Italy, discoverability in practice

by Gabriele Ciasullo, Giovanna Scaglione and Antonio Rotundo

Data discoverability is one of the main tasks, next to availability and interoperability, that public policy makers and implementers should take into due consideration in order to foster access, use and re-use of public sector information (PSI), particularly in case of open data.

Users shall be enabled to easily search and find data they need for the most different purposes. That is clearly highlighted in the introduction statements of the INSPIRE Directive, where we can read that "*The loss of time and resources in searching for existing (spatial) data or establishing whether they may be used for a particular purpose is a key obstacle to the full exploitation of the data available*".

Metadata and data portals/catalogues are essential assets to enable that data discoverability. In Italy, AgID (Agency for Digital Italy,) is in charge of managing the National Open Data Portal (dati.gov.it,) and the National Catalogue for Spatial Data (RNDT - geodati.gov.it). These catalogues are the core components of an overall PSI infrastructure to be intended as the knowledge base of all PSI, implemented by the actions defined in the Three-Year Plan for ICT in the Public Administration 2019-2021.

Recently the new version of both catalogues was released with important new features aimed at further improving the user experience in the search, access and use of public data and at facilitating the use of the functionalities for the documentation of data and services by the administrations. Some examples of these new features are:

▸ harvesting by dati.gov.it of data sources consistent with the national metadata profile DCAT-AP_IT;
▸ dataset search for each data provider, as well as for catalogue, under the new section "Administrations, Link opens in a new window" in dati.gov.it; and
▸ the pre-defined view for "priority datasets" in geodati.gov.it as defined under the INSPIRE context.

The main aim of these activities is to make public data available to an ever wider and diversified audience. In this regard, the most significant action, defined in the three-year plan mentioned above, has concerned the interoperable integration and coordination of the two catalogues through the implementation of GeoDCAT-AP, Link opens in a new window to ensure aligned, up-to-date and not conflicting descriptions of spatial data even available as open data. Thanks to that, open spatial data documented in the RNDT are also discoverable in the open data portal, without any other burden for data providers. Linked to that, specific pre-defined views were implemented in both catalogues in order to immediately identify:

▸ open data in geodati.gov.it, using DCAT-AP themes as search criteria;
▸ open spatial data coming from RNDT in dati.gov.it, using the INSPIRE themes as search criteria.

The classification of open spatial data against the DCAT-AP themes is based on the alignments between the controlled vocabularies used in ISO 19115 / INSPIRE metadata and those used in DCAT-AP defined under the ISA[2] programme. Other alignments needed for the implementation of GeoDCAT-AP and the coordination of the two catalogues are being published in the registry managed by AgID too (an example of a cross-map register is available at this URL. Finally, in order to address a wider audience, including non-GI experts, spatial data documented in RNDT is also available through the web search engines,

specifically Google Dataset Search. A relevant current example on how that discoverability, outlined in this blog post, concretely works is the COVID-19 open data, published by the Italian Department for Civil Protection, including national trends, provinces and regions data and areas, under the license CC-BY 4.0.
Metadata for those datasets were published in RNDT as spatial data (e.g. see metadata for the dataset of zones at national level for containing COVID-19 contagion) and, consequently, those datasets are also made discoverable:

▸ in dati.gov.it, as open data;
▸ in the European Data

Portal, as both dati.gov. it and geodati.gov.it are harvested;
▸ in the INSPIRE Geoportal, as geodati.gov.it is harvested;
▸ in Google Dataset Search.

In summary, Open Data are available in an 'universal' way, regardless the search tool used by the users. Additional features are being implemented in both catalogues and will be presented in a future blog post.

Interview on GeoDCAT-AP - Insights from Agency for Digital Italy
Interview with Antonio Rotundo, Geographic Information Expert, and Gabriele Ciasullo, Head of Databases and Open Data Office, Agency for Digital Italy (AgID)

Why did your agency opt for GeoDCAT-AP?
AgID (Agency for Digital Italy) has been working on a national strategy for an overall public sector information infrastructure as knowledge base of all public information resources, implemented by the actions defined in the Three-Year Plan for ICT in the Public Administration.
The core components of that infrastructure are the national catalogue for spatial data (RNDT) and the national open data portal.
The Agency decided to implement GeoDCAT-AP to achieve an interoperable integration and coordination of the two catalogues. The aim was also to ensure aligned, up-to-date and not conflicting descriptions of spatial data even available as open data.
Furthermore, it supports the general objectives to apply the once-only principle and to improve coordination between the INSPIRE implementation and eGovernment, open data and other relevant processes at the national level. This is recommended in several EU official documents, such as the eGovernment Action Plan, the
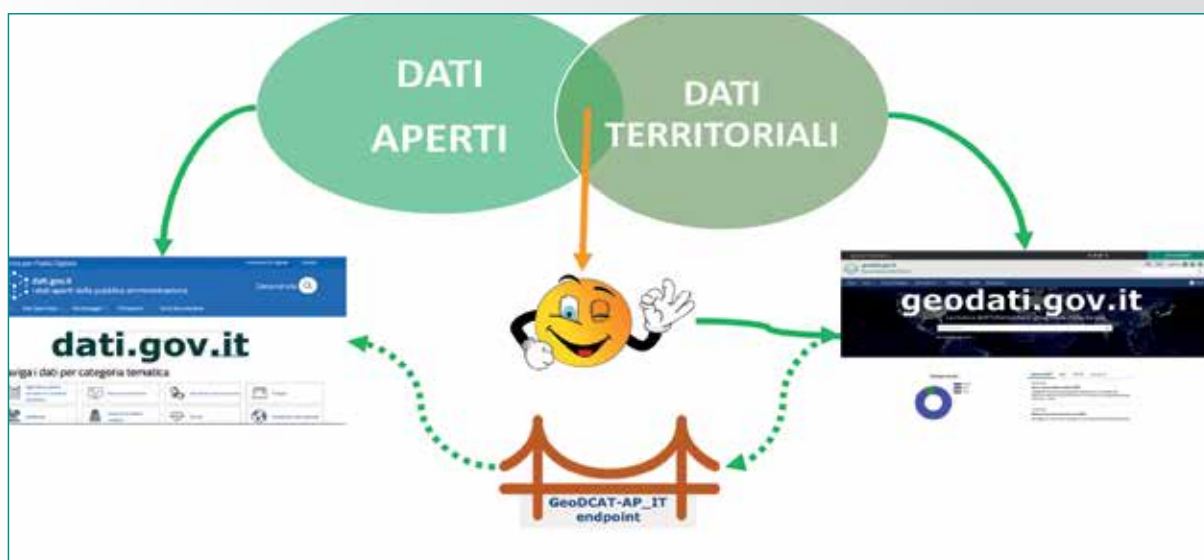
European Interoperability Framework, etc.

How was the implementation of GeoDCAT-AP carried out in Italy?
The implementation of GeoDCAT-AP in Italy (named GeoDCAT-AP_IT) was addressed as a specific action in the previously mentioned Three-Year Plan for ICT in the Public Administration. The Plan is a strategic policy document for all public administrations established by the government which oversees the digital transformation of the country.
The actions foreseen for the implementation of GeoDCAT-AP are:
▸ the definition of national guidelines (published in January 2018);
▸ the implementation of these guidelines and the development of the tools needed for their implementation (already done);
▸ and the full engagement of the Italian organisations managing local catalogues.
The tools developed include the XSLT script extended (to take into account the extensions introduced in the national metadata profiles) and the GeoDCAT-AP_IT API reusing and extending the one developed under the ISA[2] Programme.

Can you give examples of how public administrations in Italy can use GeoDCAT-AP today?

Thanks to GeoDCAT-AP, public administrations can document open spatial data ONLY in the national catalogue for spatial data that will directly enable and provide access to spatial data also in the national open data portal.

Public administrations at the local level are expected to adopt the national approach also for their catalogues. For this purpose, we make the API available for reuse under the European Union Public Licence (EUPL) to facilitate the implementation also at the local level.

Users and developers can also use the API to have a different format (RDF/XML and JSON-LD) for the metadata published in the national catalogue for spatial data.

What are the benefits of GeoDCAT-AP for your organisation? How could it benefit other public administrations?

We can identify benefits both on the side of the public administrations and on the side of the users.

The use of GeoDCAT-AP allows our organisation to make open spatial data available to a wider and diversified audience, beyond the geospatial domain.

Moreover, it allows to avoid the double burden on public administrations concerning the documentation of open geodata in both catalogues. Consequently, GeoDCAT-AP helps overcome the possible misalignments between the metadata for the same dataset published in the two catalogues.

On the side of users, GeoDCAT-AP allows to improve the user experience in the discoverability of data. The more the metadata quality increases, the more effective the search becomes.

Can you give an insight into your involvement in the work on GeoDCAT-AP?

We actively participated in the development and revision of GeoDCAT-AP specification as members of the Working Group established under the ISA² Programme. In that context, we also provided a study comparing ISO 19115:2003, INSPIRE and ISO 19115-1:2014 (the latest version of ISO 19115), included in Annex III to the specification.

Moreover, we contributed to the activity carried out for the definition of the alignments between the controlled vocabularies used in ISO 19115 / INSPIRE metadata and those used in DCAT-AP.

Finally, based on our experience in implementing GeoDCAT in Italy, we will further contribute to improve the GeoDCAT-AP specification by reporting issues and gaps as well as extensions included in the national implementation.

ABSTRACT

In Italy, AgID (Agency for Digital Italy) is in charge of managing the National Open Data Portal (dati.gov.it) and the National Catalogue for Spatial Data (RNDT, geodati.gov.it). Data discoverability is one of the main tasks, next to availability and interoperability, that public policy makers and implementers should take into due consideration in order to foster access, use and re-use of public sector information (PSI), particularly in case of open data.

AUTHOR

Gabriele Ciasullo
ciasullo@agid.gov.it
Giovanna Scaglione
Antonio Rotundo