The PDND E-Service Network: A Graph-Based Model from Italian Open Government Data

Francesco Del Re*[1]

[1]Independent researcher

April 2026

Abstract

The Italian National Digital Data Platform (PDND, *Piattaforma Digitale Nazionale Dati*) is the backbone of public sector interoperability in Italy. It enables more than seven thousand public administrations to exchange data through a catalogue of 14,102 published e-service endpoints (snapshot 2026-05). The platform publishes detailed open data on participating entities and on the e-service catalogue, but it does not publish the pointwise provider-to-consumer pairs that would allow the actual interoperability network to be reconstructed. The number of active agreements is released only as an aggregate figure. This paper documents the methodology adopted in the `pdnd-eservices-graph` project to build a directed weighted graph of the PDND ecosystem from heterogeneous public sources. Nodes and e-services come from official open datasets. Edges are reconstructed by cross-referencing the structured `attributes` field of the catalogue with institutional documentation, including circulars, operational manuals and public presentations. Large language models are used as analytical instruments to extract and normalise relational information from unstructured documents, with every inferred connection traceable to a public source. The result is a graph of 51 nodes and 86 e-service types covering approximately 89% of the catalogue endpoints by mapping and aggregation. The model exposes the topological structure of Italian public sector interoperability and provides a reproducible base for further analysis.

Keywords: PDND, public sector interoperability, bipartite graph, weighted network, entity resolution, open government data, Italy, large language models, methodology

1 Introduction

The PDND is the technical infrastructure that operationalises the *once-only* principle in Italian public administration: a public body should not ask citizens or businesses for data that another public body already holds. The platform exposes a federated catalogue of e-services and brokers the agreements through which providers grant consumers access to specific APIs.

The PDND open data programme publishes three datasets that describe the catalogue from a structural standpoint: the list of adhering entities, the catalogue of e-services with their producers, and a ranking of the most-used e-services with the count of active consumers. A fourth public source, the public dashboard *I numeri della PDND*, exposes aggregate figures such as the total number of enabled connections.

What is absent is the relational layer: the pointwise provider-to-consumer pairs that, taken together, define the topology of the interoperability network. The aggregate count of agreements

*ORCID: 0009-0009-5524-6984. Correspondence: francesco.delre@protonmail.com, github.com/engineering87, linkedin.com/in/francesco-delre

is published. The identity of the two endpoints of each agreement is not. This gap is the starting point of the work documented here.

The contribution is methodological. The PDND ecosystem has to be reconstructed from a heterogeneous set of public sources, where some are structured open datasets and others are institutional documents written for human readers. The reconstruction requires a formal model, a documented sourcing pipeline, an explicit role for large language models as extraction tools, and a clear separation between facts grounded in primary data and inferences drawn from documentation.

The companion implementation is published as an open-source project at `https://github.com/engineering87/pdnd-eservices-graph` and a public interactive visualisation is available at `https://www.pdndgraph.it`. This paper concentrates on the data model and the reconstruction methodology. The visualisation layer is mentioned only as a reference artefact.

2 The PDND Domain

The PDND distinguishes two functional roles. A *provider* (*erogatore*) publishes one or more e-services in the platform catalogue. A *consumer* (*fruitore*) requests authorisation to invoke a specific e-service for a declared finality of use. The same legal entity can be both a provider and a consumer at the same time and frequently is.

An e-service is a versioned API contract. Each e-service belongs to exactly one provider. Access from a consumer to a provider happens through a *voluntary agreement* that binds the two endpoints to a finality of use. The platform issues short-lived tokens that the consumer uses to invoke the service.

Three implications follow for any attempt to model the ecosystem as a graph. First, the ecosystem is naturally bipartite at the role level but the same entity may sit on both sides, so a single set of nodes with directed edges is the appropriate abstraction. Second, an edge between two entities is intrinsically weighted: a single pair may share many e-services. Third, the granularity of the underlying agreements is finer than the granularity of any provider-to-consumer edge, so the graph aggregates information by construction and the aggregation rule must be made explicit.

3 Source Data

Three structured datasets are used as primary sources. All three are published on `dati.gov.it` under the CC0 1.0 Universal licence and mirrored on the `italia/pdnd-opendata` repository [1] on GitHub by PagoPA S.p.A. on behalf of the Department for Digital Transformation of the Italian Presidency of the Council of Ministers.

Dataset	**Content**	**Use in this work**
`aderenti.csv`	Adhering entities, with identifiers, fiscal codes, IPA codes, certified attributes	Node set V
`eservice_a_catalogo.csv`	Catalogue of published e-services with producer identifiers and the `attributes` field	E-service inventory and seed for edge derivation
`eservice_piu_utilizzati.csv`	E-services ranked by number of active consumers (six months and total)	Validation of edge weights

Table 1: Primary structured sources from the PDND open data programme.

The catalogue file is the richest of the three from a relational standpoint, because it includes a JSON-typed field called `attributes` that lists the categories of entities authorised to consume a given e-service. A representative entry has the following shape:

```
{
  "certified": [
    { "name": "Pubbliche Amministrazioni" },
    { "name": "Comuni e loro Consorzi e Associazioni" }
  ]
}
```


This field is structurally close to a relation but semantically distinct. It declares who *may* consume the e-service, not who *is* consuming it. The distinction matters: an authorised category is a permission, not an exercised right. Treating `attributes` alone as a graph would produce a maximalist model with edges to every entity in every authorised category, which would be both incorrect and unreadable.

The aggregate dashboard *I numeri della PDND* [2] reports the global count of enabled connections in the ecosystem, on the order of 6,400 at the time of writing. This figure does not break down by pair and is used in this work only as a sanity check on the order of magnitude of the reconstructed graph.

4 Graph Formalisation

The notation used throughout the paper is summarised in Table 2.

Symbol	**Meaning**
$G = (V, E, w)$	The reconstructed network: nodes, directed edges, weight function.
V	Set of public entities represented as nodes ($\|V\| = 31$ in the current snapshot).
$V_P \subseteq V$	Subset of nodes that act as providers ($d_{\text{out}}(v) > 0$).
$V_C \subseteq V$	Subset of nodes that act as consumers ($d_{\text{in}}(v) > 0$).
$E \subseteq V \times V$	Set of directed edges (a, b), read as "a provides at least one e-service consumed by b".
S	Set of e-services represented in the model ($\|S\| = 57$).
$\text{prov}(s)$	The unique provider of e-service s.
$\text{cons}(s)$	The set of consumers of e-service s as recovered by the pipeline.
$w : E \to \mathbb{N}$	Edge weight: number of distinct e-services flowing from a to b.
$d_{\text{out}}(v)$	Out-degree of v: number of distinct consumers served by v.
$d_{\text{in}}(v)$	In-degree of v: number of distinct providers serving v.

Table 2: Notation used throughout the paper.

The reconstructed network is modelled as a directed weighted graph

$$G = (V, E, w)$$

where V is the set of public entities, $E \subseteq V \times V$ is the set of directed edges, and $w : E \to \mathbb{N}$ is the edge weight function.

A directed edge $(a, b) \in E$ encodes the relation *entity a provides at least one e-service that entity b consumes*. The graph is directed because the relation is not symmetric. The Ministry of the Interior provides ANPR services to municipalities; municipalities do not provide ANPR services to the Ministry of the Interior.

The weight function is defined as

$$w(a, b) \;=\; \big|\{\, s \in S \;:\; \text{prov}(s) = a \;\wedge\; b \in \text{cons}(s) \,\}\big|$$

where S is the set of e-services represented in the model, $\text{prov}(s)$ returns the unique provider of s, and $\text{cons}(s)$ returns the set of consumers of s as reconstructed by the pipeline described in Section 5. The weight is therefore the count of distinct e-services that flow from a to b.

Two alternative weight definitions were considered and discarded. A binary weight, equal to one whenever any e-service flows from a to b, was rejected because it discards the most informative signal in the data, namely the variety of services exchanged. A normalised weight, equal to the share of b's incoming services that come from a, was rejected because the denominator depends on the perimeter of the model and is therefore not stable across versions of the dataset.

Two structural metrics derived from G are used throughout the paper:

- the out-degree $d_{\text{out}}(v)$, that is the number of distinct entities consuming at least one e-service produced by v, which characterises providers and identifies hubs of the ecosystem;
- the in-degree $d_{\text{in}}(v)$, that is the number of distinct entities providing at least one e-service consumed by v, which characterises consumers and identifies entities with broad data needs.

Density and degree centrality are computed in the standard way for directed graphs [8] and are reported in Section 7.

5 Relationship Reconstruction Pipeline

5.1 Why reconstruction is necessary

The pointwise provider-to-consumer pairs are not published as open data. The platform publishes (i) the catalogue of e-services with their producers, (ii) aggregate counts of consumers per e-service in the most-used ranking, and (iii) the global count of enabled connections. None of these three sources gives the identity of the two endpoints of a single agreement.

The reconstruction therefore combines two complementary signals. The structured signal comes from the `attributes` field of the catalogue, which constrains the categories of admissible consumers. The unstructured signal comes from institutional documentation that names individual entities and links them to specific e-services in the context of declared use cases.

5.2 Documentary sources

Five families of documentary sources are exploited. They are listed in Table 3 together with the kind of relational evidence each provides.

Source	**Relational evidence**	**Type**
ANPR access documentation, Ministry of the Interior	Use cases and consumer categories for ANPR e-services (Notification, Communication, Verification, Death Verification)	Circulars (DAIT n. 73/2023 [3], DAIT n. 61/2025 [4])
ANCI/DTD presentation on twenty-six PDND use cases	Explicit listing of providers and consumed databases for the *Welfare as a Service* use cases	Institutional presentation, September 2023 [5]
SSU operational manual, Unioncamere	Six e-services of the SSU catalogue (Metadata, Back Office SUAP, Back Office Third Parties, ComUnica, Business Register)	Manual, version 3, November 2024 [6]
Agenda Digitale and sectoral journalism	Service rollouts, named adopters, ministerial announcements	Articles
Catalogue `attributes` field	Authorised consumer categories per e-service	Structured field in the CSV

Table 3: Documentary sources used to derive provider-to-consumer relations.

These sources are inhomogeneous in format, scope and intent. Some are designed for technical implementers, others for policy audiences, others for the general public. Their union covers the network in patches that overlap on the most visible providers (ANPR, INPS, Agenzia delle Entrate, ANAC, Unioncamere) and thin out on peripheral consumers.

5.3 Pipeline stages

The reconstruction pipeline runs in four stages, illustrated in Figure 1.

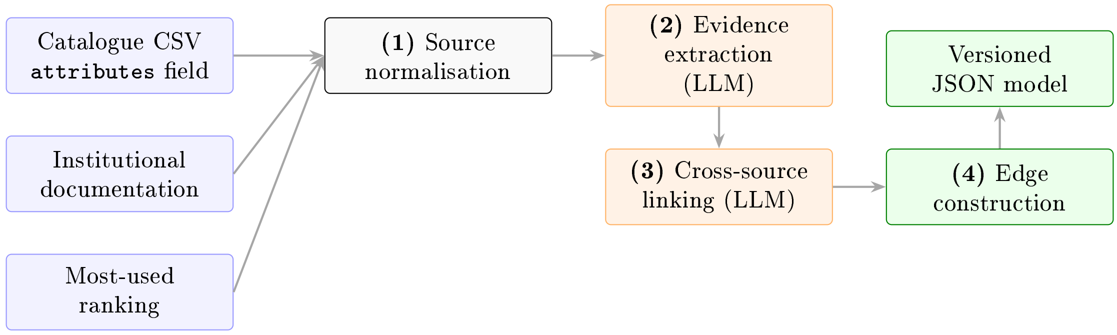

Figure 1: Reconstruction pipeline. Blue boxes are sources, orange boxes are stages where large language models are used as extraction and linking instruments, green boxes are deterministic stages.

Stage 1, source normalisation. The catalogue CSV is parsed into a relational form. The `attributes` JSON is unfolded into a list of authorised categories per e-service. The most-used ranking is joined on the e-service identifier to attach the count of active consumers. Documentary sources are collected as a corpus of plain-text passages tagged with provenance metadata.

Stage 2, evidence extraction. A large language model is prompted to read each documentary passage and emit, for every named pair of entity and e-service, a structured record of the form

$$\langle \text{provider}, \text{e-service}, \text{consumer category or named consumer}, \text{source identifier} \rangle.$$

The model is instructed to extract only assertions explicitly supported by the passage and to mark uncertain attributions for review. The output is a candidate evidence table.

Stage 3, cross-source linking. The candidate evidence is reconciled against the catalogue. Three checks are applied. The provider must match the producer recorded in the catalogue for the named e-service. The consumer category must be a subset of the authorised categories declared in `attributes`. Named consumers must resolve to entities that exist in the adhering-entities dataset. Records that fail any check are flagged. The model also performs entity resolution [7] on consumer names, since the same body appears in different documents under different denominations (for example *Comune di Bologna*, *Comune di Bologna, Settore Agenda Digitale*, and *Bologna*).

Stage 4, edge construction. The reconciled evidence is aggregated into the final graph. For each provider-to-consumer pair, the weight is computed as the count of distinct e-services that survive reconciliation. The result is serialised as the `pdnd-data.json` file consumed by the visualisation layer.

5.4 The role of large language models

Large language models occupy a clearly bounded role in the pipeline. They are used as extraction and linking instruments on the corpus of documentary sources. They are not used as generators of facts.

Concretely, the model is asked to do three things. First, to read a passage and produce structured tuples that summarise its relational content. Second, to match entity mentions across passages, given the wide variation in how Italian public bodies are named. Third, to flag passages where the relational claim is implicit or ambiguous, so that a human reviewer can decide whether to retain it.

Three guarantees follow from this scoping. Every edge in the graph is traceable to at least one tuple produced in stage 2 and reconciled in stage 3. Every tuple carries the identifier of the source passage that supports it. No edge rests on model output alone: either a structured field of an open dataset constrains it, or a documentary source explicitly states it.

This design treats the model as an instrument for systematising work that a human analyst could in principle do alone, given enough time, on a corpus of dozens of documents written in different registers and formats. The model speeds up the systematisation. The documentary grounding remains the basis of every claim.

6 Modelling Choices and Aggregations

6.1 Aggregation of small municipalities

The PDND catalogue contains 14,102 published e-service endpoints in the snapshot used for this paper. The majority are standardised services, such as the digital noticeboard (*Albo Pretorio*, 1,451 endpoints), the one-stop business desk interface (*SUAP*, 1,429 endpoints), and the protocol service (*Protocollo Informatico*, 1,296 endpoints), that each of the approximately seven thousand five hundred adhering municipalities publishes individually. Representing every municipality as a separate node would produce a graph of several thousand nodes whose structure is dominated by replication, not by interoperability.

The model adopts an explicit aggregation rule. Six large municipalities (Milan, Rome, Naples, Bologna, Genoa, Padua) are kept as individual nodes because they publish non-standard, locally specific e-services in addition to the standardised ones. All other adhering municipalities are collapsed into a single aggregate node labelled *Comuni (aggregati)*. The aggregate node is treated as a single endpoint of edges. An edge from INPS to the aggregate node does not assert that all seven thousand five hundred municipalities consume the service, but that the service is structurally available to the category. This semantic distinction is recorded in the metadata accompanying the node.

6.2 Selection of represented entities

The graph includes fifty-one entities, selected with three criteria. All central public administrations that provide e-services documented either in the catalogue or in institutional sources are included. Six large municipalities are included as individual nodes. All regions and autonomous provinces with at least twenty-five published endpoints in the catalogue are included as individual nodes (eighteen regional administrations in total), with two roles: they consume from central administrations and they expose their own regional services to municipalities and to other central bodies. The aggregate municipalities node closes the model.

The selection threshold for the regional layer was set at approximately twenty-five published endpoints. This avoids fragmenting the graph with administrations that have a marginal catalogue presence and at the same time keeps the regional layer almost complete: only a few small administrations remain unrepresented.

The selection is intentionally conservative. Entities for which no documentary evidence of either provision or consumption could be located are not included, even if they appear in `aderenti.csv`, in order to keep the graph faithful to verifiable relations.

6.3 Categorisation

Each entity is assigned to a category according to its prevalent institutional function. The categories are: *Ministero*, *Previdenza*, *Fisco*, *Digitale*, *Lavoro*, *Imprese*, *Anticorruzione*, *Trasporti*, *Cultura*, *Statistica*, *Tecnologia*, *Regione*, *Comune*, *Comuni Aggregati*. The categorisation is used in the visualisation layer for colouring and filtering. It is descriptive, not analytic, and does not enter the graph definition.

7 The Resulting Network

The reconstructed graph contains $|V| = 51$ nodes, $|E| = 343$ directed edges and 86 distinct e-service types covering approximately 89% of the 14,102 published endpoints in the official catalogue. Table 4 summarises the structural metrics computed on the current snapshot.

Metric	Value
Nodes $\|V\|$	51
Directed edges $\|E\|$	343
Distinct e-service types modelled	86
Provider nodes (with $d_{\text{out}} > 0$)	48
Consumer nodes (with $d_{\text{in}} > 0$)	43
Mixed-role nodes (both roles)	40

Table 4: Aggregate metrics of the reconstructed network.

7.1 Schematic structure

Figure 2 renders the qualitative shape of the network. Provision is concentrated in a small set of central hubs. Consumption is distributed across three structural classes that follow the aggregation rules described in Section 6. The diagram does not attempt to draw the full edge set: it conveys the topology rather than the individual links.

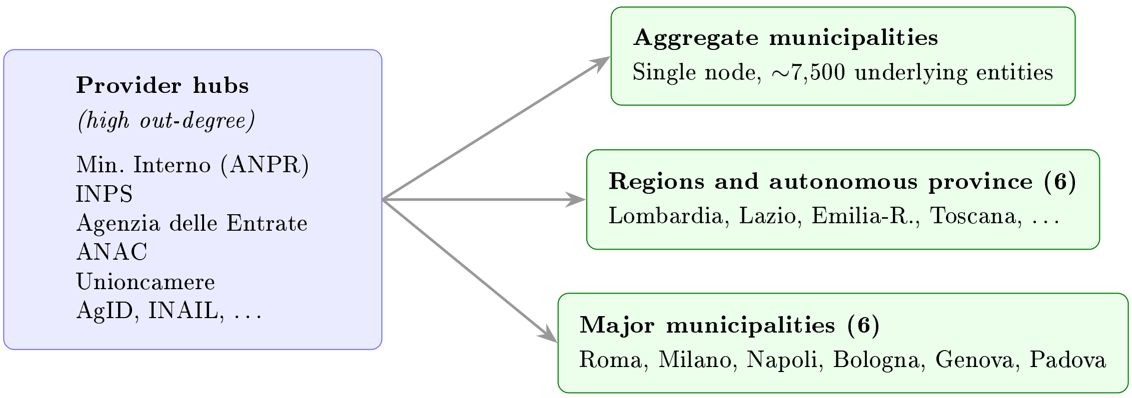

Figure 2: Schematic structure of G. A small set of central hubs provides e-services to three structural classes of consumers. The classes reflect the aggregation rules of Section 6: a single aggregate node for the seven thousand five hundred small municipalities, six regions and one autonomous province, and six major municipalities represented individually. Diagram is qualitative.

7.2 Degree distribution

Figure 3 reports the in-degree and out-degree distributions. Before reading the shape of the two histograms, the aggregation rule of Section 6 has to be kept in mind: the seven thousand five hundred small municipalities are collapsed into a single node, so the node-level degree counts in this graph are not directly comparable to the per-entity counts that an unaggregated graph would produce. Specifically, the aggregate node concentrates on a single point the in-degree and out-degree that would otherwise be distributed across thousands of nodes at low degree, and it inflates by construction the upper tail of both distributions.

With that caveat, the distributions are different in shape. The out-degree distribution has a marked peak around four-five (fifteen nodes at out-degree four, ten at five), corresponding to the regional cluster: each region produces one regional service and is also a consumer of the standardised central services, which gives it an out-degree concentrated in a narrow band. Three nodes have no outgoing edges, eight nodes have no incoming edges, and the upper tail at degree eleven or more contains seven providers and ten consumers — these are the central administrations and the most active entities. The in-degree distribution is broader, with eleven nodes at in-degree two and a long tail upward, reflecting the variety of consumption profiles: some bodies consume from a handful of providers, others from many. The bimodal pattern visible in both distributions, with a cluster at low degree and a separate cluster at high degree, is in part a real feature of the ecosystem and in part an artefact of aggregation; an unaggregated graph would distribute the upper tail across many more low-degree nodes.

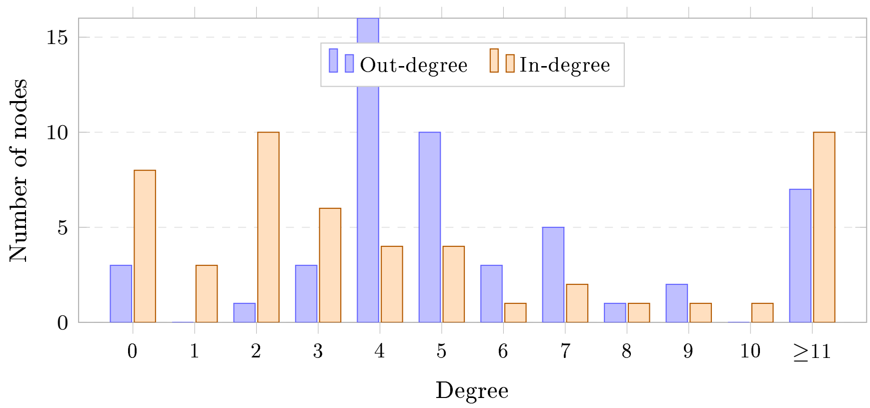

Figure 3: Distribution of node degrees in G. Both distributions show a peak at degree zero, a broad middle, and a tail at degree eleven or more. The shape is partly a feature of the ecosystem and partly a side effect of the aggregation rule of Section 6, which collapses the small municipalities into a single node and inflates the upper tail of the distribution. The rightmost bin aggregates all degrees of eleven and above.

7.3 Top providers and top consumers

Figure 4 ranks the seven highest-out-degree providers and the seven highest-in-degree consumers. The aggregate municipalities node appears in both rankings: this is a direct consequence of the aggregation rule and not a property of any single municipality. Treating the same node as one endpoint of both an outgoing and an incoming relation is the price paid to keep the diagram readable; the alternative would have been a graph of several thousand nodes dominated by replication, where individual hubs would not be visible at all.

INPS and the Ministry of the Interior (ANPR) lead the provider ranking with thirty-three and thirty-four distinct consumers respectively, followed by Unioncamere and the Italian Revenue Agency at seventeen, the aggregate municipalities node at fourteen, the Ministry of Labour at thirteen, and AgID at eleven. INPS and ANPR dominate because their core services (ISEE, demographic verification, social security) are consumed by every region and large-municipality node in the model, plus by the aggregate node and by other central administrations.

The consumer ranking shows that the most active receivers are the aggregate municipalities node at thirty-seven distinct providers, INPS at thirty-four, ISTAT at thirty-three, the Ministry of Labour at twenty-five, the Region of Lombardy at twenty-one, the MEF at seventeen and the Comune di Milano at fourteen. Three of the seven top consumers (Comuni aggregati, INPS, Ministero del Lavoro) also appear among the top providers, and for INPS and Min. Lavoro the dual role reflects a genuine property of the ecosystem rather than an artefact of aggregation: both bodies provide their own services and consume from many regional and central counterparts.

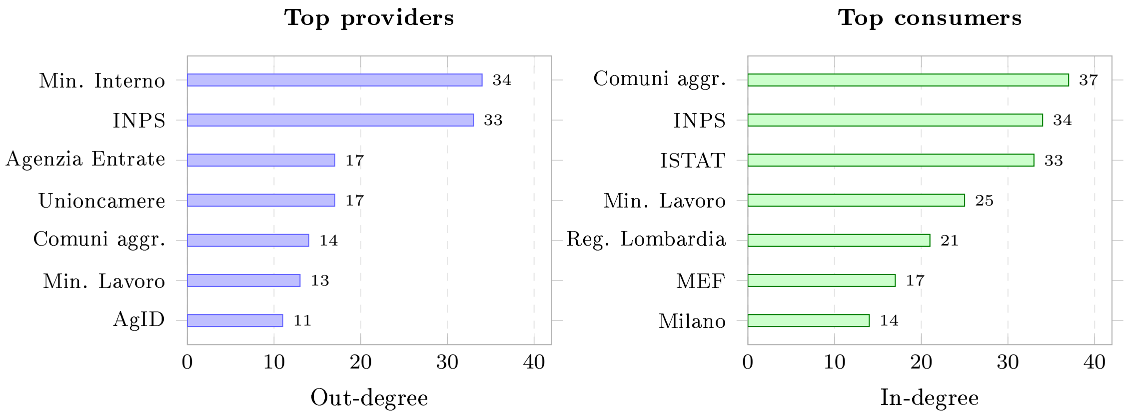

Figure 4: Top-7 providers ranked by out-degree (left) and top-7 consumers ranked by in-degree (right). *Comuni aggr.* is the aggregate node that stands for the seven thousand five hundred adhering municipalities not modelled individually, and it appears in both rankings as a side effect of the aggregation. Three of the seven top consumers (INPS, Min. Lavoro, Comuni aggr.) also rank among the top providers; for INPS and Min. Lavoro this reflects a genuine dual role.

7.4 Order of magnitude

The reconstructed network contains 343 directed edges, well below the global figure of approximately 6,400 enabled connections reported by the official aggregate dashboard. The discrepancy is expected. The model represents structural relations between entities and aggregated categories, while the official figure counts every individual agreement, including those between thousands of small municipalities and the same provider. The collapse of the seven thousand five hundred small municipalities into a single aggregate node accounts for the bulk of the difference. Measured against the 14,102 published e-service endpoints in the catalogue, the model covers approximately 89% by direct mapping and aggregation.

8 Validation

Three layers of validation are applied to the model.

Source-traceability check. Every edge in the JSON output carries an array of source identifiers that point to the documentary passages or structured fields that support it. A script verifies that no edge is empty of provenance. This check is part of the build pipeline of the visualisation.

Authorisation consistency. Every edge from a named consumer to an e-service is checked against the `attributes` field of the same e-service in the catalogue. An edge is retained only if the consumer belongs to one of the authorised categories. This check rules out hallucinated consumers.

Aggregate sanity check. The total count of edges and the distribution of edge weights are compared against the public aggregate counts on *I numeri della PDND*. Order-of-magnitude consistency is required. A reconstructed model that produced ten thousand edges or one hundred edges, given the official figure of about 6,400 individual agreements, would be flagged as inconsistent.

The validation is qualitative on the level of individual edges, in the sense that no exhaustive ground truth is available. It is quantitative on the level of aggregates, where official figures provide a sanity baseline.

9 Reproducibility

The artefacts required to reproduce the graph are listed in Table 5. The dataset `pdnd-data.json` is the canonical output of the pipeline and is the object that any downstream analysis should consume. The structured upstream sources are versioned by date because the official portal does not assign explicit version numbers to the snapshots. The documentary sources are versioned by their official numbering when available (DAIT circulars, Unioncamere manual) and by retrieval date otherwise. The large language model used in stages 2 and 3 of the pipeline is reported by family and version, in line with current practice for studies that delegate part of the extraction to a generative model.

Artefact	Reference
Repository	`https://github.com/engineering87/pdnd-eservices-graph`
Canonical dataset	`src/data/pdnd-data.json` (current snapshot of G)
Methodology document	`METODOLOGIA.md` in the repository
Upstream open data	`italia/pdnd-opendata`, snapshot of April 2026
DAIT circulars	n. 73 of 31 May 2023; n. 61 of 23 June 2025
Unioncamere manual	SSU Operational Manual, version 3, November 2024
ANCI/DTD presentation	*Twenty-six PDND use cases*, September 2023
Aggregate dashboard	*I numeri della PDND*, retrieved April 2026
LLM family used	Claude (Anthropic), Sonnet/Opus class, accessed April 2026

Table 5: Artefacts and references required to reproduce the reconstructed graph.

A formal schema of the canonical dataset is provided in Appendix A. Every record carries the source identifiers that justify it, so a reviewer can trace any edge in G back to the documentary passage or the structured field that supports it.

10 Limitations

Six limitations are explicit and relevant to any use of the dataset.

The provider-to-consumer pointwise data is not open. The model relies on documentary reconstruction for the relational structure. The quality of the reconstruction is bounded by the coverage of the documentary sources, which is uneven across providers and over time.

The aggregation of small municipalities trades faithfulness for readability. The seven thousand five hundred individual bodies are collapsed into a single node, which has two consequences. First, any research question that requires per-municipality resolution cannot be answered on this graph and would need a different model with all municipalities as separate nodes. Second, the aggregation distorts the observed topology by construction: it concentrates on a single point the in-degree and out-degree that would otherwise be distributed across thousands of low-degree nodes, it inflates the upper tail of the degree distribution, and it makes the aggregate node appear simultaneously among top providers and top consumers. The structural readings reported in Section 7 should be interpreted in light of this distortion.

The model covers approximately 89% of the published catalogue endpoints. The residual 11% (about 1,600 endpoints) is concentrated on a long tail of producers with fewer than twenty endpoints each, dominated by individual universities, music conservatories and academic institutions, plus a small number of regional in-house ICT companies. These bodies are not currently represented as individual nodes because their inclusion would weight the graph towards the academic sector without changing the structural picture of inter-administrative interoperability. A future extension may aggregate them under an academic-sector node, in line with the convention used for the small municipalities.

The graph is a snapshot. The catalogue updates daily and agreements come and go. The model captures the state of the ecosystem at the time of compilation. Versioning of the JSON file is the mechanism that allows past states to be retrieved.

The categorisation reflects an institutional reading of the Italian public administration that is fit for navigation but not for analytic typologies. Researchers interested in functional taxonomies should derive their own labels from the underlying `attributes` and `producerId` fields.

E-services with status `SUSPENDED` or in beta are not represented unless documentary sources explicitly attest to their use. This excludes some forward-looking parts of the catalogue.

11 Future Work

Three directions extend this work without altering its methodological core.

The first is automation of the ingestion of the structured sources. The catalogue and the most-used ranking can be re-fetched on a schedule and the model can be re-built incrementally, with diff reports against the previous snapshot. This would make the graph a living object.

The second is the formal export of the graph in standard interchange formats such as GraphML and GEXF, so that the dataset can be loaded directly in network analysis tools (Gephi, NetworkX, igraph) without parsing the project-specific JSON. This is purely a serialisation extension and does not affect the model.

The third, and the most consequential, is opening a structured dialogue with PagoPA and the Department for Digital Transformation around the publication of provider-to-consumer pairs as open data, with appropriate aggregation rules. The reconstruction effort documented in this paper would become unnecessary if the platform itself published what it already records internally for the management of agreements.

12 Code and Data Availability

The implementation that consumes the JSON model and renders the interactive visualisation is published as open source at https://github.com/engineering87/pdnd-eservices-graph under the AGPL-3.0 licence. The methodological documentation, including the full list of documentary sources, is in the file `METODOLOGIA.md` of the same repository. The reconstructed dataset is the file `src/data/pdnd-data.json`, distributed under the same terms as the implementation. The upstream PDND open data are released under CC0 1.0 by PagoPA S.p.A. on behalf of the Department for Digital Transformation. A public deployment of the visualisation is available at https://www.pdndgraph.it.

13 Acknowledgements

This work draws on the open data programme of the Italian Department for Digital Transformation and on the maintenance work of PagoPA on the `italia/pdnd-opendata` repository. The institutional documents listed in Section 5 are themselves a public good, and their authors deserve credit for the level of operational detail they make available.

References

[1] PagoPA S.p.A., *PDND Open Data*. GitHub repository, retrieved April 2026. https://github.com/italia/pdnd-opendata.

[2] Department for Digital Transformation, *I numeri della PDND*. Public dashboard, retrieved April 2026. https://www.interop.pagopa.it/numeri.

[3] Ministry of the Interior, Department for Internal and Territorial Affairs, *Circular DAIT no. 73 of 31 May 2023: ANPR access for municipalities*. Retrieved April 2026. https://www.anagrafenazionale.interno.it/area-tecnica/accesso-ai-dati/.

[4] Ministry of the Interior, Department for Internal and Territorial Affairs, *Circular DAIT no. 61 of 23 June 2025: ANPR access extension to regions and autonomous provinces*. Retrieved April 2026.

[5] ANCI and Department for Digital Transformation, *Twenty-six PDND use cases: Welfare as a Service and beyond*. Institutional presentation, September 2023, retrieved April 2026. https://www.anci.it/wp-content/uploads/Presentazione-PDND_ANPR_ANCI_sett.2023.pdf.

[6] Unioncamere, *SSU Operational Manual: E-Service Catalogue*. Version 3, November 2024, retrieved April 2026. https://catalogo.impresainungiorno.gov.it/assets/config/files/manuale_operativo_Eservice_CatalogoSSU.pdf.

[7] P. Christen, *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Springer, 2012. doi:10.1007/978-3-642-31164-2.

[8] M. E. J. Newman, *Networks: An Introduction*. Oxford University Press, 2010. ISBN 978-0-19-920665-0. doi:10.1093/acprof:oso/9780199206650.001.0001.

[9] F. Del Re, *pdnd-eservices-graph*. Open-source project, AGPL-3.0 licence. GitHub repository: https://github.com/engineering87/pdnd-eservices-graph. Live deployment: https://www.pdndgraph.it.

A Canonical Dataset Schema

The canonical output of the pipeline is the file `src/data/pdnd-data.json` of the repository. The file contains a single object with two top-level arrays: `nodes` and `edges`. Each node represents a public entity included in the model. Each edge represents a directed provider-to-consumer relation, weighted by the number of distinct e-services that flow between the two endpoints. Listing 1 reports an annotated excerpt that exemplifies both record types.


```
{
  "nodes": [
    {
      "id": "min-interno",
      "label": "Ministero dell'Interno",
      "category": "Ministero",
      "role": "provider",
      "description": "Provider of ANPR e-services."
    },
    {
      "id": "comuni-aggregati",
      "label": "Comuni (aggregati)",
      "category": "Comuni Aggregati",
      "role": "consumer",
      "description": "Aggregate node for the ~7,500 Italian municipalities not
          represented individually."
    }
  ],
  "edges": [
    {
      "source": "min-interno",
      "target": "comuni-aggregati",
      "weight": 4,
      "services": [
        "ANPR-Notifica",
```

```
        "ANPR-Comunicazione",
        "ANPR-Verifica",
        "ANPR-VerificaDecesso"
      ],
      "sources": ["DAIT-73-2023", "attributes:eservice_a_catalogo"]
    }
  ]
}
```

Listing 1: Annotated excerpt of `pdnd-data.json`.

The fields of a node record are the stable identifier (`id`), the display label (`label`), the institutional category used for visual grouping (`category`), the role flag that records whether the entity acts as provider, consumer or both (`role`), and a free-text description.

The fields of an edge record are the source and target node identifiers, the integer weight, the explicit list of e-services that contribute to the weight (`services`), and the array of source identifiers (`sources`) that document each edge. The `sources` array is the implementation of the traceability guarantee discussed in Section 5.4: every edge points back to at least one structured field of an open dataset or to at least one documentary passage.