Big Multilingual Linked Data

European Data Forum 2015 (EDF2015)
16-17 November 2015, Luxembourg

Structure the data, the how is secondary

* Objective: Clean structured multilingual data

* One model for the three traditions:
  - Big data
  - Multilingual corpora
  - Linked data

* Desired characteristics:
  - Big data – database – terabytes-size
  - Small data – record – kilobytes-size
  - n-lingual – not just bilingual
  - Web friendly – online/offline – URI
  - Longterm friendly – program independence

* Raw data example:
  - Official Journal of the European Union

M. T. Carrasco Benitez
Directorate-General for Translation of the European Commission
http://dragoman.org/edf2015