Breaking public administrations’ data silos:  
*The case of Open-DAI, and a comparison between open data platforms.*

Raimondo Iemma, Federico Morando, Michele Osella

CeDEM14
http://nexa.polito.it/iemma2014open
Objectives

- **Open data platforms:**
  - operate as 'buses' between data centers and the outside world;
  - still, data publication entails a good share of manual work (e.g., taking decisions).

- **Paper:**
  - Isolating **requirements** for state-of-the-art OD platforms.
  - Discuss whether and how four existing **platforms** comply with such requirements.
  - Set **further steps** to enrich/customize our benchmarking.
Requirements for OD portals / 1

- Source: 20 widely diffused policy documents:
  - European Legislation (e.g., PSI Directive, INSPIRE);
  - National / local legislation (e.g., ITA Code for Digital Administration);
  - EU Digital Agenda-related plans (e.g., eGOV action plan);
  - Technical guidelines (e.g., UK Open Standard Principles);
  - Tender specifications (e.g., EU Open Data portals);
  - Independent studies.

- Refinement in two steps
  1) Eliciting a long-list of preliminary requirements drawing on aforementioned sources.
  2) Distilling the short-list of refined requirements.
Publication features

- Automatically extracts from legacy databases
- Uses APIs at the data level
- Promotes the use of standard metadata
- Enables catalogue federation
- Allows to perform a data quality check
- Is designed to be / is integrated with a front-end
- Releases API to reuse data
- Enables browsing at the data level
Architectural features

- Released as open-source software
- Available in a cloud environment
- Available at data holder's premises

Add-ons

- Allows to gather feedback on data (including 'forked' datasets)
- Encompasses a ticketing system
Open Data portals

- **Socrata Open Data portal**: commercial cloud-based service for data publishing, metadata management, data catalogue federation, and exposure of data as services.

- **CKAN** (acronym for Comprehensive Knowledge Archive Network): an open-source data management platform maintained by the Open Knowledge Foundation.

- **Open-DAI**: EU-funded open-source platform designed to expose data as services, directly pulling from legacy databases of the data holder.

- **ENGAGE**: EU FP7-funded data infrastructure supporting scientific collaboration and research.
Benchmarking results

- **Synoptic table**: in the article; and in .csv here: https://github.com/seralf/checklist-piattaforme-opendata

- **Publication features**: direct link with legacy databases is a plus for Open-DAI, particularly suitable for frequently changing data. However, Socrata has strong browsing features, and CKAN is more interoperable (e.g., performs quality check through OpenRefine)

- **Data features**: OK Linked Open Data by the two EU-backed projects.

- **Architecture**: complete options by Open-DAI;

- **Add-ons**: 'ticketing' systems by ENGAGE, tracking bugs and encouraging derived datasets.

- In general: CKAN and Socrata have less sustainability problems (already widely used).