BREAL – Big data REference Architecture and Layers

Using BREAL to share components and services – Use cases

**Keywords:** Big Data, standards, enterprise architecture

# Introduction

BREAL serves the purpose of guiding Big Data investments by National Statistical Institutes (NSIs) and helping the development of standardized solutions and services to be shared within the ESS and beyond.

BREAL can be used as a reference framework for enterprise architecture, as a ‘language’ to describe projects using Big data sources, or as helping NSIs top management to plan investments.

# Projects using Big Data sources

During the ESS Net ‘Big Data’, some sub-projects used Big Data sources in order to complete, to improve or to build statistical information.

One project focuses on online job vacancies, and aims to produce statistical estimates for this point. During a first phase of the project, suitable techniques and concrete methodologies were developed. During a second phase, the project implement these points, especially using a mix of web scraping and data coming from the CEDEFOP (European Centre for the Development of Vocational Training).

A second project deals with enterprise characteristics and their improvement using web scraping, text mining and inference techniques. The goal was to improve or update information (internet presence, kind of activity, ownership structure…) in national business registers. As the first project, techniques and methodologies were defined during the first phase, and the implementation during the second one.

# Aligning the projects with BREAL for sharing

Once the projects have implemented the techniques and methodologies needed, an effort was made to describe their process and to align these descriptions with BREAL.

This alignment contributes to improve the BREAL model. It also gives representation of the two projects with the same model (BREAL) using several layers: an application architecture and an information architecture.

Regarding the application architecture, it put in evidence which components can be shared and help to define the way they can be shared. Especially it allows to define an operational model with different sharing ways, from the algorithm implemented in each NSI to a component proposed as a service for every NSI.