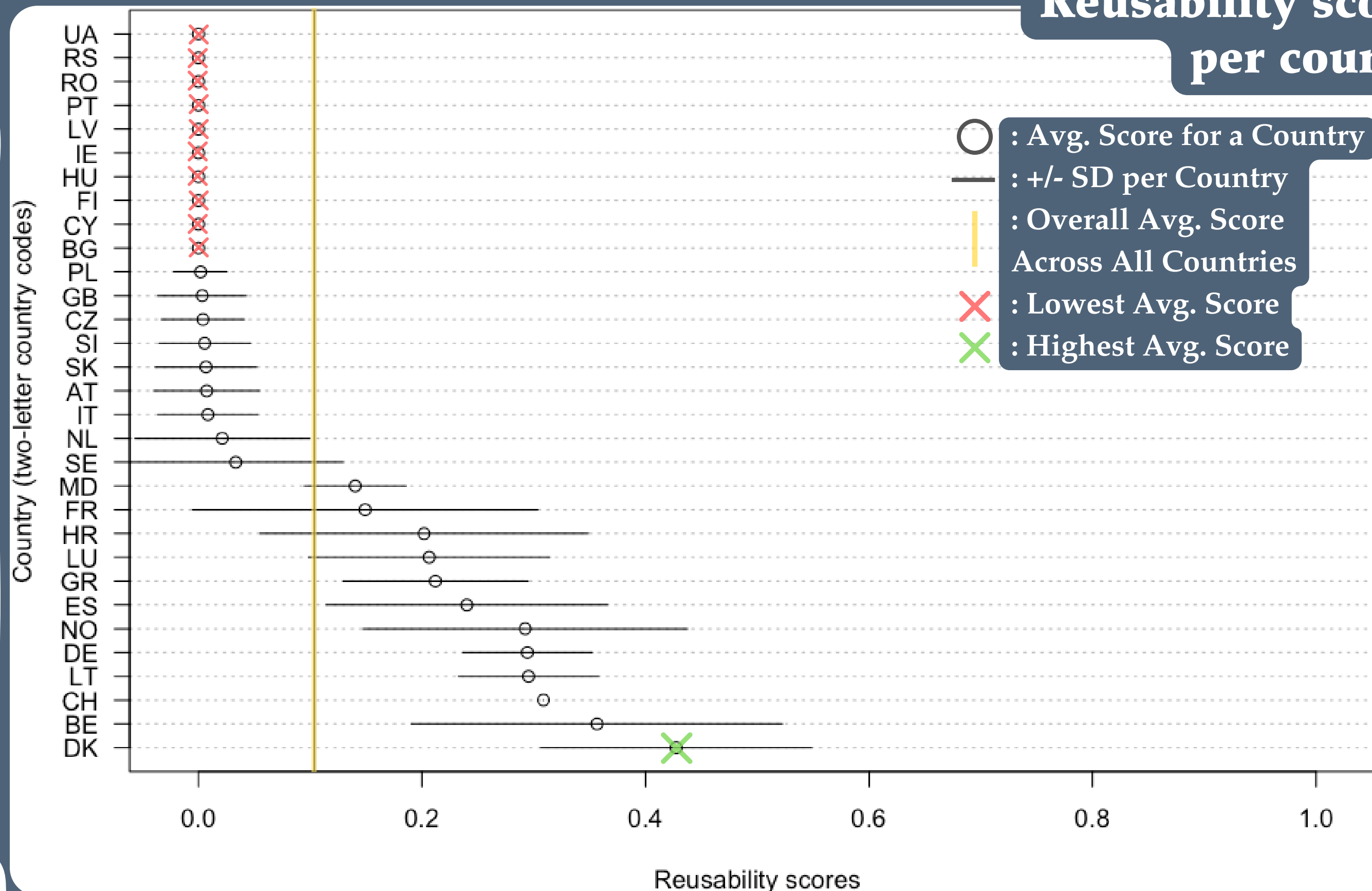
Contextuality

is the extent to which users can obtain additional information about the data, such as origin, quality, copyright, and publication date.

Contextuality scores per country

Legend:
- ○ : Avg. Score for a Country
- — : +/- SD per Country
- | : Overall Avg. Score Across All Countries
- ✕ : Lowest Avg. Score
- ✕ : Highest Avg. Score

Y-axis: Country (two-letter country codes)
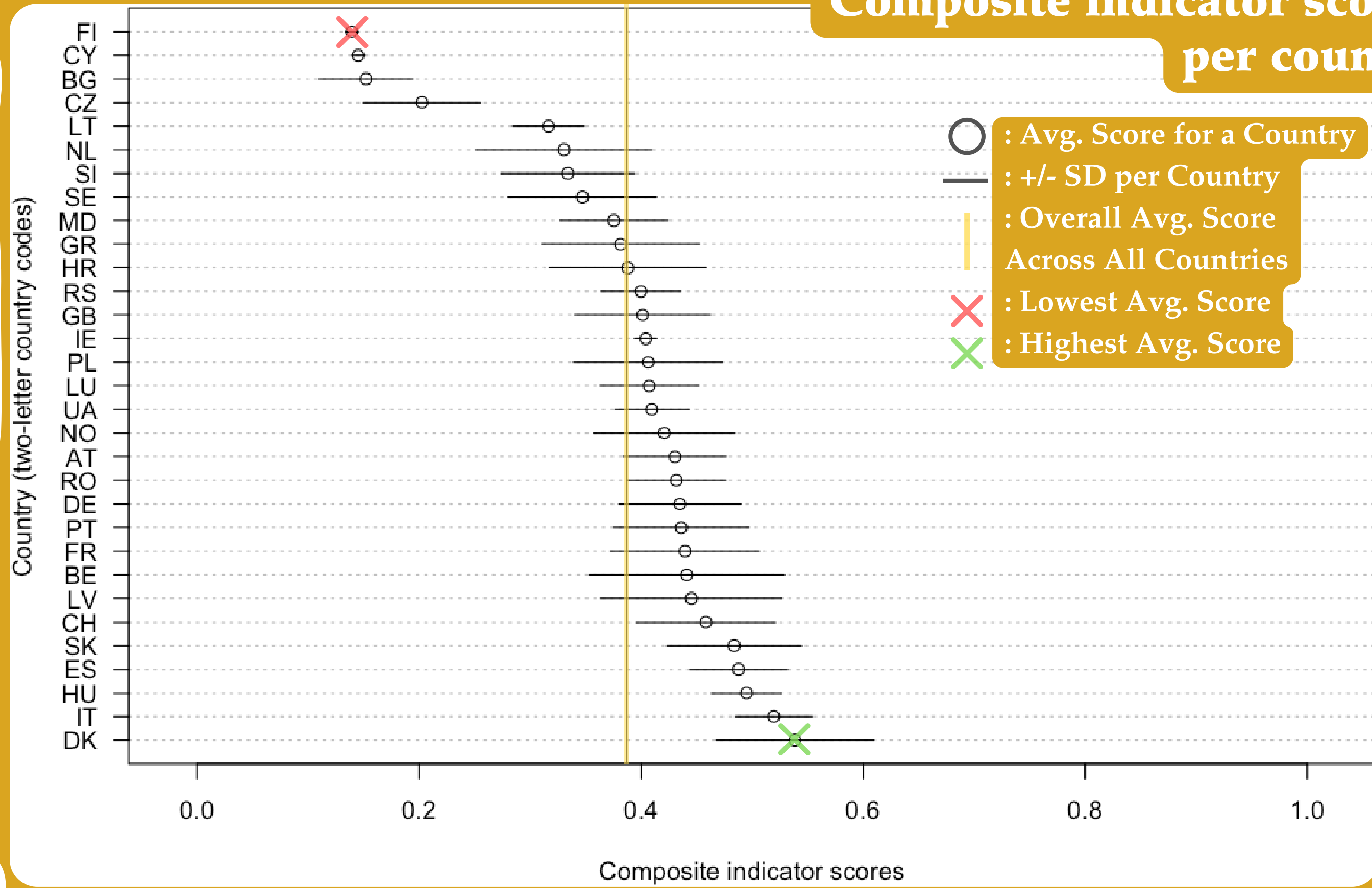X-axis: Contextuality scores

EU OPEN DATA DAYS

Composite indicator

aggregates all previously presented dimensions, providing an overall evaluation of metadata quality for each country.

Composite indicator scores per country

Legend:
- ○ : Avg. Score for a Country
- — : +/- SD per Country
- | : Overall Avg. Score Across All Countries
- ✕ (red) : Lowest Avg. Score
- ✕ (green) : Highest Avg. Score

Y-axis (Country, two-letter country codes), top to bottom: FI, CY, BG, CZ, LT, NL, SI, SE, MD, GR, HR, RS, GB, IE, PL, LU, UA, NO, AT, RO, DE, PT, FR, BE, LV, CH, SK, ES, HU, IT, DK

X-axis: Composite indicator scores (0.0, 0.2, 0.4, 0.6, 0.8, 1.0)

WHAT WE NEED TO REMEMBER

01
UNDERSTAND METADATA QUALITY

Recognize its importance and the diverse composite indicator scores across open datasets.

02
DEVELOP STRATEGIES AND ACTION PLANS

Promote best practices for publishing datasets as open data.

03
PROVIDE PUBLISHING GUIDANCE AND ENSURE QUALITY

Offer clear publishing guidance, implement quality control, and enforce metadata quality assurance on OGD portals while adhering to standards.

04
UTILIZE THE COMPOSITE INDICATOR

Use the developed composite indicator for benchmarking; comparing datasets, portals, countries, or publishers.

THANK YOU
FOR YOUR ATTENTION!

foi