Perspectives on the Implementation of the CESSDA Metadata Model

Vipavc Brvar Irena, Bolko Irena

Social Science Data Archives, Univestity of Ljubljana

EDDI 17, Lausanne, 5. - 6. December 2017

About the project

Partners: FSD (lead) + ADP, CASD, DDA, GESIS, NSD, SND and UKDS

The Portfolio aims to support resource discovery, question banks, preservation and data access.

It includes core metadata (study-level and variable / question level)

and controlled vocabularies for relevant metadata fields

-> enabling cross-language information retrieval

DDI compliant

Purpose of Impact Analysis

Portofolio Version 1 took into consideration needs of the Product and Service Catalogue and the Euro Question Bank, as well as daily workflow of the SPs.

CESSDA SP gave feedback and expressed their reservations towards the whole model. Evaluation and more detailed analysis of the use of mandatory fields was made.

Answers and comments also provided pathway for the further work on the development of the Portfolio in Phase 2 of the project.
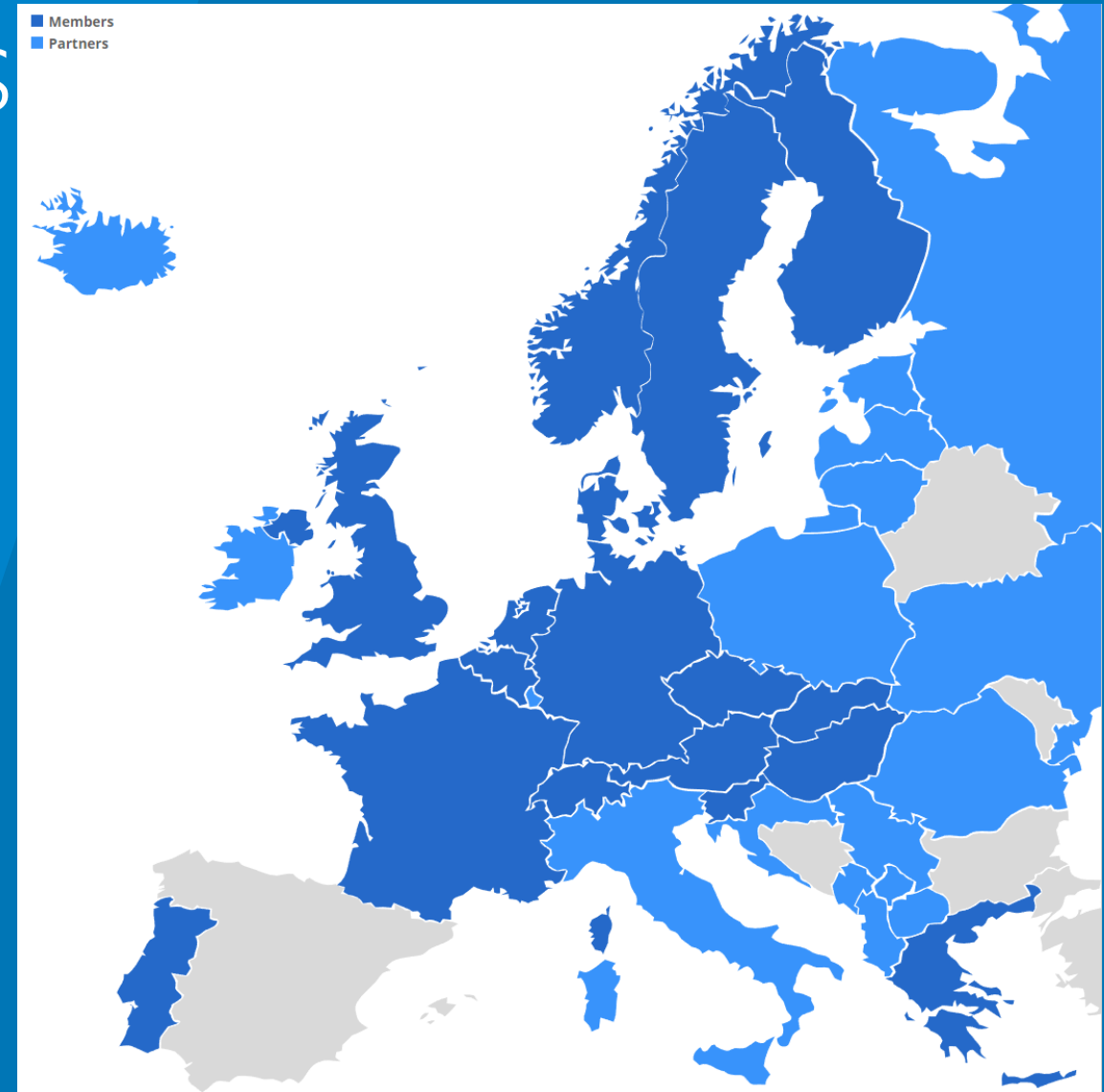
Evaluators – CESSDA SPs

» Austria
» ~~Belgium~~
» Czech Republic
» Denmark
» France
» Finland
» Germany
» Greece
» Hungary
» Netherlands
» Norway
» Slovenia
» Sweden
» Switzerland (Observer)
» UK

13+2

Dataverse

depositor 'CDSP - Centre de Données Socio-Politiques' from Sciences Po

Portugal , Hungary (not member at the time), Slovakia (don't have archive)

Members
Partners

cessda eric

Methodology – questionnaire

Phase 1:

- Evaluation of Mandatory elements - 15 - (crucial for understanding and finding studies)
- Adjusted System Usability Scale – 11 items
- To what extend SPs provide metadata

  study description, variable description and question text + in English?

cessda eric

Methodology – administration

February 8th 2017 – Portfolio draft posted on Basecamp
-> collecting comments

February 24th - Personalized email invitations send – Google Sheet
……. After
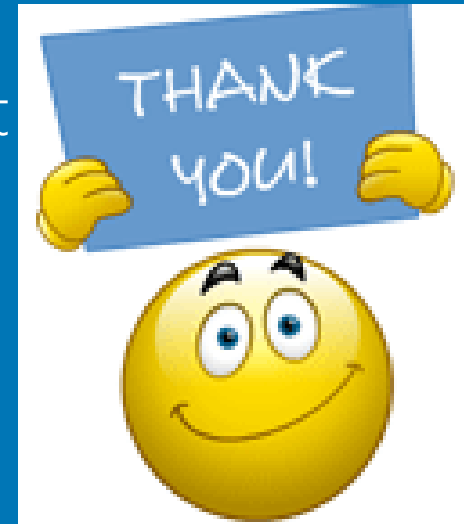
All possible answers received by April 10th

cessda eric

Providing Metadata

	Available in national language (n=15)	Available in English (n=14)
Study description metadata	15	11
Variable description metadata	13	7
Question text metadata	12 + 1 planned	5

Source: Impact analysis survey, 2017

Variable label metadata (in EN)

About 6% of the data holdings are available with variable description.

Only small part (less than 10 %) of content is available in English.

Some variables (but not all) are described in English.

Variable descriptions in English available for 25 % of our data.

In national language or in English depending on the study .

We provide the element in English if the data is natively in English.

Variable descriptions are available in codebooks, which are stored alongside the data; the language is not standardized, it depends on the survey.

We do not translate variable level, unless principal investigator provide us with documentation and data files in original and English language.

Only for selected NESSTAR published studies.

MANDATORY ELEMENTS

Mandatory Element	CV for this element
Study Number	NO
Title Study	Use ISO 3166 2-letter code as country pairs: language - country
Principle Investigator Reference	NO
Publisher	NO
Publication Date	ISO 8601
Kind of Data	DDI CV: KindOfDataFormat
Class Identifier	NO
Classification System Name	NO
Study Area Country	ISO 3166 2-letter code
Universe	NO
Type of Sampling Procedure	DDI CV: SamplingProcedure
Type of Mode Collection	DDI CV: ModeOfCollection
Data access	CV "AccessClass" – (open/restricted)
Title of study documentation	NO
Producer of the study documentation	NO

cessda eric

Source: Impact analysis survey, 2017   n=14
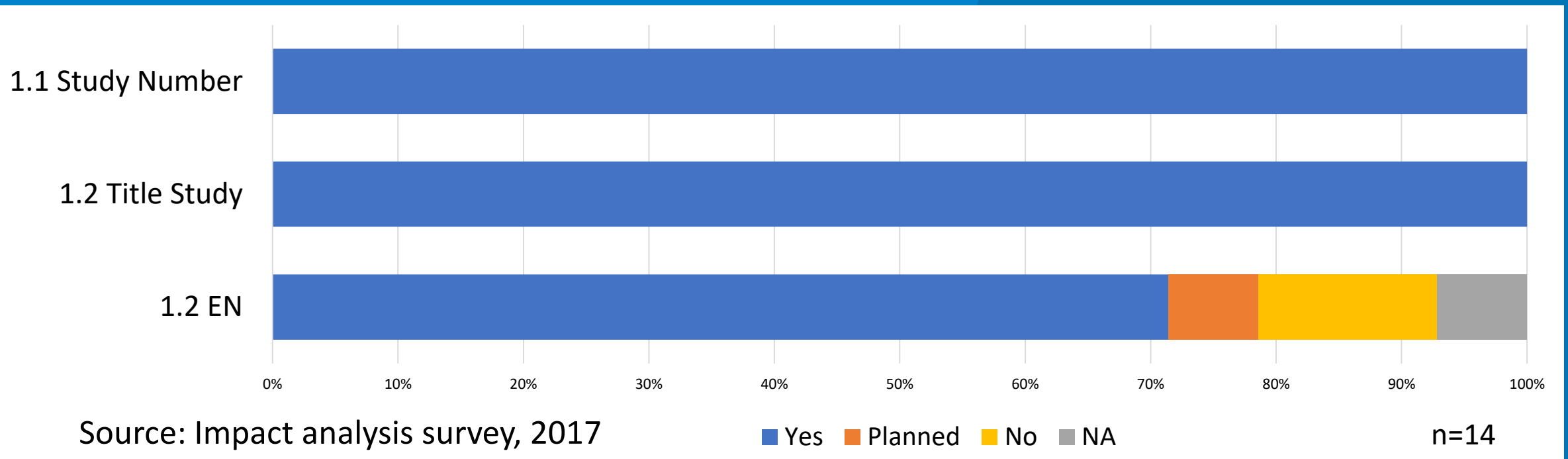
Study number:

DOI or URN used instead of classical Study Number.

Will we have different study no. for English / other language description?

Title Study:

Is the field repeatable? International or multilingual studies might have more titles.

cessda eric

Principal Investigator Reference:

Which registry system is going to be used / proposed? Should we be able to use multiple (national, international)?

What do we do with PI that don't have "their number" (students, retired professors)?
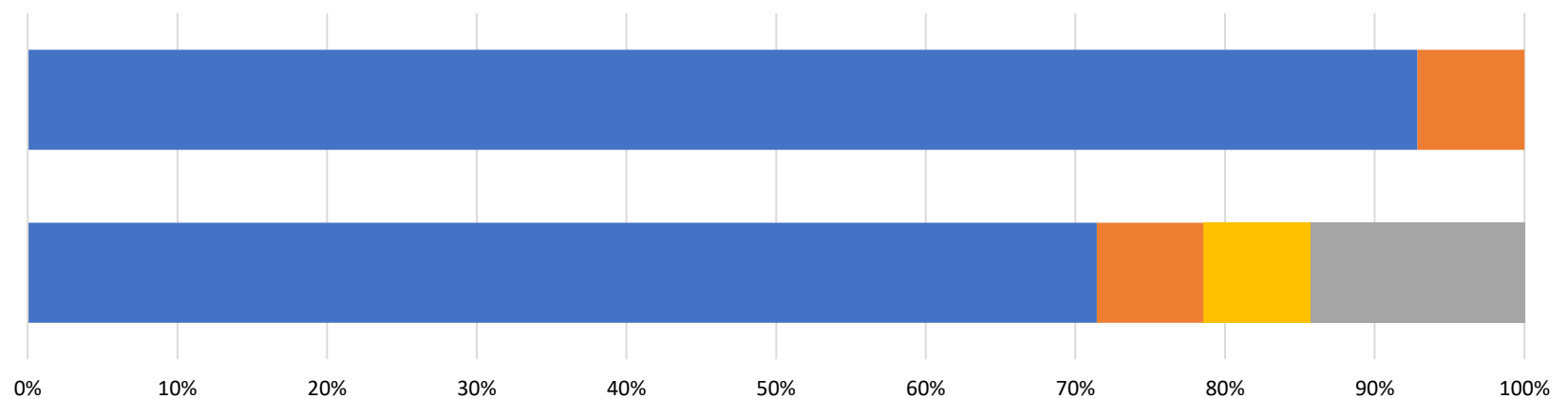
There is a need to use this fields for persons and organizations.

DataCite use exact, and a bit different description of this field and its attributes. Do we need it for CESSDA?

Chart
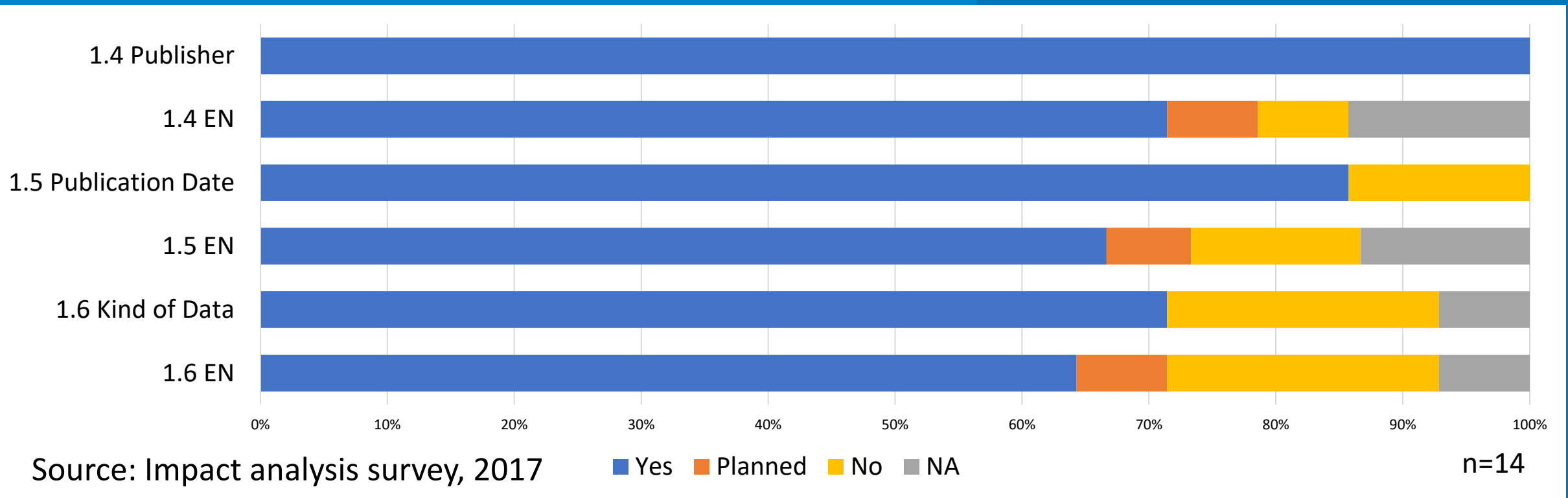
1.3.1 Principle Investigator Reference — Yes (~93%), Planned (~7%)

1.3.1 EN — Yes (~71%), Planned (~7%), No (~7%), NA (~14%)

Legend: ■ Yes ■ Planned ■ No ■ NA

Source: Impact analysis survey, 2017

n=14

Source: Impact analysis survey, 2017

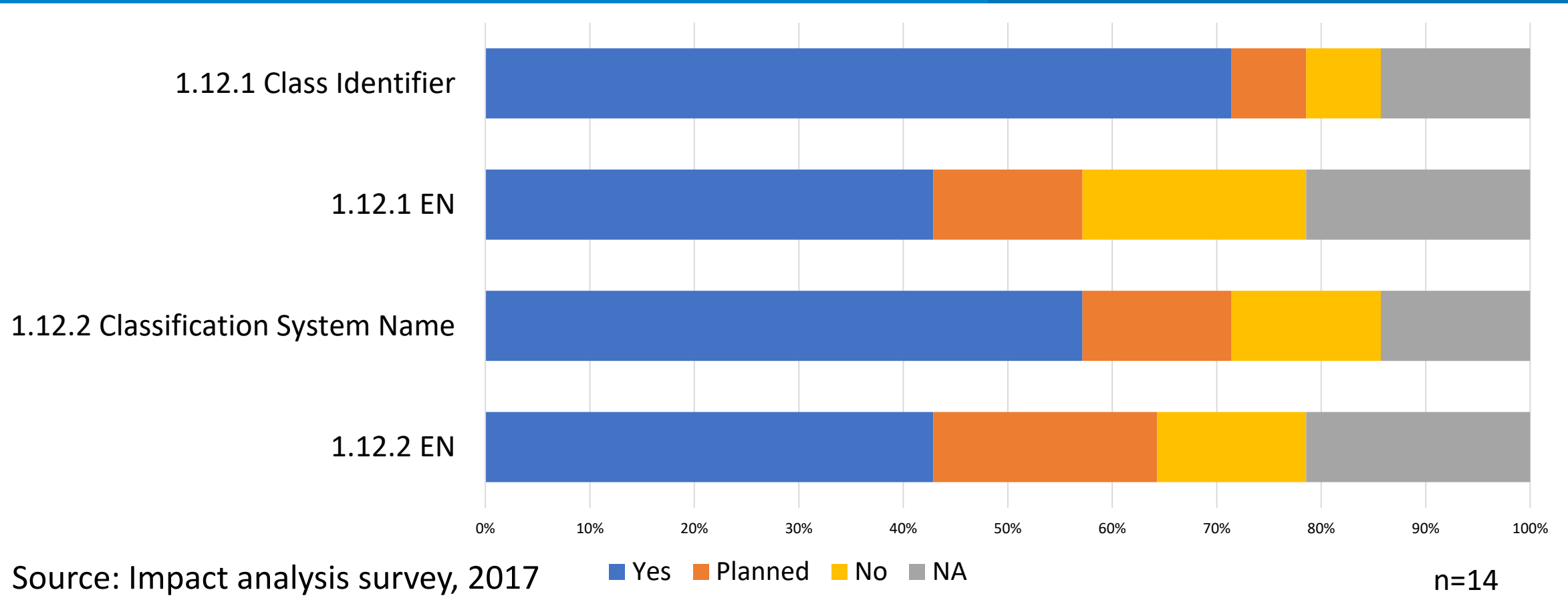Publiser: Do we need English translation of the names of organizations?
Publication or Creation date for citation?
Clear definition of each field is needed.
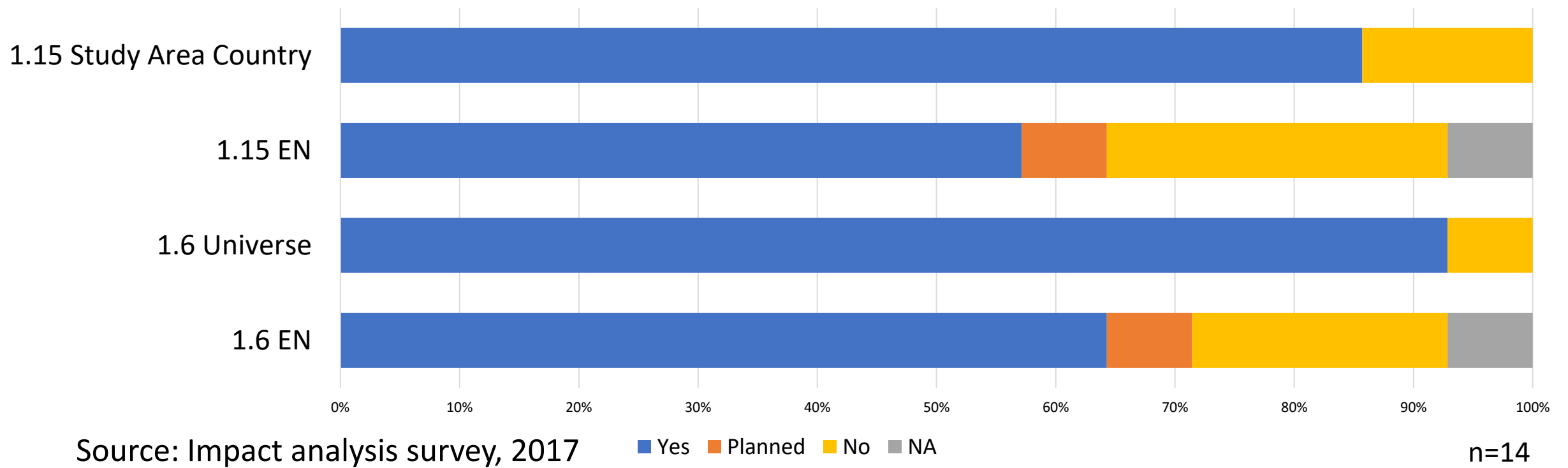Date: Need for YYYY-MM-DD?
Kind of Data: Confusion with the use of this and a-like fields / Clear definition of this element is needed. CV?
Will the element be defined on the level of study or dataset?

cessda eric

Chart: Impact analysis survey 2017

Categories (bars):
- 1.12.1 Class Identifier
- 1.12.1 EN
- 1.12.2 Classification System Name
- 1.12.2 EN

Legend: Yes, Planned, No, NA

Source: Impact analysis survey, 2017   n=14

Classification: Several countries do not provide these metadata. If SP use the element, it is likely that CV is used and content is translated in English.

Will CESSDA Top Class be mandatory? Where is it accessible?

cessda eric

Source: Impact analysis survey, 2017 · n=14

Yes · Planned · No · NA

Chart categories: 1.15 Study Area Country, 1.15 EN, 1.6 Universe, 1.6 EN

Country: Element used; plan to use ISO code

Most studies are done inside one country. Lower level elements (region, town) are needed for classification of the study and statistical analysis of collected data.
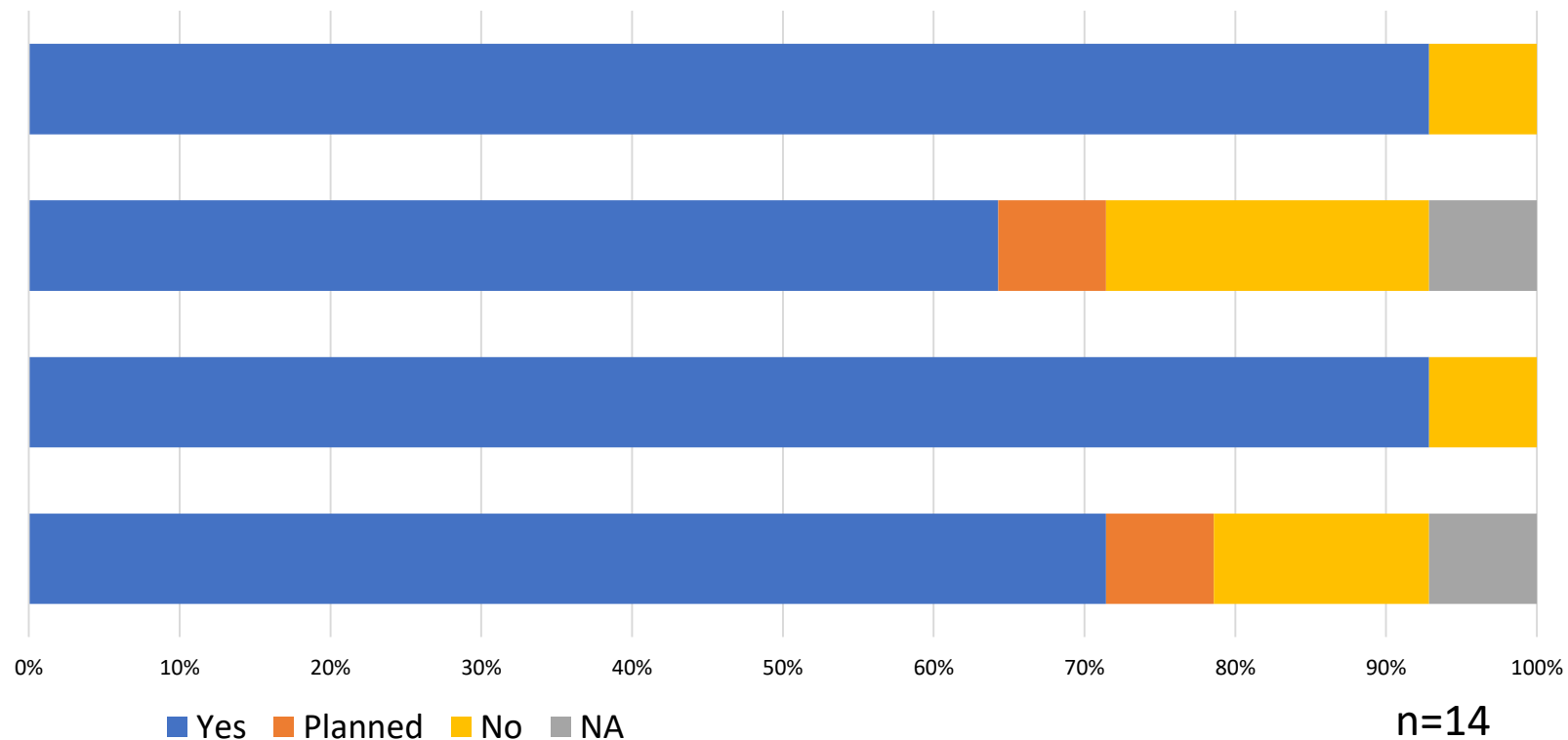
Universe: Free text; some use it on the level of dataset and not study.

cessda eric

Chart: Impact analysis survey results

| Category	Values
1.18.2.1 Type of Sampling Procedure	Yes ~93%, No ~7%
1.18.2.1 EN	Yes ~64%, Planned ~7%, No ~22%, NA ~7%
1.18.4 Type of Mode Collection	Yes ~93%, No ~7%
1.18.4 EN	Yes ~71%, Planned ~7%, No ~14%, NA ~7%

Legend: Yes, Planned, No, NA

Source: Impact analysis survey, 2017   n=14

Sampling: Use DDI CV or their own.

Will the element be defined on the level of study or dataset?

Qualitative and administrative studies?

Mode: Study / Dataset

cessda eric
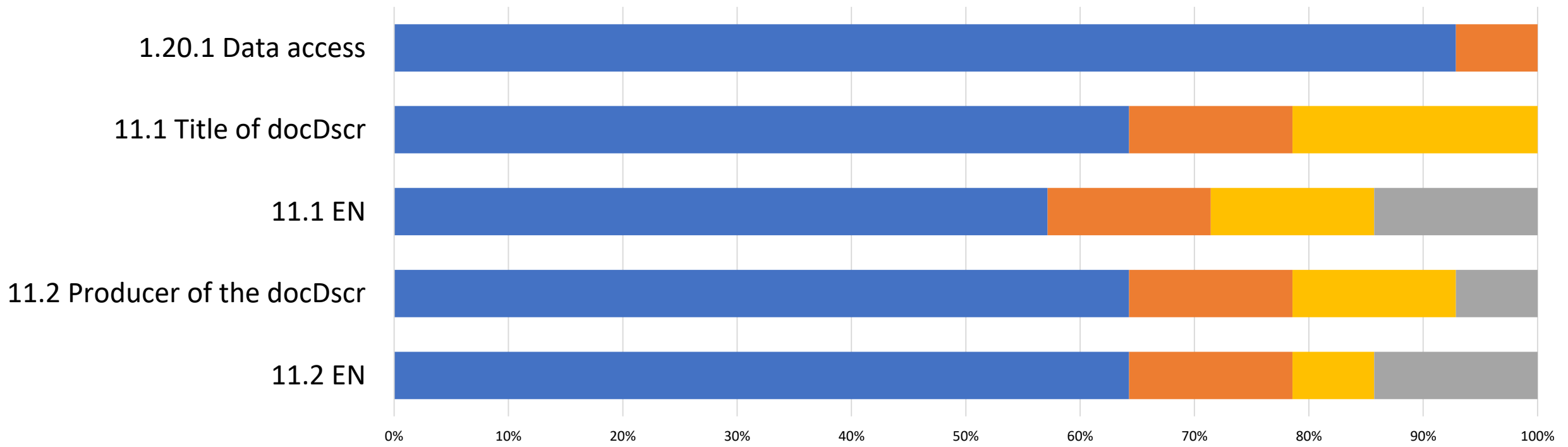
Source: Impact analysis survey, 2017 n=14

Data Access: CESSDA policy. Need for clear definition. Relation to the „restriction„

docDscr:

Clear description needed. What would we need to translate English?

Developed vs. less developed

Development level on the basis of CESSDA Maturity model - SAW project

We have deloped and middle developed countries.

Not much difference on the level of mandatory fields.

Middle developen – newer – had access to tools and good practices from older / larger archives.

… translation in English

cessda eric

Dataverse

„ an open source software application to share, cite and archive data„

DANS leading this in CESSDA SP (since lately also AUSSDA)

Would easily feed the needs.
    Currently CV not supported. Could be added.
    Multilinguality crucial (currently 2 languages)

CESSDA project in 2018 – DataverseEU

Sytem usability scale

Used to do general evaluation of the Portfolio.

Expressed generally positive attitude towards usage of the Portfolio.

middle developed SPs will require more help of CESSDA MO and reported they will need some more time to learn how to use the Portfolio

DDI-L and DDI-C version of Portfolio needed.

Tools needed.

cessda eric

Sytem usability scale

Source: Impact analysis survey, 2017

cessda eric
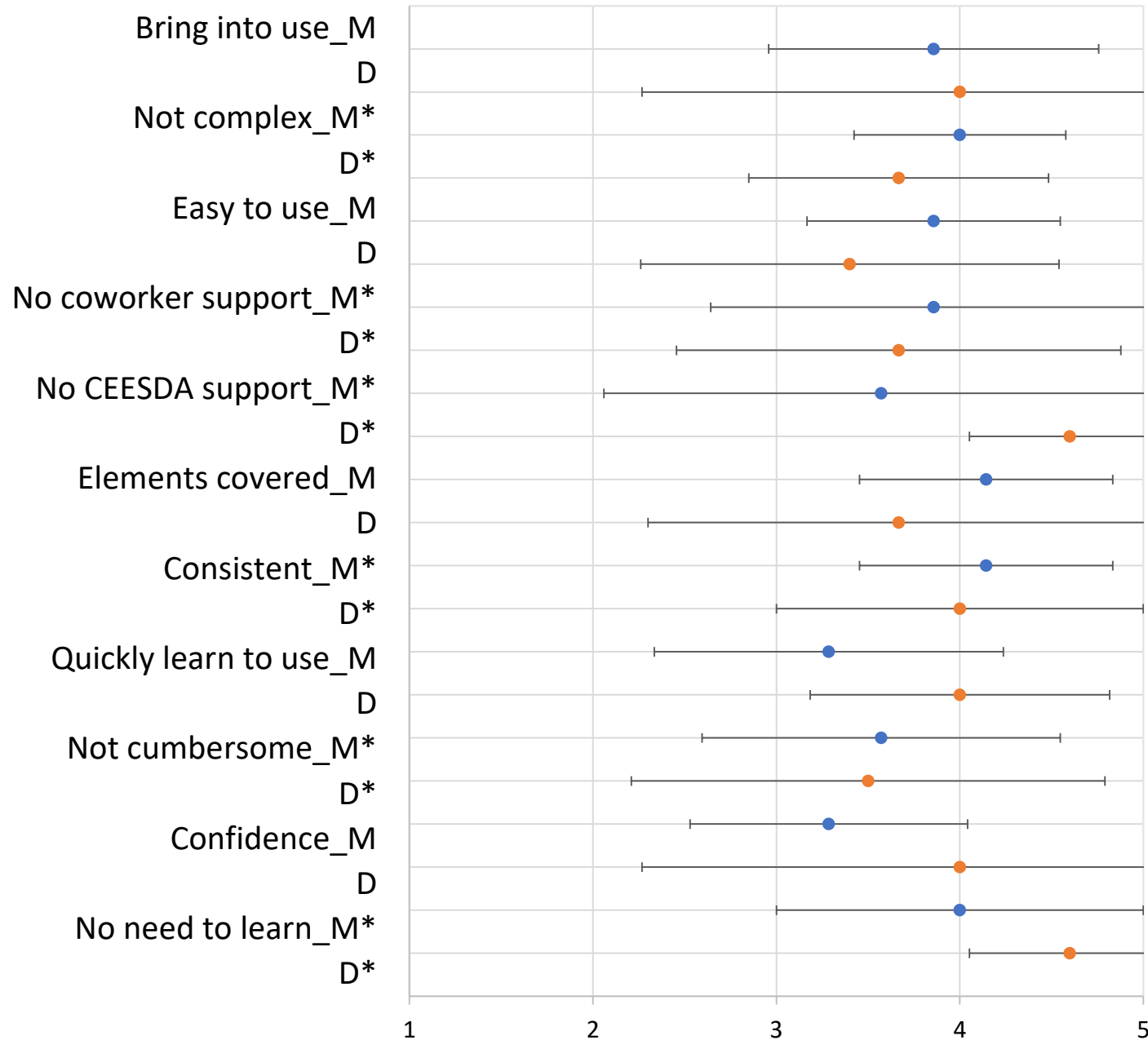
Summary

The main aim of the survey was to analyse the **impact of the proposed Portfolio solution on the** <u>workflow</u> **and** <u>processes</u> of CESSDA SP's and furthermore its impact on CESSDA's policies, products and services.

-> not all SP provide metadata for mandatory metadata, but could in the future
-> CVs are not commonly used, but can be if needed

Clear guidance and tools needed.

cessda eric

Phase 2

SP doing real testing – preparing XML description of typical Study using the Portfolio elements -> discuss possible open issues in 2019

What is needed to be able to offer context for Mandatory Fields and using CV?

Are user guidelines clear enough?

Testing CESSDA Partners , depositors, Institutional repositories – matching their description to Portfolio Mandatory Fields

cessda eric