

Towards the Discovery of Person-Level Data

International Workshop on Semantic Statistics

22 October 2013, Sydney, Australia

Thomas Bosch¹, Benjamin Zopilko¹, Joachim Wackerow¹, Arofan Gregory²

¹GESIS – Leibniz Institute for the Social Sciences, Germany

{first name.last name}@gesis.org

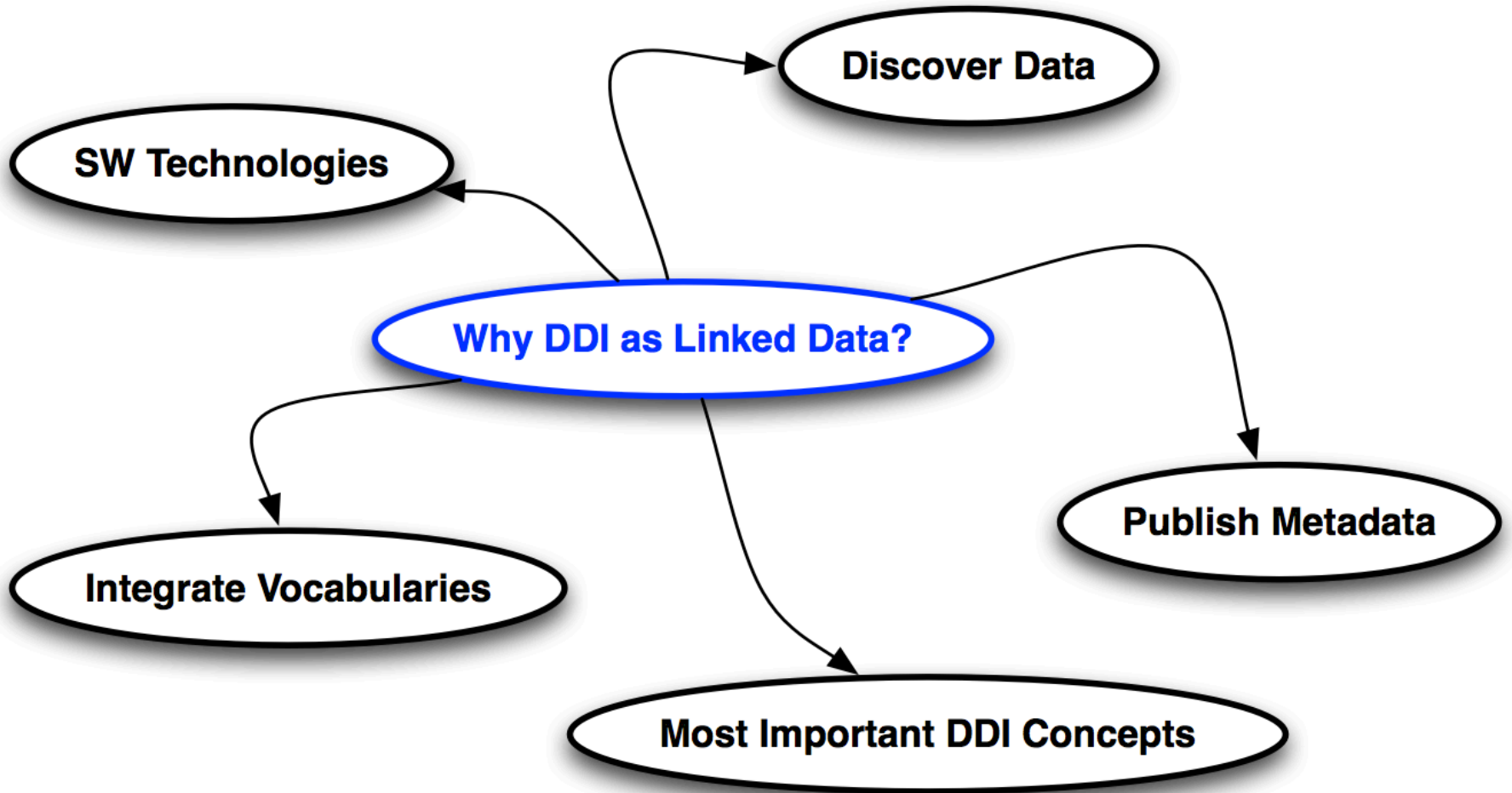
²Open Data Foundation, USA

agregory@opendatafoundation.org

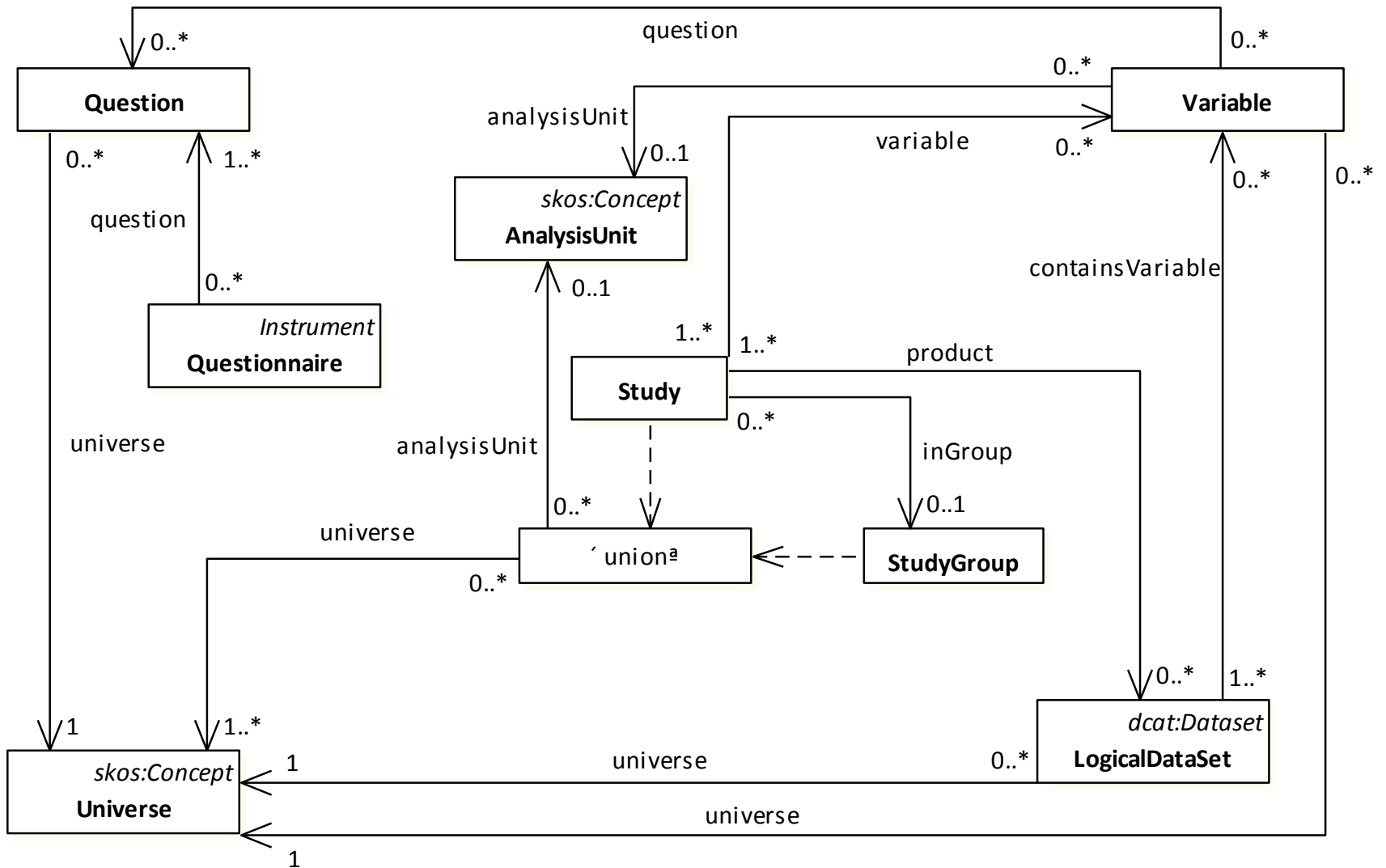
Member of the

Leibniz
Leibniz Association

Why DDI as Linked Data?

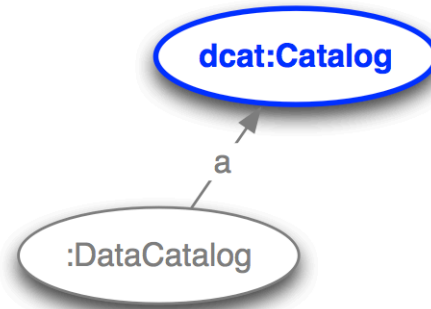


Overview

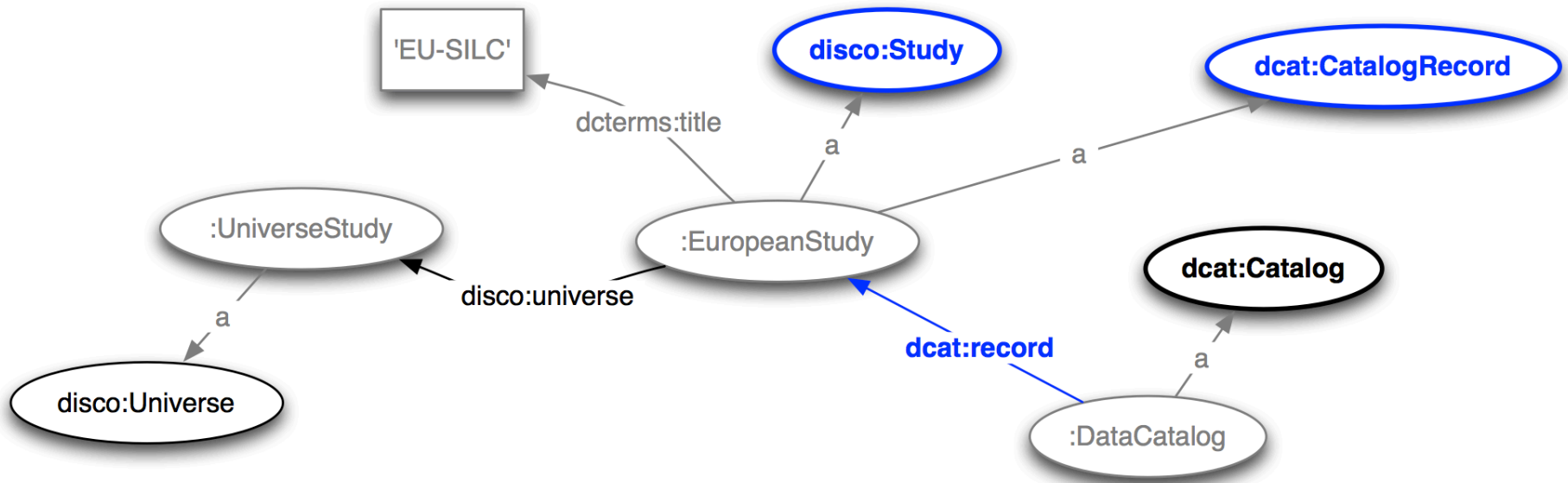


Use Cases

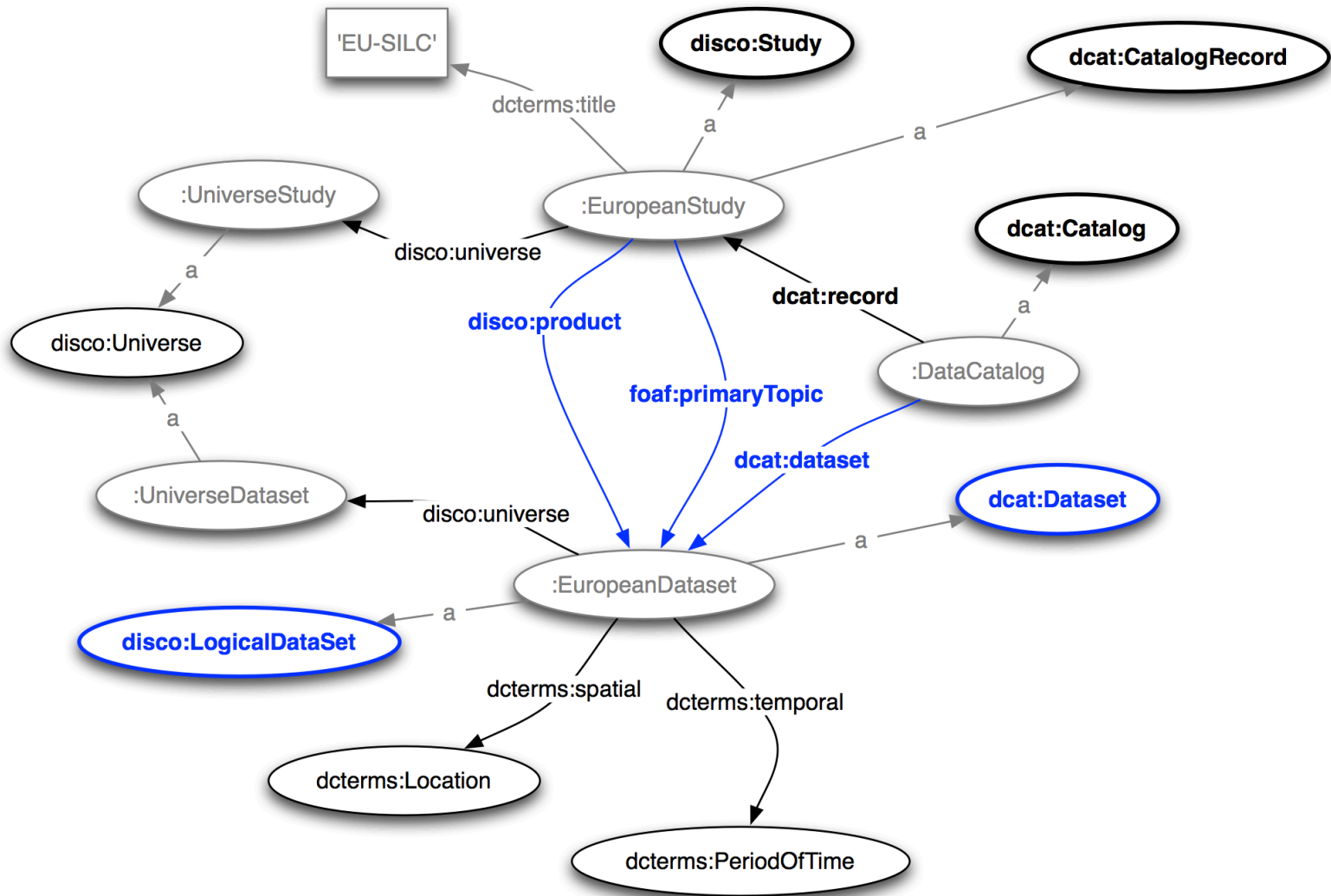
Where to search for specific **data**?



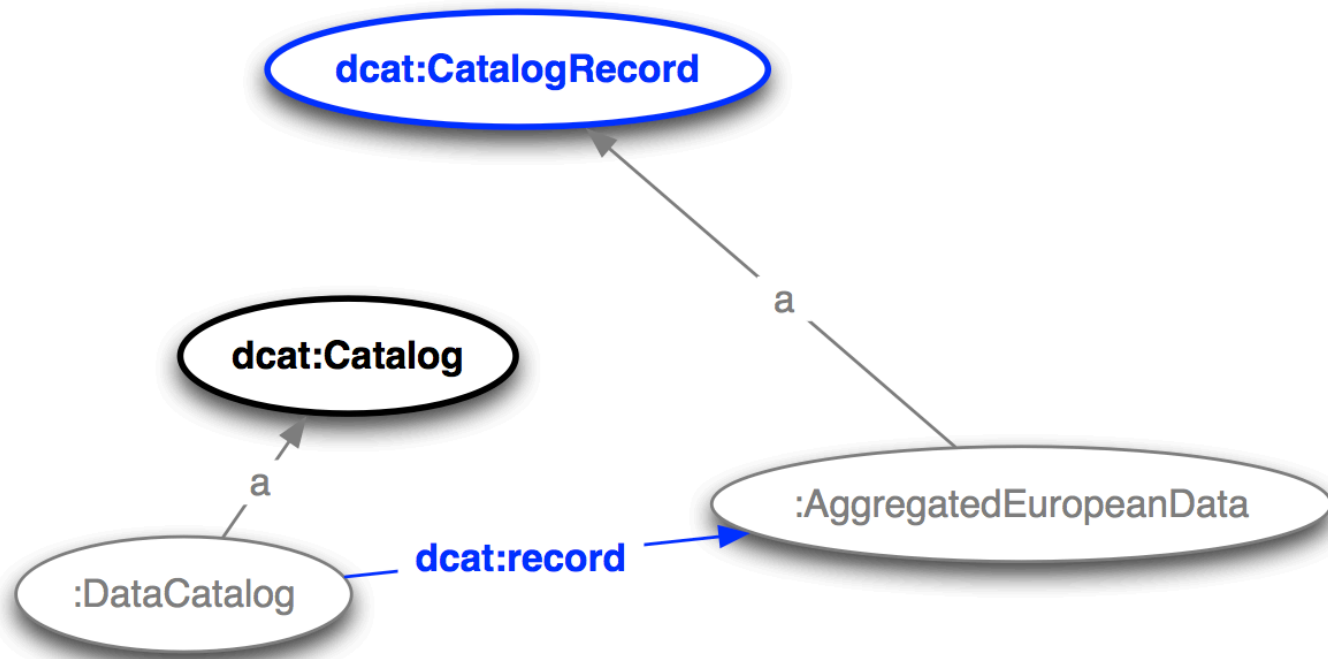
What **microdata** according to specific metadata exists?



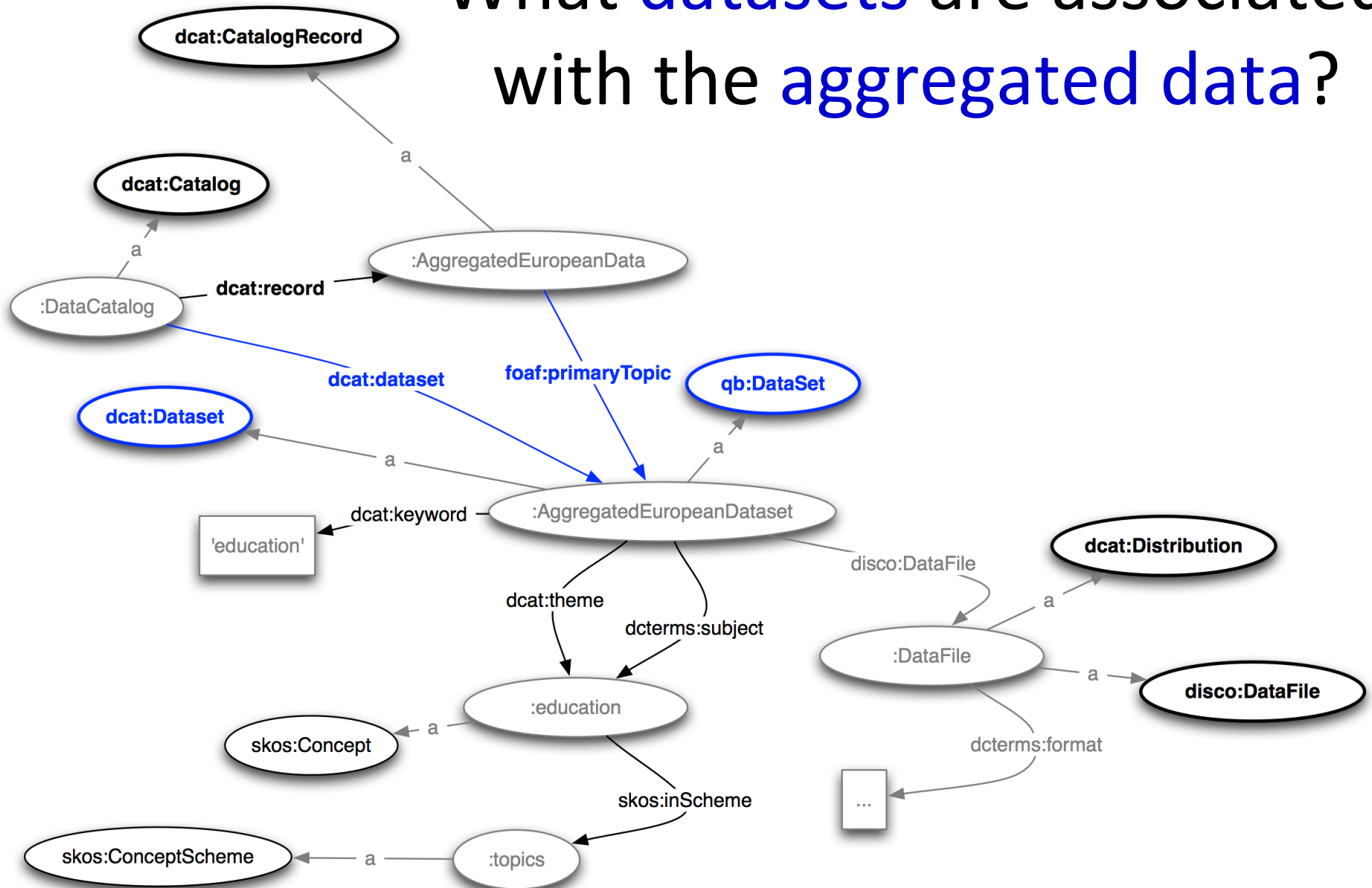
What **datasets** are associated with the **microdata**?



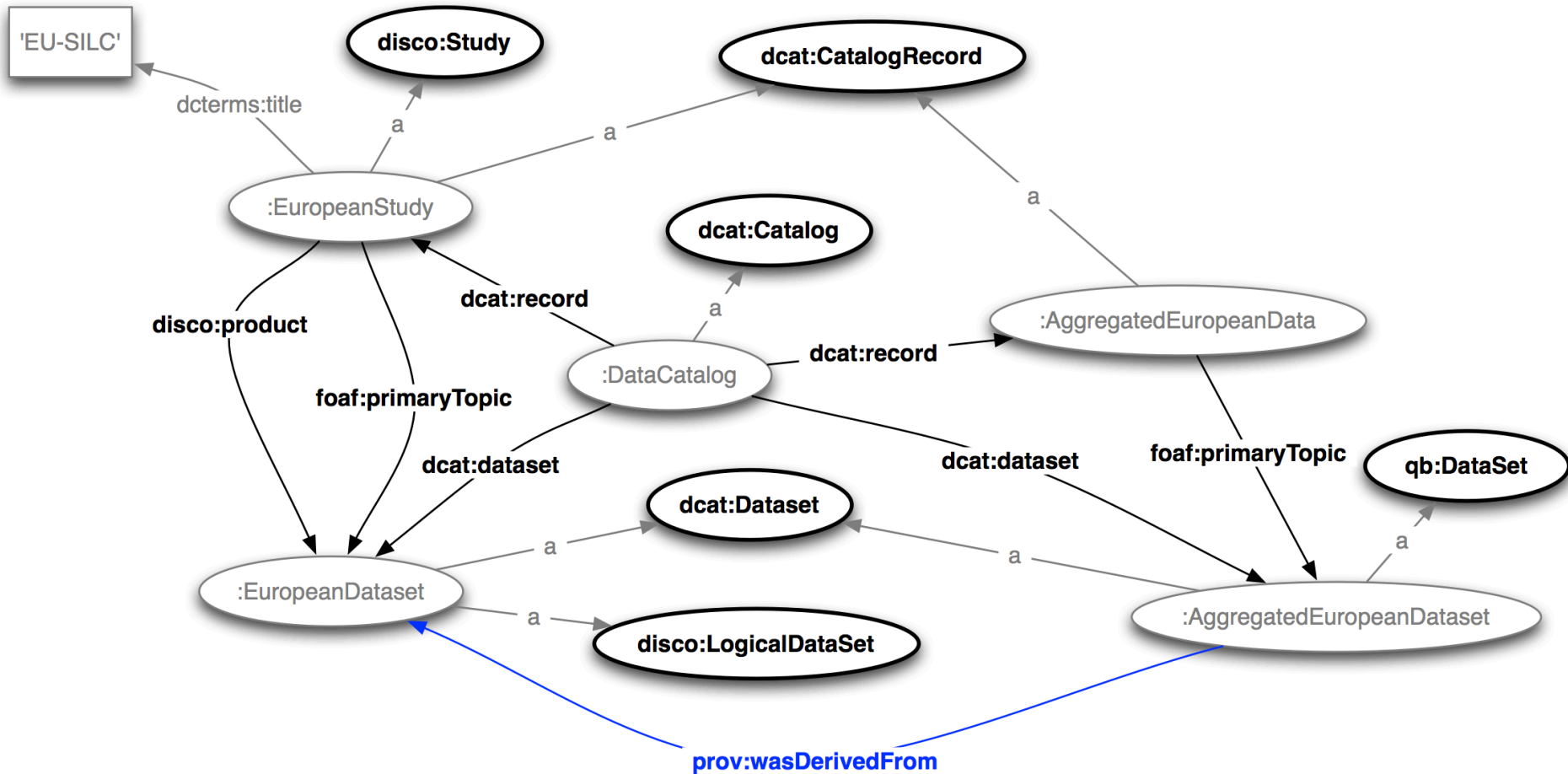
What **aggregated** data according to specific metadata exists?



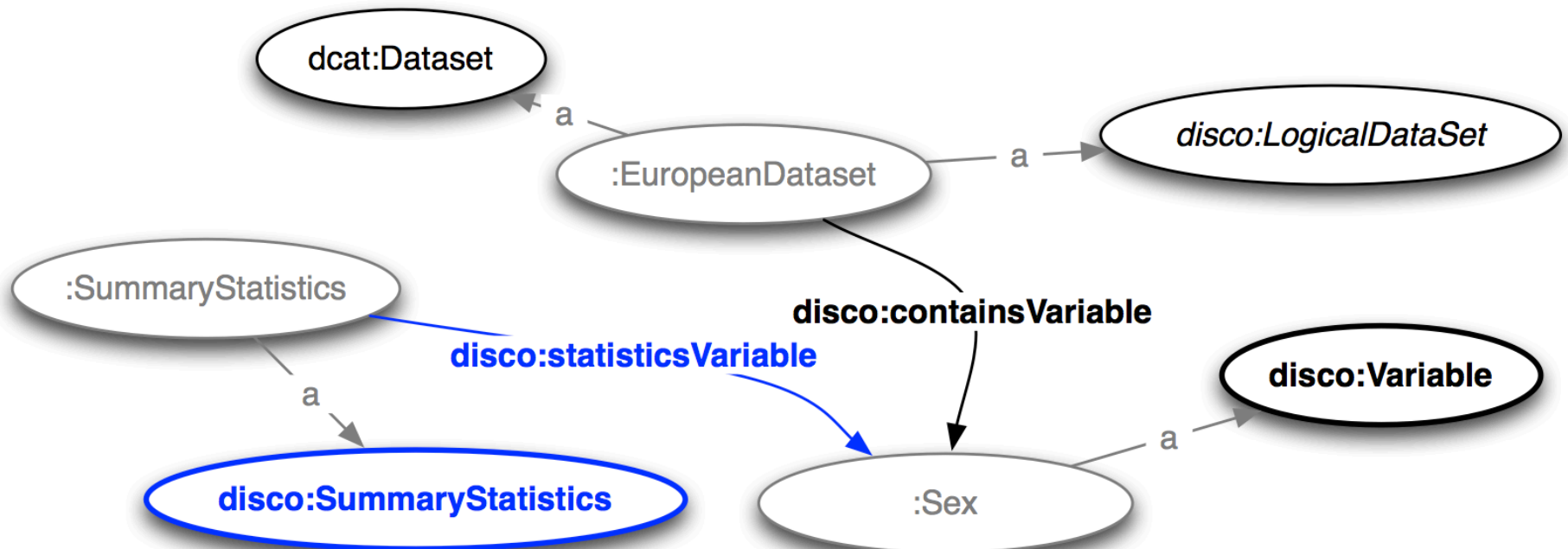
What datasets are associated with the aggregated data?



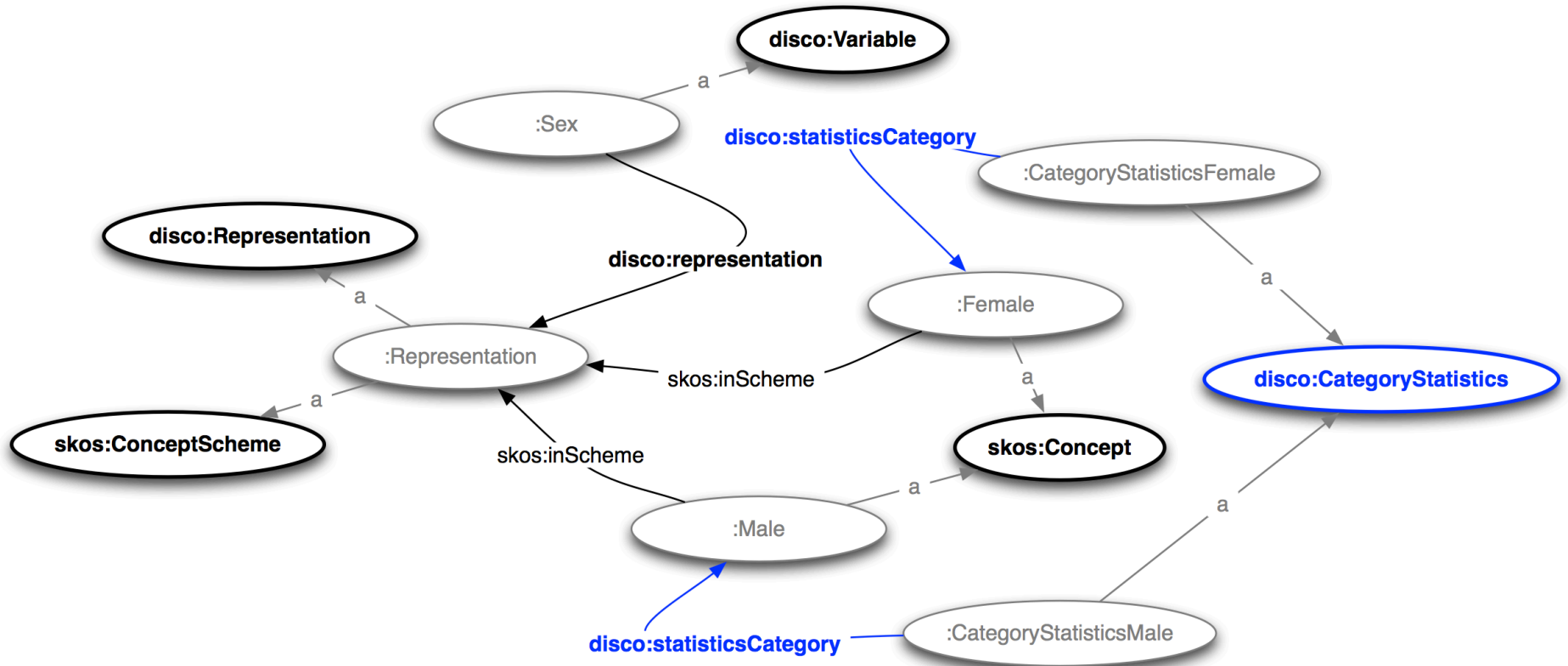
From which microdata datasets is the aggregated dataset derived?



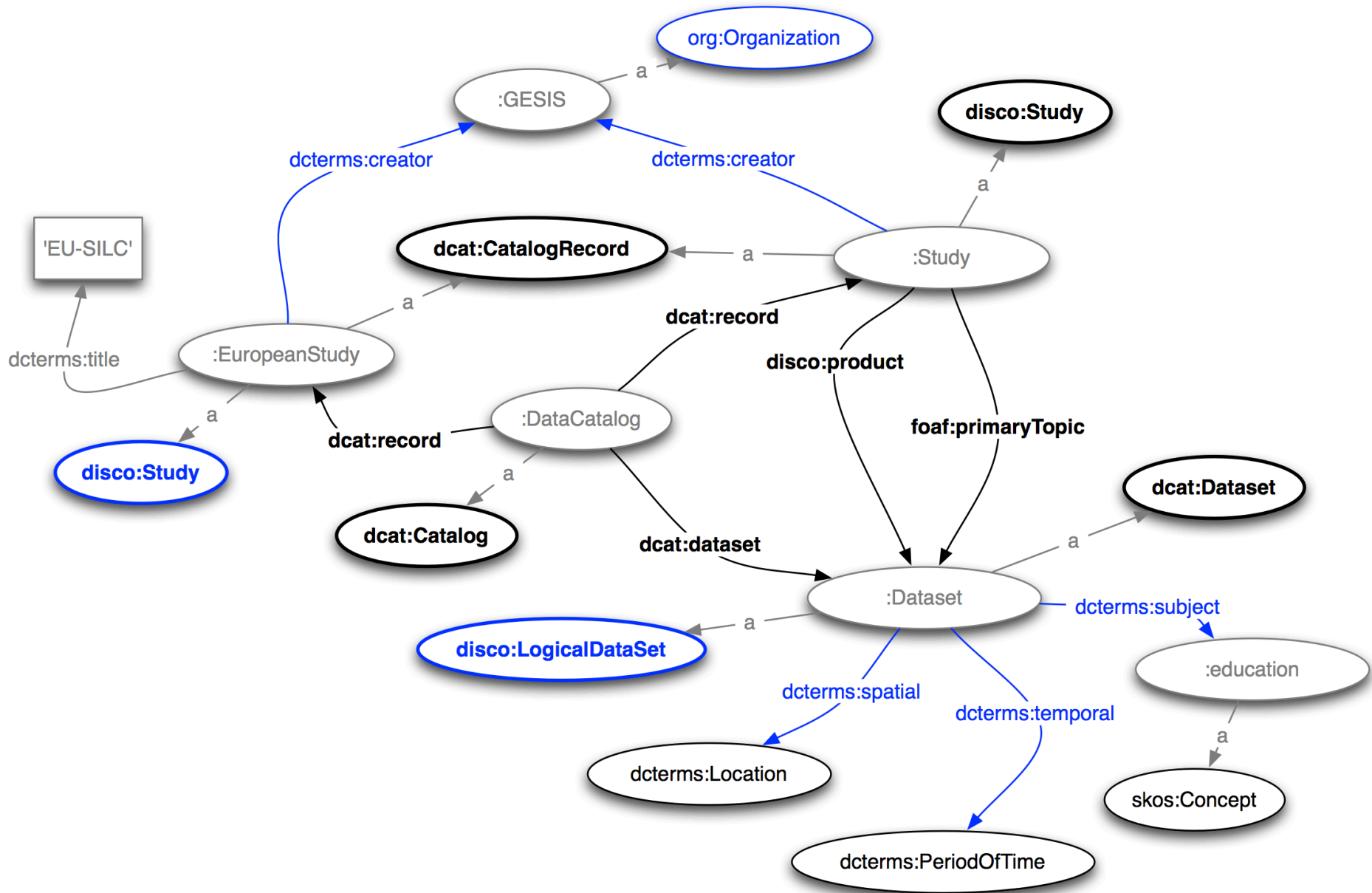
What **summary statistics** does a variable have?



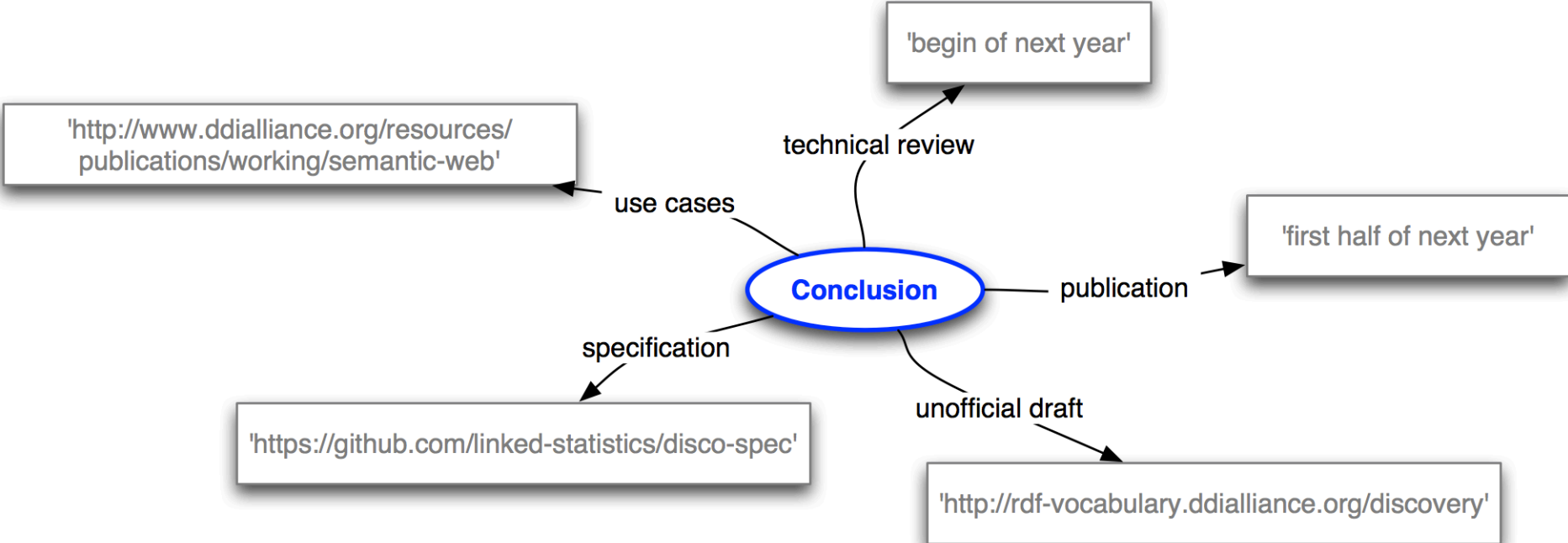
What **category statistics** does a variable representation have?



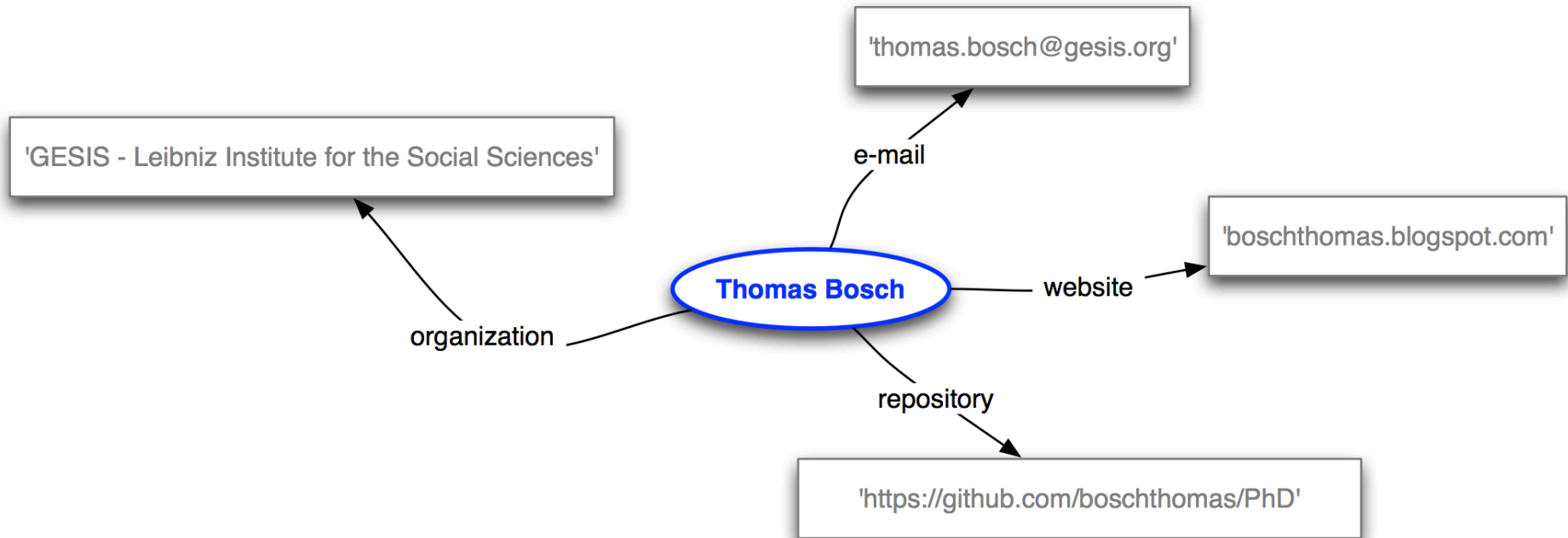
What microdata datasets are created
by the research institute 'GESIS'



Conclusion



Thank you for your attention...



Backup Slides

Why DDI as Linked Data?

- Users discover data in the Linked Open Data Cloud using DDI metadata
- Users can search for data
- Data providers publish searchable / accessible metadata
- The Discovery specification contains the most important DDI concepts for the discovery purpose
- We integrated well elaborated vocabularies
- We use SW technologies

Overview

- Study
- LogicalDataSet: the dataset where we save the actual data
- Universe: for whom is the study applied to? (e.g. all women in Germany)
- Analysis Unit (e.g. persons or households)
- Instrument: How do we want to measure? (e.g. 'What is your sex?')
- Concept: What do we want to measure?
- Variable: Where do we save what we measured? (e.g. sex)

Future Work

- Physical description of rectangular data especially CSV data
 - We integrate this description in discovery
- How originate aggregated data on the basis of microdata?
 - Aggregation method is described in the form machines can process it
 - We see a need that this area should be explored further in order to describe the relationship between aggregate data and microdata more detailed

Acknowledgements

26 experts from the statistical community and the Linked Data community coming from 12 different countries contributed to this work. They were participating in the events mentioned below.

- [1st workshop](#) on 'Semantic Statistics for Social, Behavioural, and Economic Sciences: Leveraging the DDI Model for the Linked Data Web' at Schloss Dagstuhl - Leibniz Center for Informatics, Germany in September 2011
- Working meeting in the course of the 3rd Annual European DDI Users Group Meeting ([EDDI11](#)) in Gothenburg, Sweden in December 2011
- [2nd workshop](#) on 'Semantic Statistics for Social, Behavioural, and Economic Sciences: Leveraging the DDI Model for the Linked Data Web' at Schloss Dagstuhl - Leibniz Center for Informatics, Germany in October 2012
- Working meeting at GESIS - Leibniz Institute for the Social Sciences in Mannheim, Germany in February 2013