

One Page Overview of ODN for Data Publishers and Data Consumers

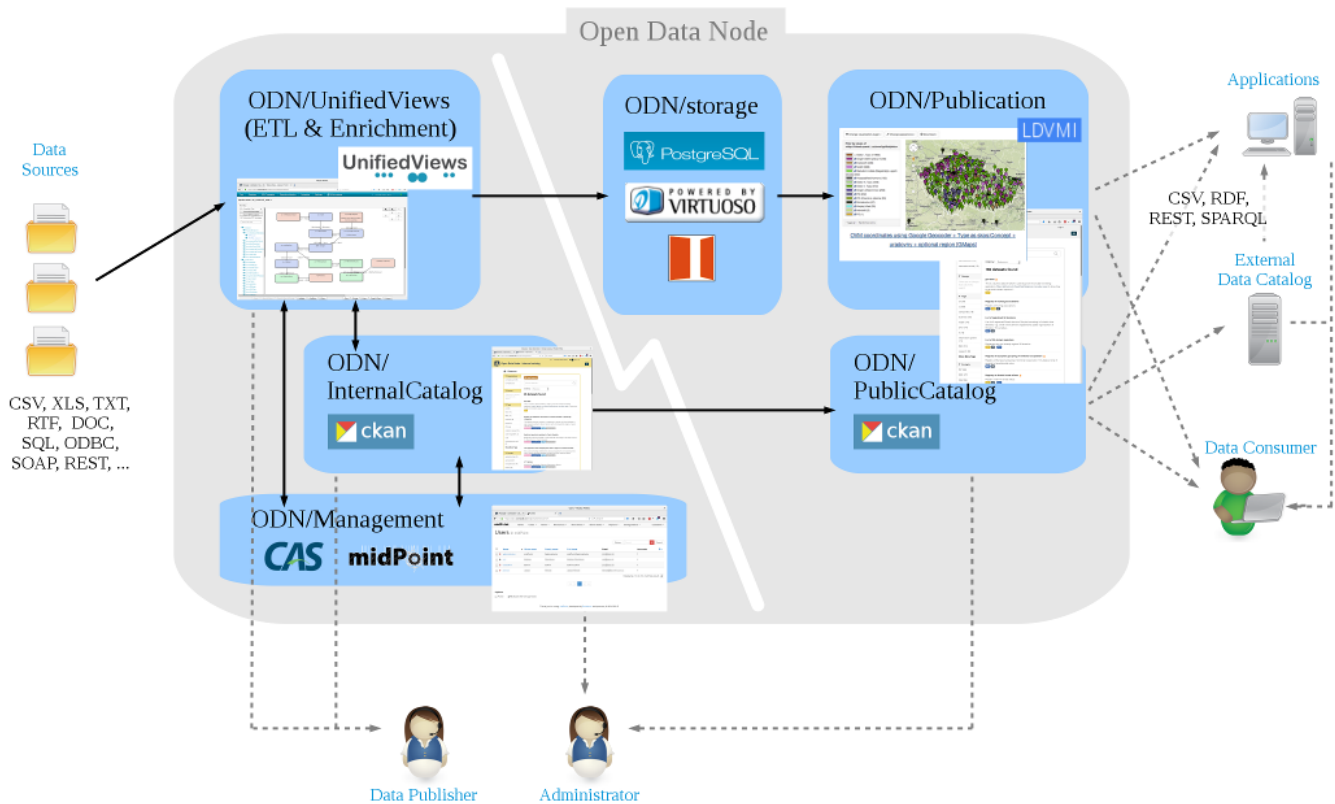


Figure 1 depicts main modules of ODN. Typical workflow for data publishing and data consumption is as follows:

1. Data publisher creates new dataset record in ODN/InternalCatalog.
 1. Data publisher specifies metadata of the dataset record, such as its name, description, license, topic, etc.
 2. Data publisher associates the dataset record with a data publishing pipeline (created either by using wizard or by directly using ODN/UnifiedViews, an ETL tool for RDF data).
 1. Data publisher specifies sources of the dataset (Excel spreadsheets, tables in relational database, REST service etc.).
 2. Data publisher specified how the data sources should be transformed (Certain rows/columns removed/adjusted in the tabular data sources, data sources are anonymised, integrated with other data sources, cleansed etc.)
 3. Data publisher specifies how the data should be published (in which formats they will be available to data consumers - CSV/RDF dumps, REST API, SPARQL Endpoint API)
2. Data publisher clicks "Publish" button to publish data sources for the given dataset as defined.
 1. Publishing pipeline is run in ODN/UnifiedViews, which ensures extraction of data sources, its transformation, and creation of output published data.
 2. ODN/InternalCatalog is updated with links to newly published data.
 3. ODN/Publication module prepares REST API, SPARQL Endpoint API for the published data.
3. Data publisher verifies the publication process, makes the dataset publicly available.
 1. As a result, dataset record in ODN/InternalCatalog is pushed to publicly available ODN/PublicCatalog.
4. Data consumer may access published data (using ODN/PublicCatalog)
 1. Data consumer may download dumps of published data
 2. Data consumer may access published data via REST API/SPARQL Endpoint API
 3. Application developer may access Apache Atom feed to get published data which changed from the last time.
5. Data publisher may update data at any time (by re-running publication process).
 1. Update may be automated (by setting a schedule for repeated re-run of the publication process)
6. (Optional) Data publisher may also configure search strategy for the published dataset, so that data consumers may search the dataset using tailored search engine.

