

foi

UNIVERSITY OF ZAGREB
FACULTY OF ORGANIZATION AND INFORMATICS
VARAŽDIN

A Deep Dive into Metadata Quality: Open Datasets Across Europe

Barbara Šlibar

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

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!

METADATA

are data about the data.

BŠ_EU_OPEN_DATA_DAYS_2025.pptx 2 MB
Modified: Today, 14:57

Add Tags...

General:

- Kind: PowerPoint Presentation (.pptx)
- Size: 1.972.571 bytes (2 MB on disk)
- Where: Macintosh HD ▸ Users ▸ bslibar ▸ Desktop
- Created: Friday, 10 January 2025 at 14:20
- Modified: Friday, 10 January 2025 at 14:57

Stationery pad

Locked

More Info:

- Keywords: open data; metadata; quality
- Last opened: Friday, 10 January 2025 at 14:56

Name & Extension:

BŠ_EU_OPEN_DATA_DAYS_2025.pptx

DATA

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

EU OPEN DATA DAYS 2025 v1

THEORETICAL STRUCTURE of the composite indicator

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.

TOP LEVEL

MIDDLE LEVEL

BOTTOM LEVEL

foi

12

Slide 5 of 13 English (United Kingdom) Accessibility: Investigate

Level	Category	Indicator	Count
TOP LEVEL	CI	Findability	5
		Retrievability	3
		Interoperability	5
		Reusability	3
MIDDLE LEVEL	Findability	Completeness	5
		Conformance	1
	Retrievability	Completeness	2
		Conformance	1
		Retrievability	1
	Interoperability	Accuracy	1
		Completeness	2
		Conformance	1
		Openness	2
	Reusability	Completeness	5
		Conformance	1
		Openness	1
	Contextuality	Completeness	7
Conformance		1	
Timeliness		1	
BOTTOM LEVEL			12

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**.

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.

THEORETICAL STRUCTURE OF THE COMPOSITE INDICATOR

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

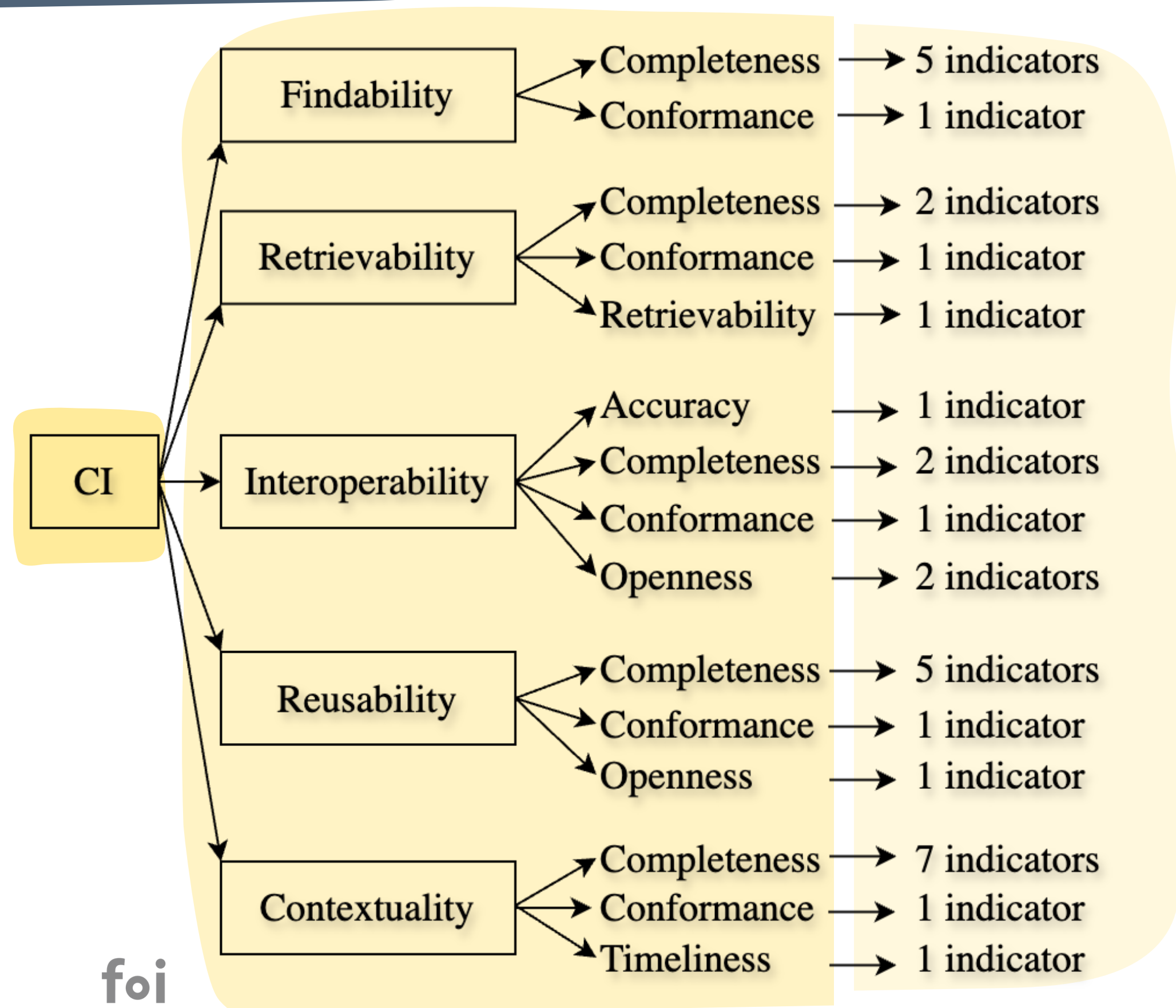
CRITERIA FOR DEVELOPING THE COMPOSITE INDICATOR:

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.



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.

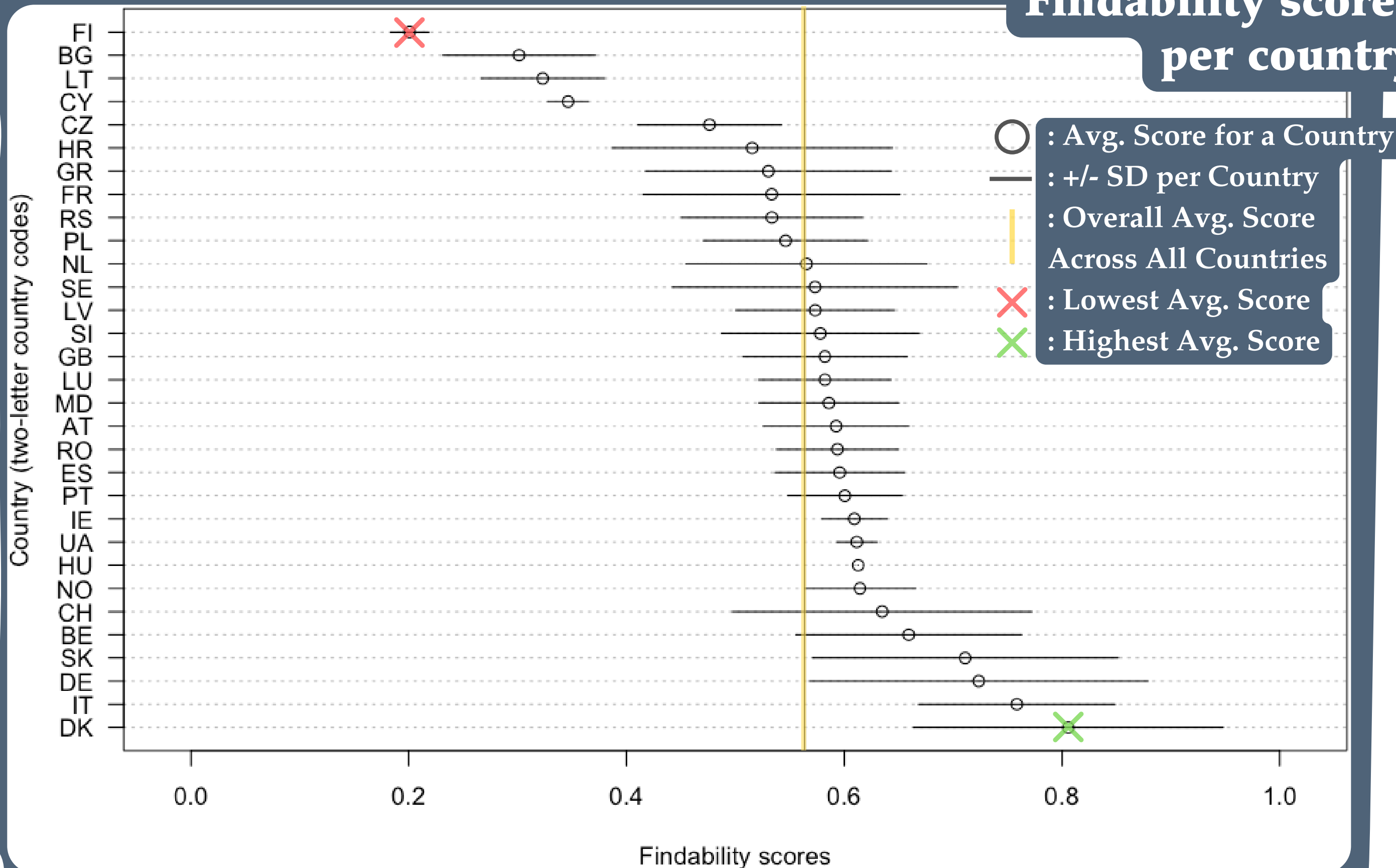
Countries **ordered by the number of datasets** available on October 11, 2024, from highest to lowest.

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

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.

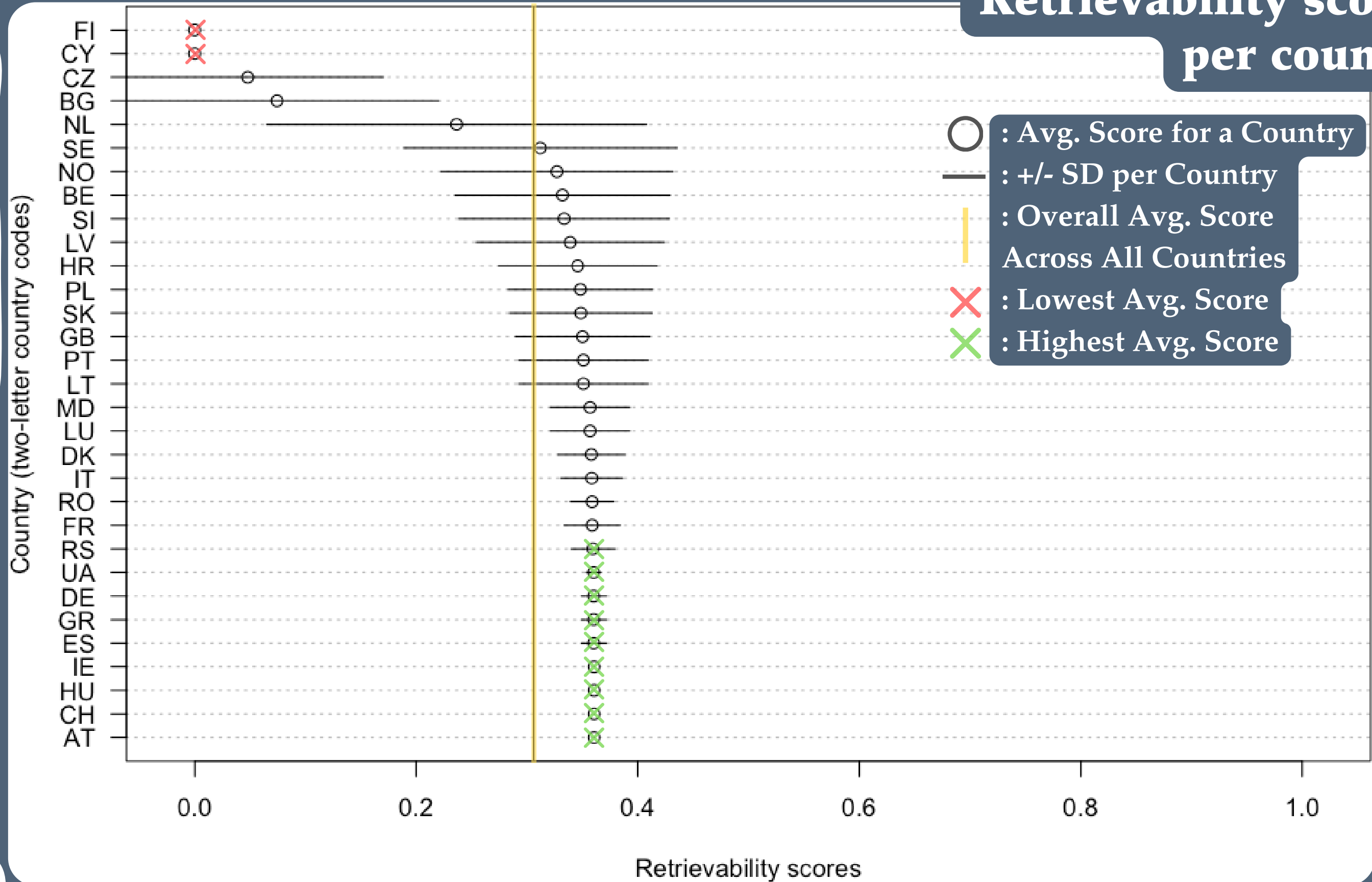
Findability scores per country



Retrievability

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

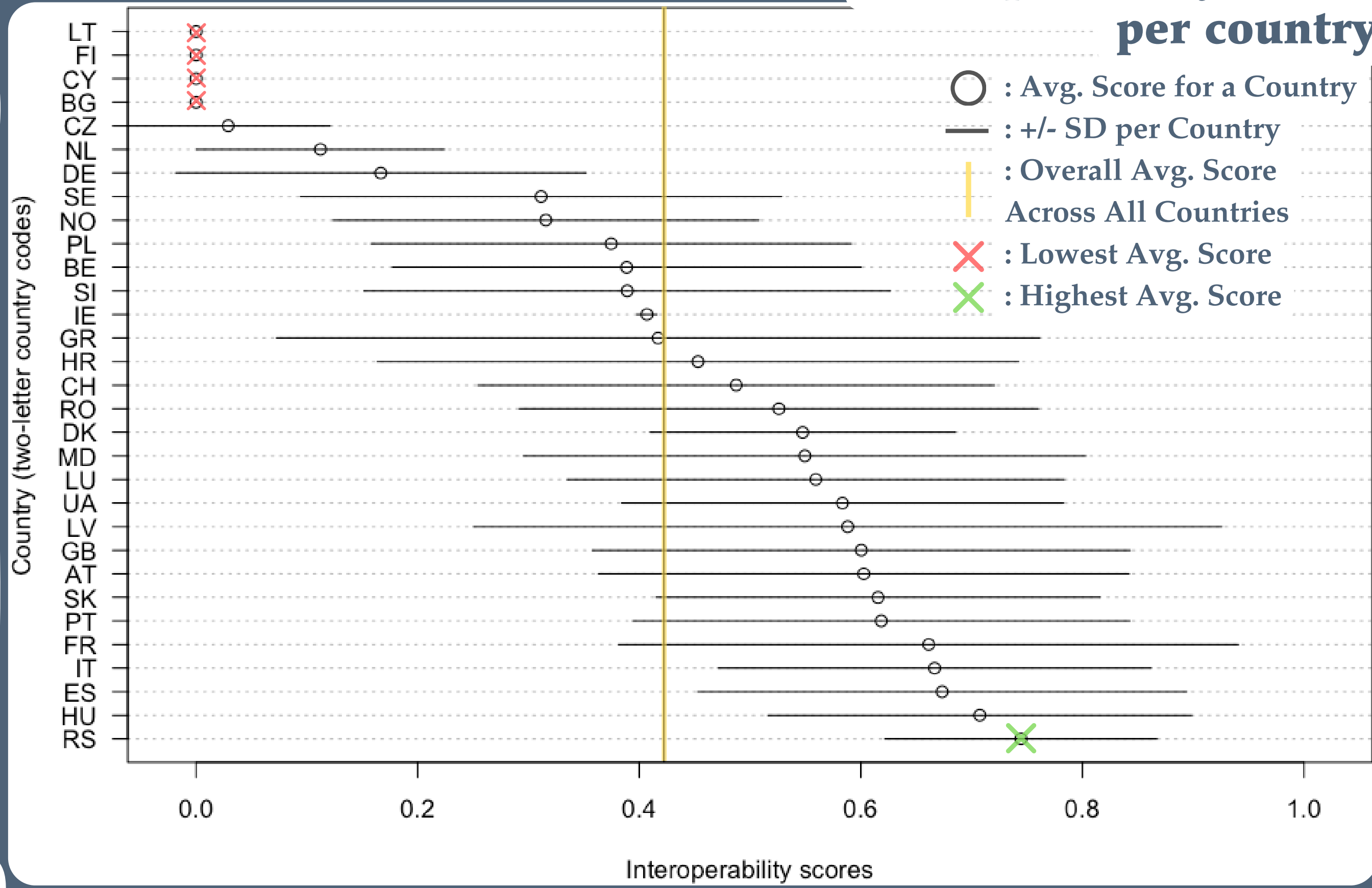
Retrievability scores per country



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.

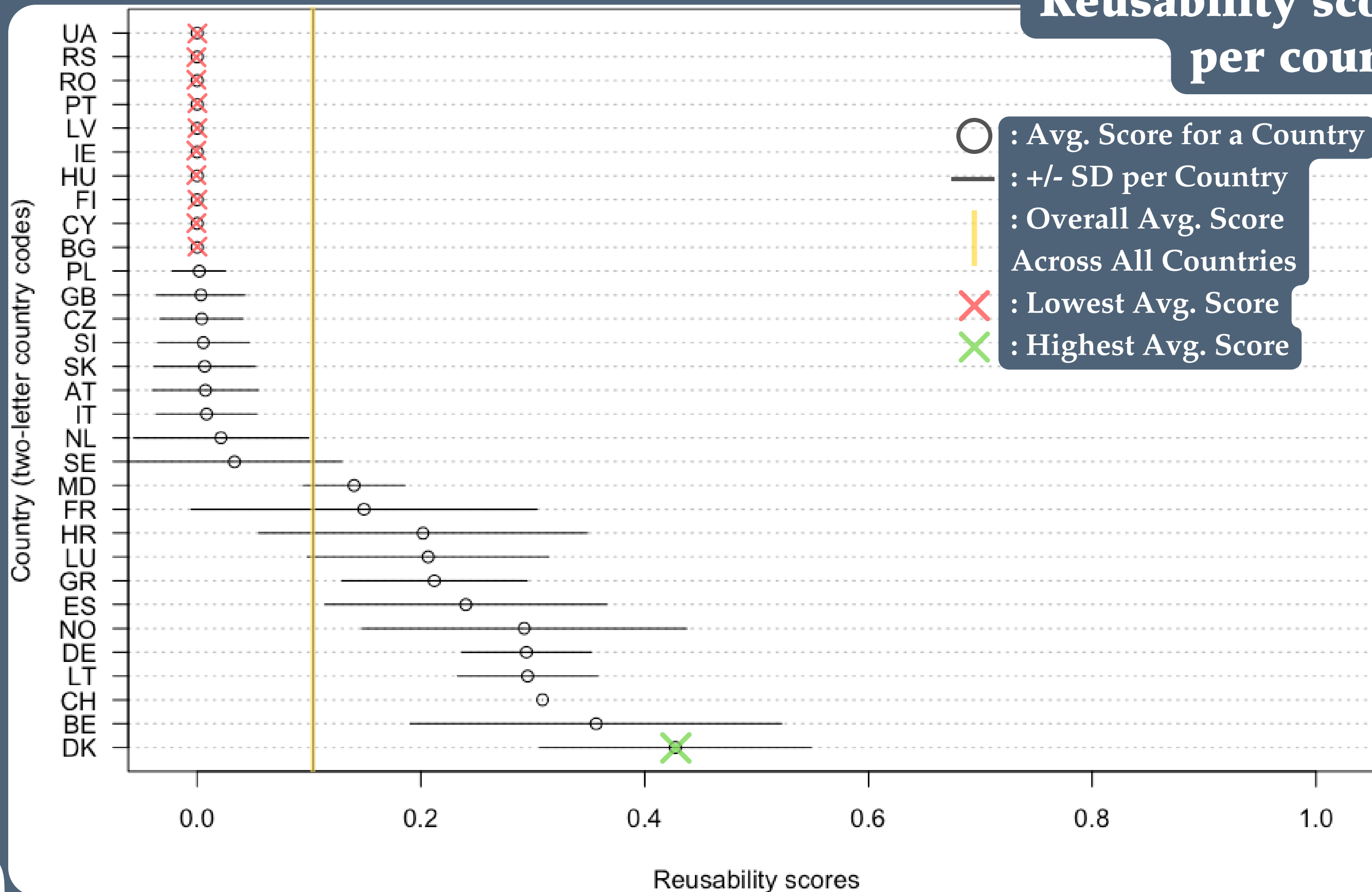
Interoperability scores per country



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.

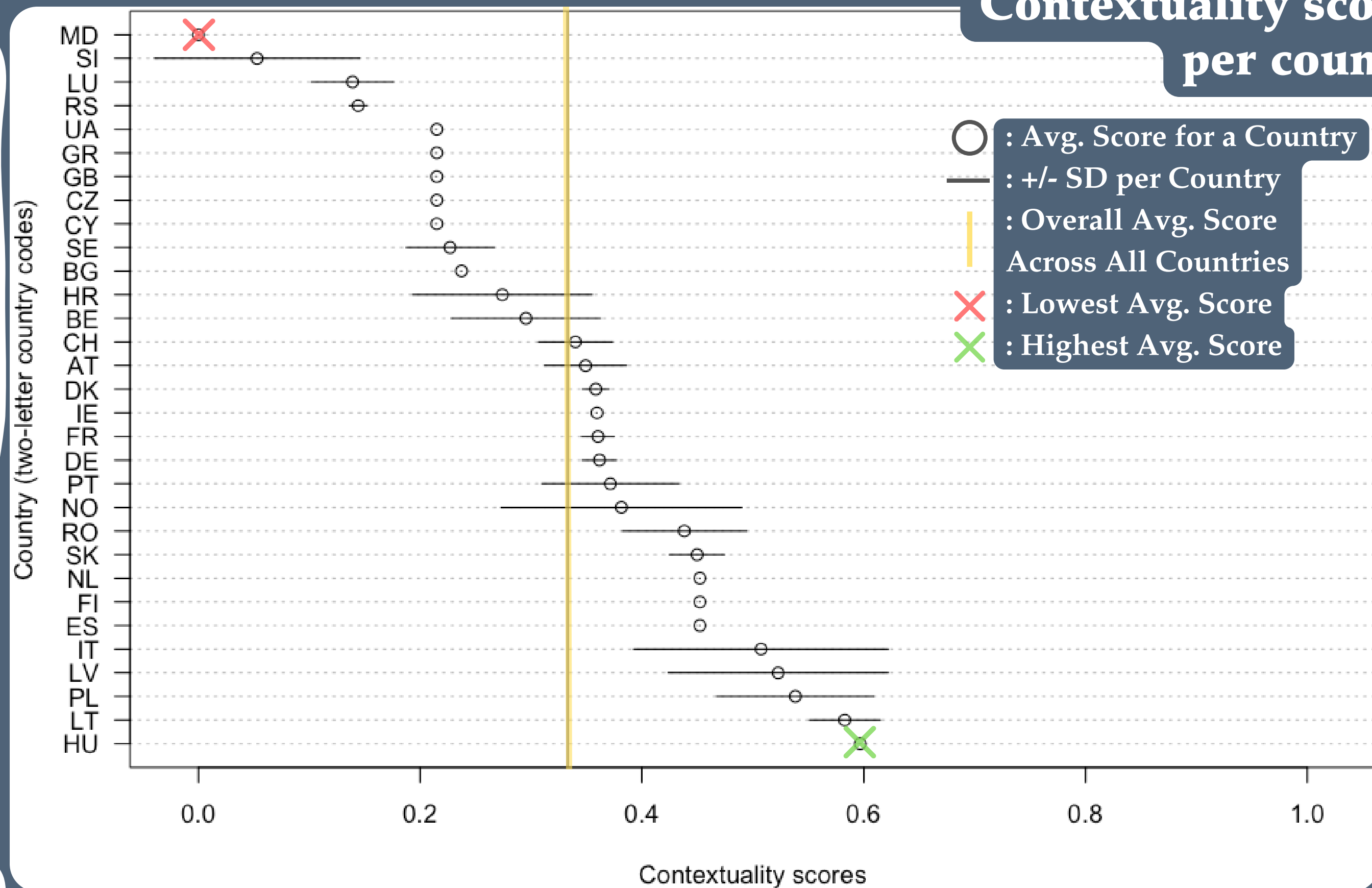
Reusability scores per country



Contextuality

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

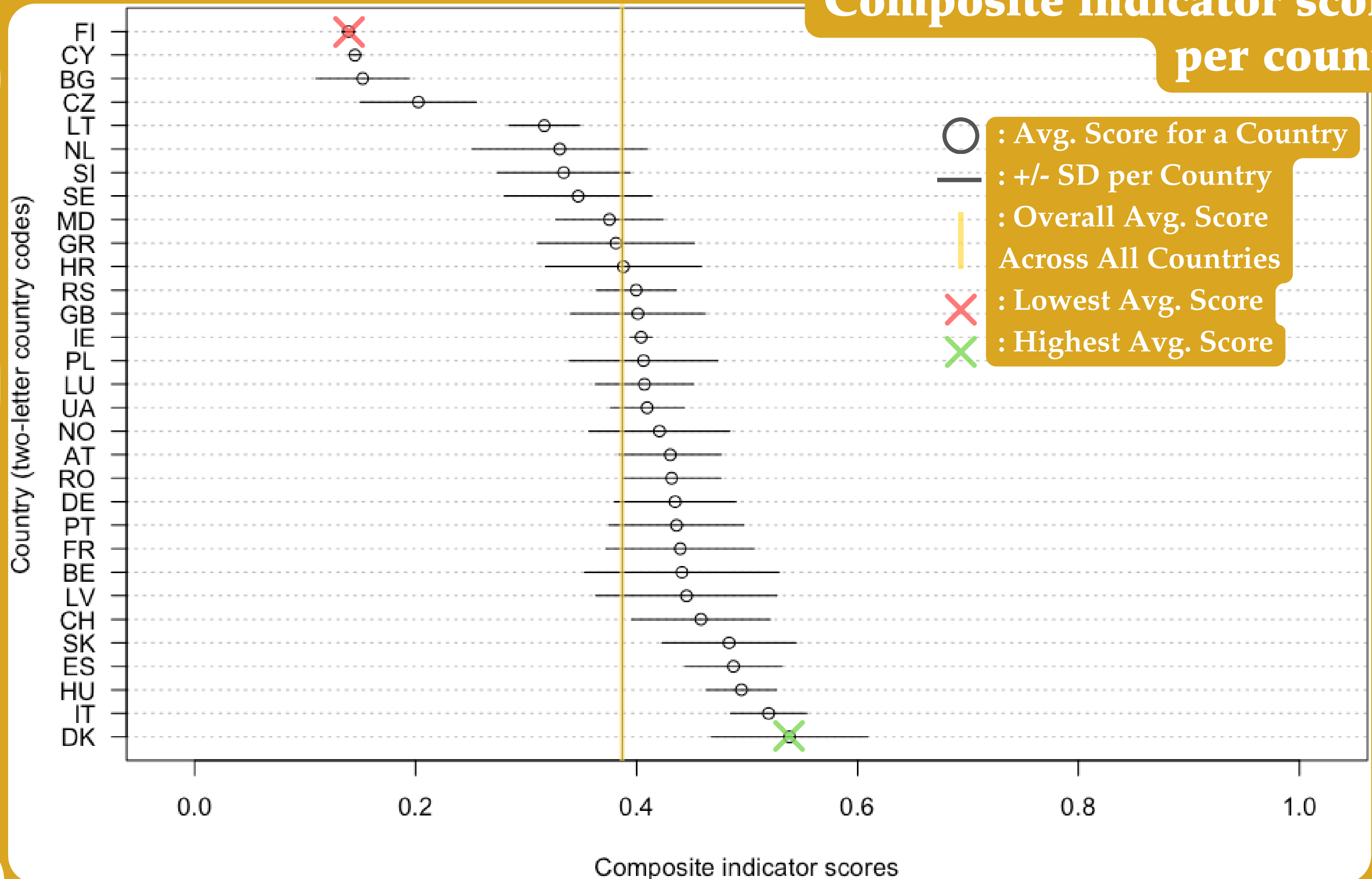
Contextuality scores per country



Composite indicator

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

Composite indicator scores per country



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