

# Dealing with Open Data at ISTAT

First Steps Towards a Perfect Data Portal:

a Cutting-edge Approach for Dealing with Open Data

Giovanni A. Barbieri

Stefano De Francisci

Italian National Statistical Institute (Istat)

- Our proposal is simple: [...] the federal government [...] is to provide data that is easy for others to reuse, rather than to help citizens use the data in one particular way or another
- Open infrastructures that enable citizens to make their own uses of the data
- Reverse the current policy, which is to regard government websites themselves as the primary vehicle for the distribution of public data, and open infrastructures for sharing the data as a laudable but secondary objective [Robinson, Yu, Zeller and Felten 2008]

- Government information that is nominally publicly available is in fact difficult to access either because it is not online or, if it is online, because it is not available in useful and flexible formats [Brito 2008]
- “Structured data”
  - Associated structured XML file would allow a user to sort the data by ascending or descending date, alphabetically by headline or author, by number of words, and in many other ways

# Why is Linked Data an Opportunity?

## Linked Data as a **semantically rich paradigm** for data representation

- Rich enough for the strict requirements of Official Statistics
- Formal and well-defined data structures, i.e. ontologies

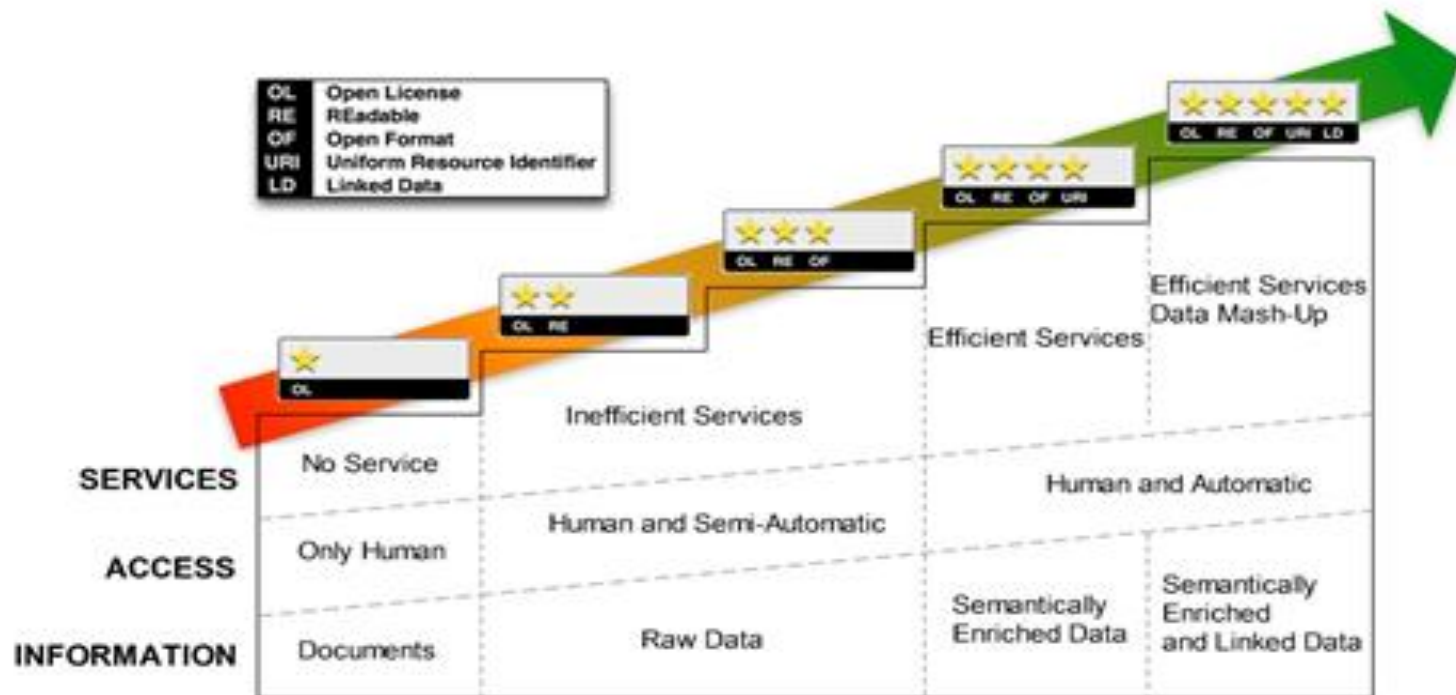


## Linked Data as an **international standard (W3C)**

- Tools availability and independence
- **Beyond** statistical users
- RDF: Resource Description Framework (W3C) (subject-predicate-object)

# The 5-Star Model

|                  |       |                             |
|------------------|-------|-----------------------------|
| Data             |       |                             |
| Open Data        | *     | Open License                |
| Open Data        | **    | Readable                    |
| Open Data        | ***   | Open Format                 |
| Linked Open Data | ****  | Uniform Resource Identifier |
| Linked Open Data | ***** | Linked Data                 |



# Italian Strategies: Digital Agenda and Open Data

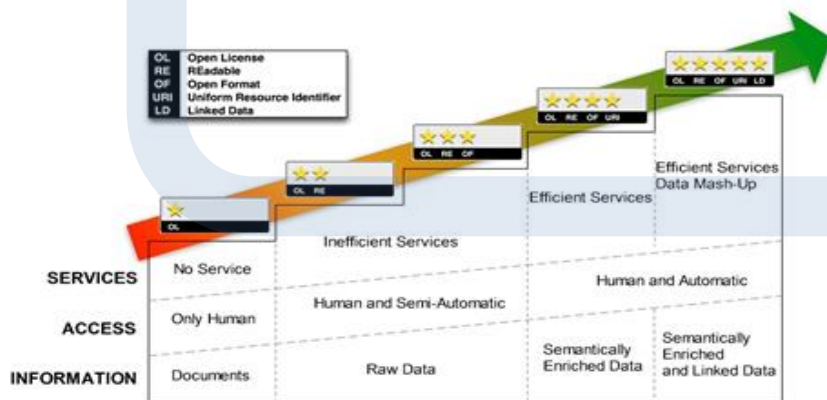


Agenzia per l'Italia Digitale  
 Presidenza del Consiglio dei Ministri

Publication and sharing of national public-sector information



National guidelines for the development of public-sector information



Linked data paradigm for enabling semantical interoperability within the public sector

# The National Open Data Portal

## National Open Data Portal

<http://www.dati.gov.it/>

Published and maintained by AgID

More than 10,000 datasets

Provided by 76 public bodies

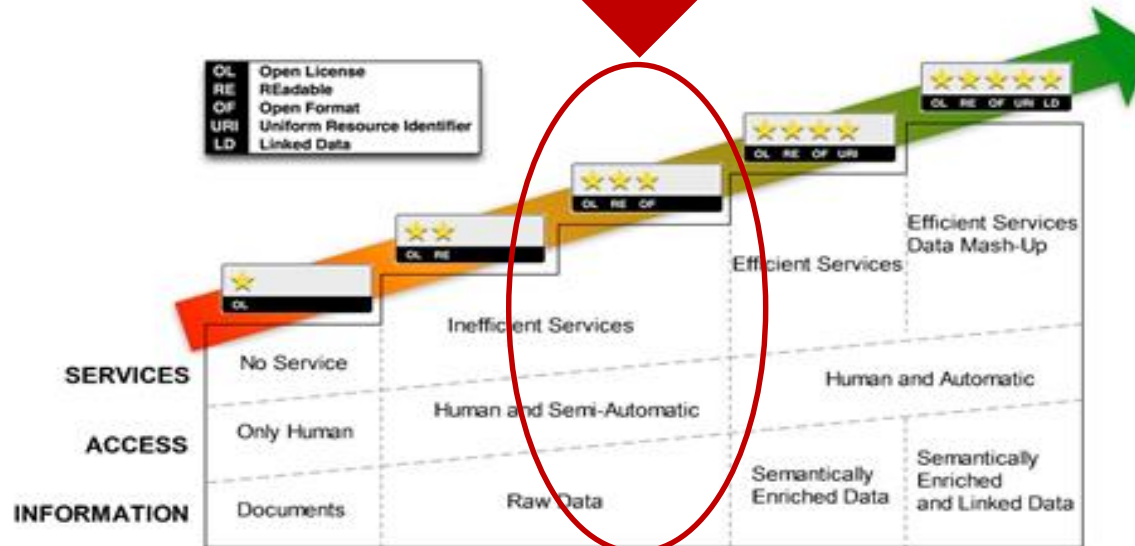
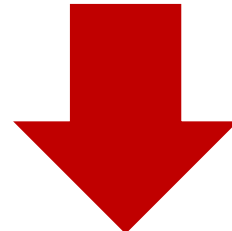
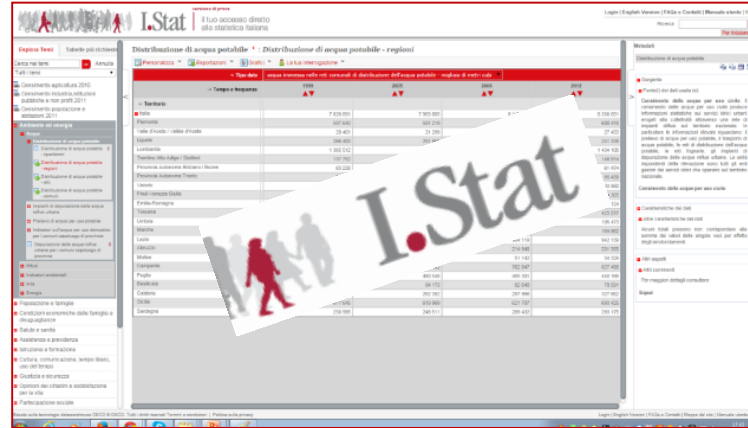
## Istat makes available through the AgID portal more than 700 different datasets



| Territorio                      | 1999      | 2005      | 2008      | 2012      |
|---------------------------------|-----------|-----------|-----------|-----------|
| Italia                          | 7 626 074 | 7 955 001 | 8 142 013 | 8 108 073 |
| Puglia                          | 387 892   | 387 276   | 388 086   | 389 070   |
| Vale d'Aosta - Valle d'Aoste    | 18 493    | 21 296    | 22 524    | 27 412    |
| Liguria                         | 298 431   | 291 891   | 298 193   | 297 598   |
| Lombardia                       | 1 162 022 | 1 438 882 | 1 457 075 | 1 424 423 |
| Basilicata                      | 137 982   | 168 138   | 169 816   | 169 852   |
| Piemonte                        | 42 239    | 62 311    | 62 768    | 61 414    |
| Piemonte Autonoma Valle d'Aosta | 74 514    | 75 889    | 74 942    | 75 470    |
| Veneto                          | 622 762   | 629 962   | 622 767   | 622 882   |
| Friuli-Venezia Giulia           | 159 036   | 159 188   | 159 188   | 159 327   |
| Emilia-Romagna                  | 458 024   | 482 882   | 471 812   | 459 228   |
| Toscana                         | 529 894   | 492 882   | 469 287   | 452 282   |
| Lazio                           | 65 094    | 65 281    | 65 441    | 65 415    |
| Marche                          | 188 281   | 188 281   | 188 281   | 188 282   |
| Umbria                          | 69 423    | 69 523    | 69 523    | 69 523    |
| Abruzzo                         | 198 182   | 202 182   | 202 182   | 202 182   |
| Molise                          | 47 082    | 47 082    | 47 082    | 47 082    |
| Campania                        | 292 882   | 292 882   | 292 882   | 292 882   |
| Puglia                          | 482 182   | 482 182   | 482 182   | 482 182   |
| Basilicata                      | 62 182    | 62 182    | 62 182    | 62 182    |
| Calabria                        | 112 182   | 112 182   | 112 182   | 112 182   |
| Sicilia                         | 612 182   | 612 182   | 612 182   | 612 182   |
| Sardegna                        | 242 182   | 242 182   | 242 182   | 242 182   |

<http://dati.istat.it>

# A First Step, but Not Enough





# Other Open Data Released by Istat

In addition to I.Stat, Istat currently publishes open data according to some main data models and formats

## Excel

Metadata described in a dedicated sheet

## CSV

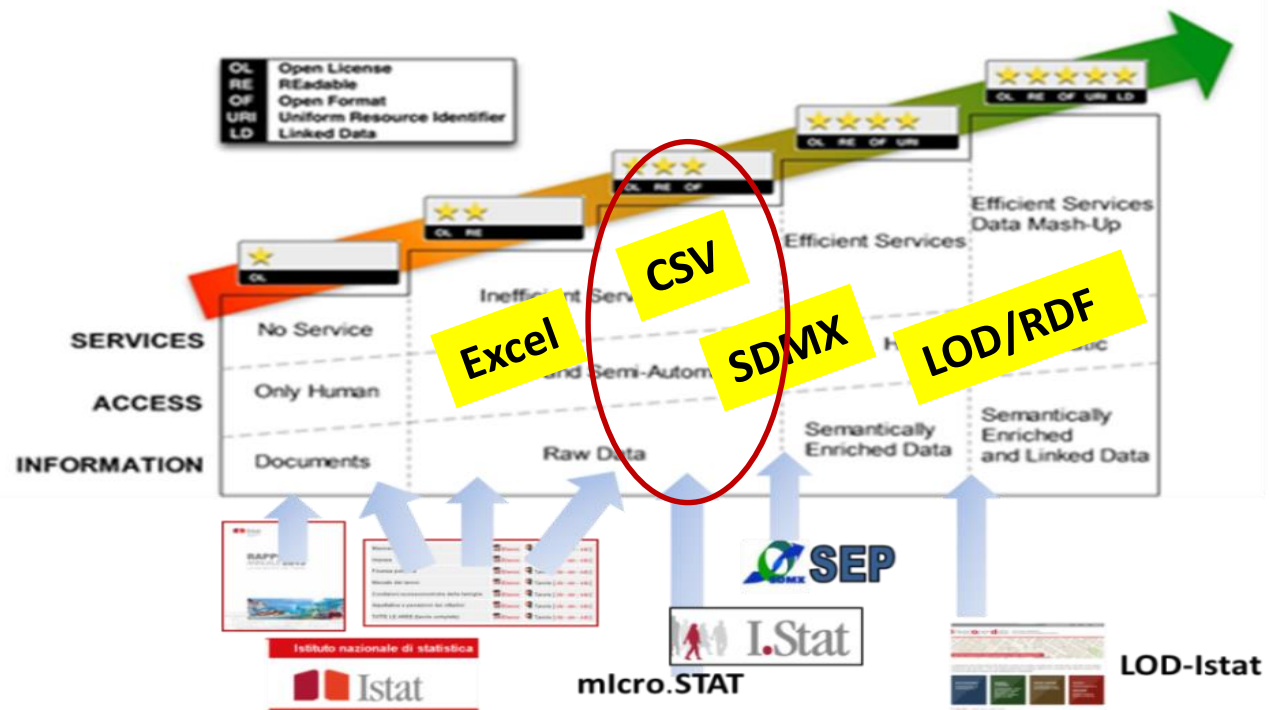
Mainly tabular, with a header that contains metadata

## SDMX

Description of data and metadata according to shared standards

## Linked Open Data/RDF

Data and metadata semantically described and linked to other (external) data



The National Institute of Statistics | Tips | About | Census Variables Webservice | Browse data | ITA | ENG

linkedopendata | PIATTAFORMA SPERIMENTALE PER I LOD DELL'ISTITUTO NAZIONALE DI STATISTICA

Dati del Censimento della popolazione e delle abitazioni 2011

The Linked Open Data (LOD) web site, allows to accesses and browse data in open format based on technology and Semantic Web standards. The LOD can be inquired directly from any application and responds to the need, expressed by the community of users, to have standardized and interoperable data.

**TERRITORIAL LEVEL**  
Wizard queries: territorial data year 2011

**CENSUS VARIABLES**  
Wizard queries:  
Aggregation by **territorial sub municipal partitions**  
Aggregation by **Labour Market Areas**

**SPARQL ENDPOINT**  
Custom queries: direct access to ENDPOINT SPARQL

**DOWNLOAD AREA**  
Territorial sub municipal partitions **dataset**  
Labour market areas **dataset**  
Territorial and census data **ontology**

linkedopendata | Contact us | Privacy | Legal notice | WebInfo

## Platform for

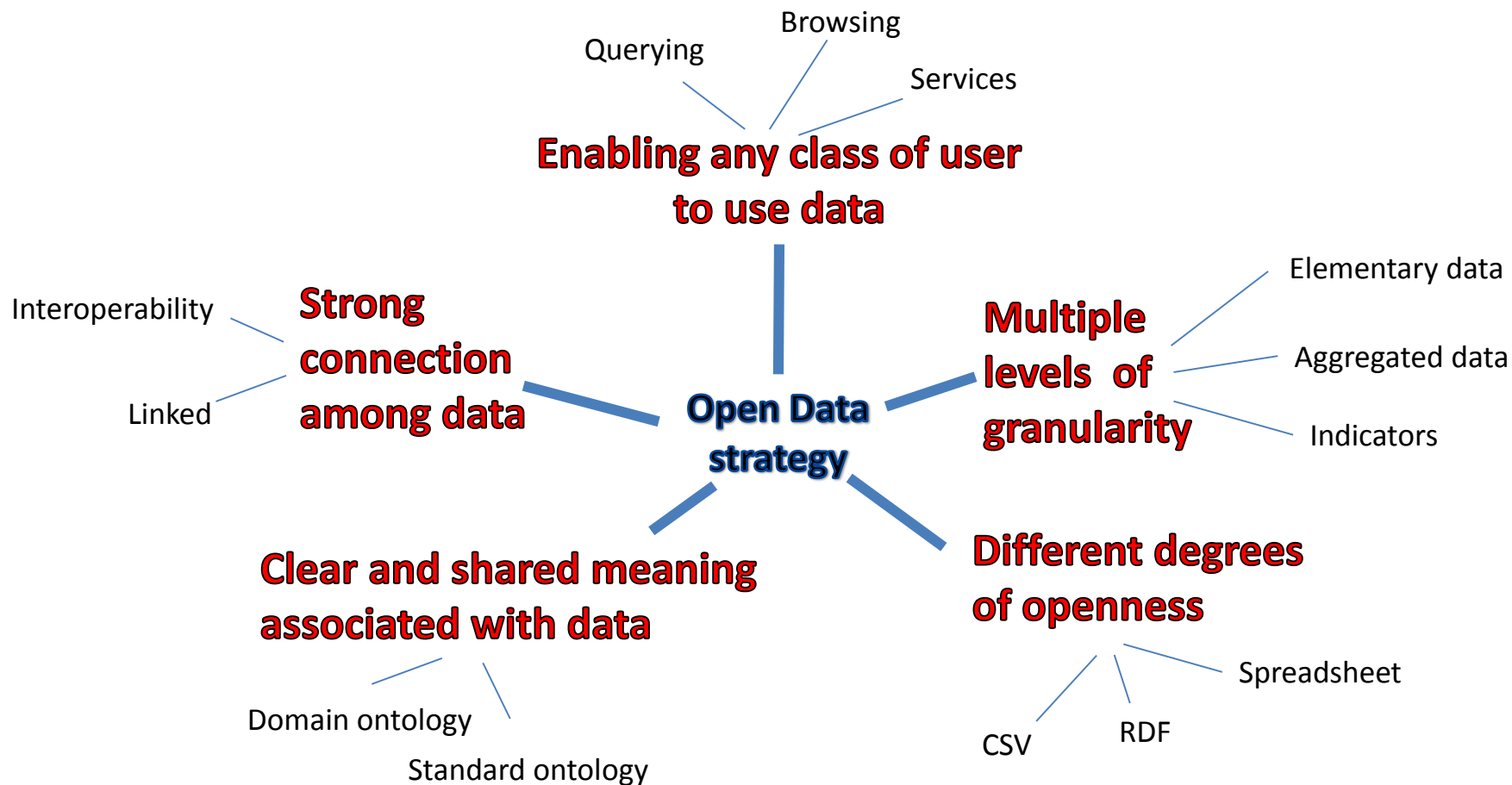
- Selecting
- Navigating
- Searching
- Querying
- Visualizing

## Open Data

## The platform allows

- Direct access to data via Web Services
- M2M solutions (e.g. GIS-LOD)
- Data conversion
- Export to productivity tools
- Visualization by means of external tools

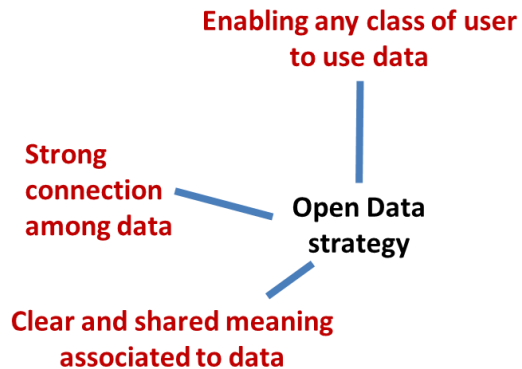
To increase the value of statistical data, some advanced features can be addressed through the LOD Portal



# Steps to a «perfect» data portal

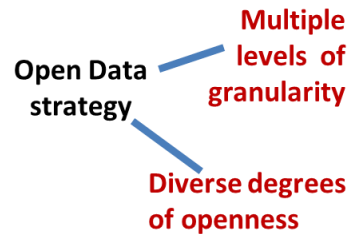
## STEP 1

**Give each class of users (human or not) the most appropriate way to use the data**



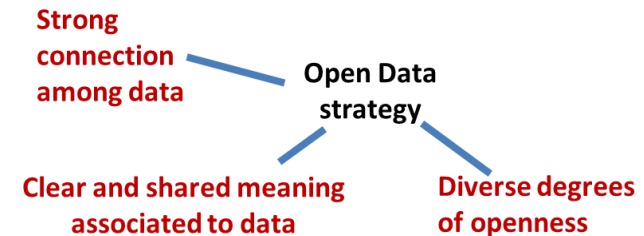
## STEP 2

**Make data in open format, whatever the level of openness**



## STEP 3

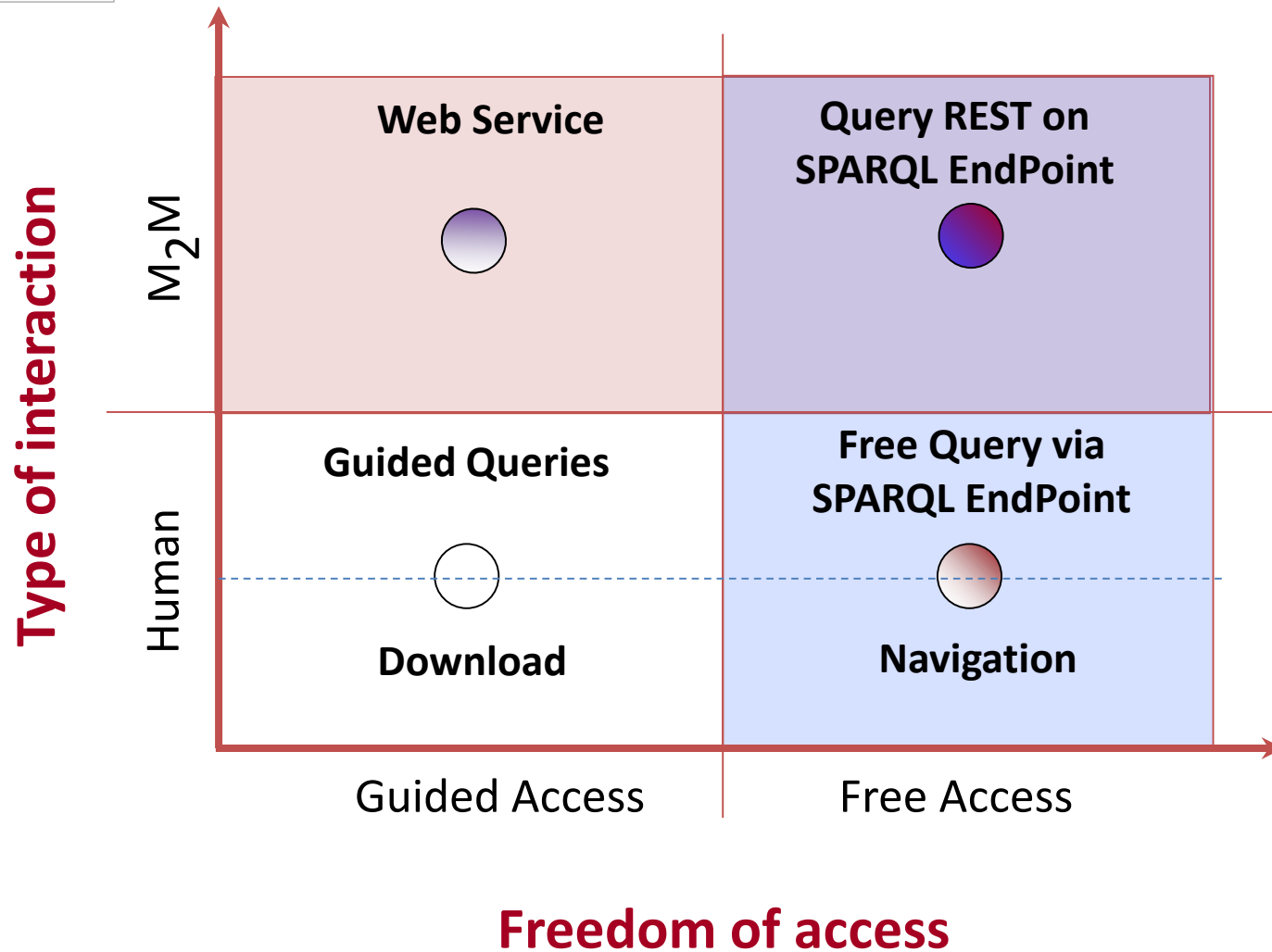
**Enrich the data with a semantic layer, regardless of the release on public web sites**



# Interaction Modes

STEP 1

Give each class of users the most appropriate way to use the data

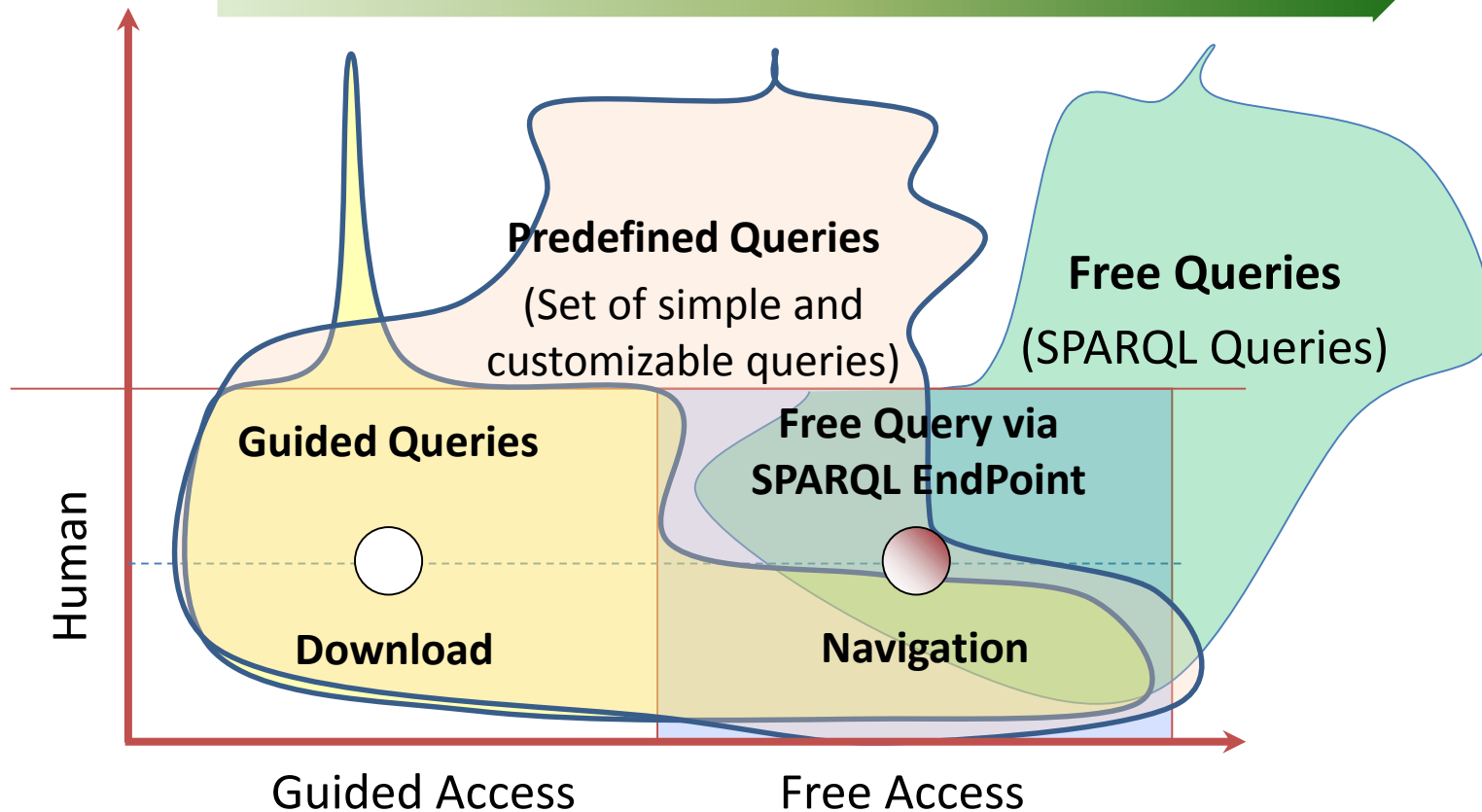


# Interaction Modes

(Human) User technical skills



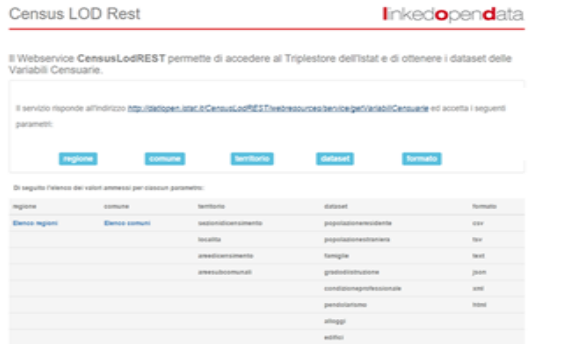
Type of interaction



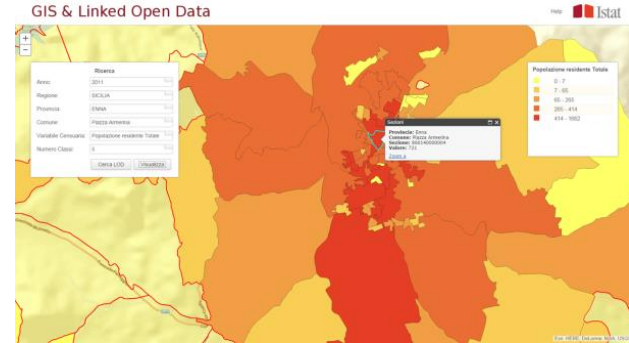
Freedom of access

# Possible uses of the portal

STEP 1 Give each class of users the most appropriate way to use the data



Interoperability



Integration

Graphical User Interface

BASI TERRITORIALI **VARIABILI CENSUARIE** SPARQL ENDPOINT DATASET ONTOLOGIE

VARIABILI CENSUARIE

Livello Territoriale  
Seleziona una provincia

Seleziona un comune

Livello Aggregazione Dati

Variabili Censuarie

Hai selezionato

Testo della query

Human-Computer-Interaction

Interactive browsing



## STEP 2 Make data in open format, whatever level of openness

**Datasets in traditional open formats (CSV, Spreadsheet, etc.) can be accessed and downloaded**



**Choice of Region**

- Piemonte
- Valle d'Aosta/Vallée d'Aoste
- Lombardia
- Trentino Alto Adige/Südtirol
- Veneto

**Choice of specific dataset**

- Sicilia
- Variabili censuarie per Area di censimento anno 2011
- Variabili censuarie per Area di censimento anno 2011
- Variabili censuarie per Sezione di censimento anno 2011

The screenshot shows a spreadsheet with columns labeled A through P22. The data includes codes for regions, provinces, and municipalities, along with various numerical indicators. The spreadsheet is titled 'Sicilia' and shows data for the year 2011.

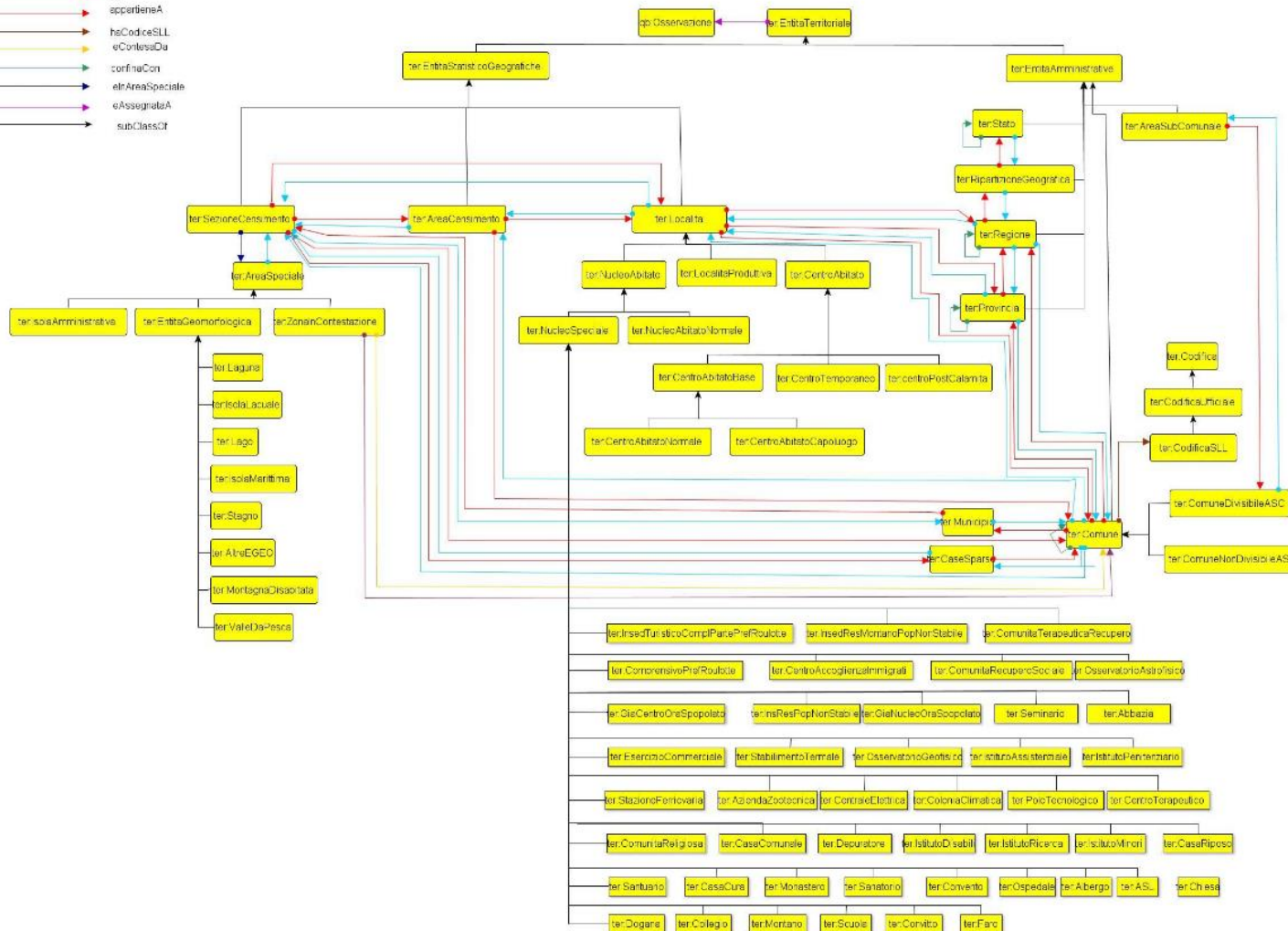
| A  | B      | C       | D      | E         | F      | G      | H      | I            | J     | K   | L       | M   | N   | O   | P   | Q   | R  | S  | T  | U   | V   | W   | X   | Y   | Z   | AA  | AB  | AC  | AD  | AE  | AF  | AG  | AM  |    |    |    |    |
|----|--------|---------|--------|-----------|--------|--------|--------|--------------|-------|-----|---------|-----|-----|-----|-----|-----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|----|----|----|
| 1  | CODREG | REGIONE | CODPRO | PROVINCIA | CODCOM | COMUNE | PROCOM | SEZ2011      | INSEZ | ACE | CODCASC | P1  | P2  | P3  | P4  | P5  | P6 | P7 | P8 | P9  | P10 | P11 | P12 | P13 | P14 | P15 | P16 | P17 | P18 | P19 | P20 | P21 | P22 |    |    |    |    |
| 2  | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000001 | 1     | 3   | 10001   | 149 | 74  | 75  | 53  | 81  | 4  | 10 | 1  | 32  | 40  | 2   | 0   | 0   | 3   | 2   | 6   | 5   | 13  | 11  | 7   | 9   | 6   |    |    |    |    |
| 3  | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000002 | 2     | 2   | 10001   | 221 | 121 | 100 | 95  | 114 | 2  | 10 | 0  | 62  | 56  | 2   | 1   | 0   | 10  | 15  | 21  | 20  | 17  | 8   | 11  | 19  | 17  |    |    |    |    |
| 4  | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000003 | 3     | 1   | 10001   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0  |    |    |    |
| 5  | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000004 | 4     | 1   | 10001   | 171 | 77  | 94  | 73  | 78  | 2  | 17 | 1  | 31  | 41  | 1   | 4   | 0   | 4   | 9   | 15  | 6   | 7   | 7   | 11  | 21  | 21  |    |    |    |    |
| 6  | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000005 | 5     | 1   | 10001   | 416 | 215 | 201 | 144 | 219 | 8  | 7  | 90 | 111 | 1   | 13  | 0   | 17  | 21  | 15  | 22  | 15  | 21  | 22  | 31  | 31  | 31  |    |    |    |    |
| 7  | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000006 | 6     | 1   | 10001   | 214 | 100 | 114 | 86  | 97  | 2  | 29 | 0  | 45  | 50  | 0   | 5   | 0   | 3   | 5   | 8   | 18  | 15  | 13  | 10  | 11  | 11  |    |    |    |    |
| 8  | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000007 | 7     | 1   | 10001   | 155 | 76  | 79  | 60  | 73  | 2  | 20 | 0  | 38  | 36  | 1   | 1   | 0   | 8   | 3   | 8   | 12  | 8   | 7   | 9   | 12  | 11  | 11 |    |    |    |
| 9  | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000008 | 8     | 1   | 10001   | 56  | 20  | 36  | 27  | 24  | 1  | 1  | 1  | 12  | 0   | 0   | 0   | 1   | 2   | 4   | 4   | 3   | 0   | 2   | 3   | 4   | 4   | 4  |    |    |    |
| 10 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000009 | 9     | 1   | 10001   | 146 | 66  | 80  | 57  | 67  | 1  | 17 | 4  | 27  | 34  | 1   | 2   | 4   | 3   | 5   | 7   | 7   | 9   | 13  | 4   | 8   | 8   | 8  |    |    |    |
| 11 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000010 | 10    | 1   | 10001   | 115 | 42  | 73  | 38  | 39  | 2  | 34 | 2  | 22  | 18  | 1   | 1   | 0   | 4   | 5   | 7   | 6   | 7   | 10  | 9   | 11  | 11  | 11 |    |    |    |
| 12 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000011 | 11    | 1   | 10001   | 103 | 47  | 56  | 50  | 50  | 1  | 22 | 23 | 1   | 0   | 1   | 2   | 5   | 6   | 0   | 3   | 4   | 5   | 5   | 10  | 10  | 10  | 10 |    |    |    |
| 13 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000012 | 12    | 1   | 10001   | 112 | 45  | 67  | 7   | 15  | 25 | 7  | 1  | 1   | 1   | 0   | 3   | 11  | 7   | 1   | 4   | 4   | 6   | 12  | 9   | 7   | 7   | 7  |    |    |    |
| 14 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000013 | 13    | 1   | 10001   | 113 | 47  | 77  | 77  | 77  | 77 | 77 | 77 | 77  | 77  | 77  | 77  | 77  | 77  | 77  | 77  | 77  | 77  | 77  | 77  | 77  | 77  | 77 | 77 |    |    |
| 15 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000014 | 14    | 1   | 10001   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  |    |
| 16 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000015 | 15    | 1   | 10001   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0  |
| 17 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000016 | 16    | 1   | 10001   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0  |
| 18 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000017 | 17    | 1   | 10001   | 369 | 177 | 192 | 146 | 175 | 5  | 36 | 3  | 80  | 88  | 1   | 10  | 19  | 14  | 20  | 31  | 24  | 20  | 27  | 24  | 24  | 24  | 24 | 24 | 24 |    |
| 19 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000018 | 18    | 1   | 10001   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0  |
| 20 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000019 | 19    | 1   | 10001   | 437 | 205 | 232 | 167 | 224 | 5  | 40 | 1  | 82  | 112 | 2   | 9   | 0   | 22  | 28  | 31  | 16  | 18  | 27  | 26  | 39  | 34  | 34 | 34 | 34 |    |
| 21 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000020 | 20    | 1   | 10001   | 368 | 165 | 203 | 131 | 188 | 4  | 40 | 5  | 65  | 94  | 3   | 1   | 2   | 11  | 19  | 18  | 15  | 18  | 19  | 20  | 26  | 29  | 29 | 29 | 29 |    |
| 22 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000021 | 21    | 1   | 10001   | 131 | 76  | 77  | 63  | 66  | 1  | 16 | 7  | 15  | 19  | 2   | 2   | 6   | 4   | 9   | 11  | 14  | 9   | 6   | 13  | 13  | 13  | 13 | 13 | 13 |    |
| 23 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000022 | 22    | 1   | 10001   | 14  | 68  | 75  | 56  | 72  | 2  | 11 | 2  | 27  | 38  | 1   | 0   | 1   | 0   | 3   | 7   | 7   | 5   | 10  | 10  | 10  | 10  | 10 | 10 | 10 |    |
| 24 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000023 | 23    | 1   | 10001   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0  |
| 25 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000024 | 24    | 2   | 10001   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0  |
| 26 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000025 | 25    | 2   | 10001   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0  |
| 27 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000026 | 26    | 1   | 10001   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0  |
| 28 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000027 | 27    | 1   | 10001   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0  |
| 29 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000028 | 28    | 2   | 10001   | 312 | 158 | 154 | 133 | 158 | 4  | 14 | 3  | 73  | 79  | 3   | 1   | 2   | 22  | 22  | 25  | 21  | 17  | 14  | 19  | 35  | 35  | 35 | 35 | 35 |    |
| 30 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000029 | 29    | 2   | 10001   | 199 | 103 | 96  | 76  | 111 | 0  | 9  | 3  | 43  | 56  | 0   | 3   | 1   | 9   | 13  | 18  | 18  | 17  | 20  | 11  | 20  | 11  | 20 | 11 | 20 | 11 |
| 31 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000030 | 30    | 2   | 10001   | 279 | 139 | 140 | 108 | 137 | 1  | 11 | 0  | 59  | 79  | 1   | 0   | 0   | 22  | 14  | 18  | 17  | 12  | 19  | 22  | 36  | 17  | 22 | 36 | 17 | 22 |
| 32 | 19     | Sicilia | 81     | Trapani   | 1      | Aicamo | 81001  | 810010000031 | 31    | 2   | 10001   | 117 | 56  | 61  | 48  | 63  | 0  | 3  | 24 | 32  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0  |

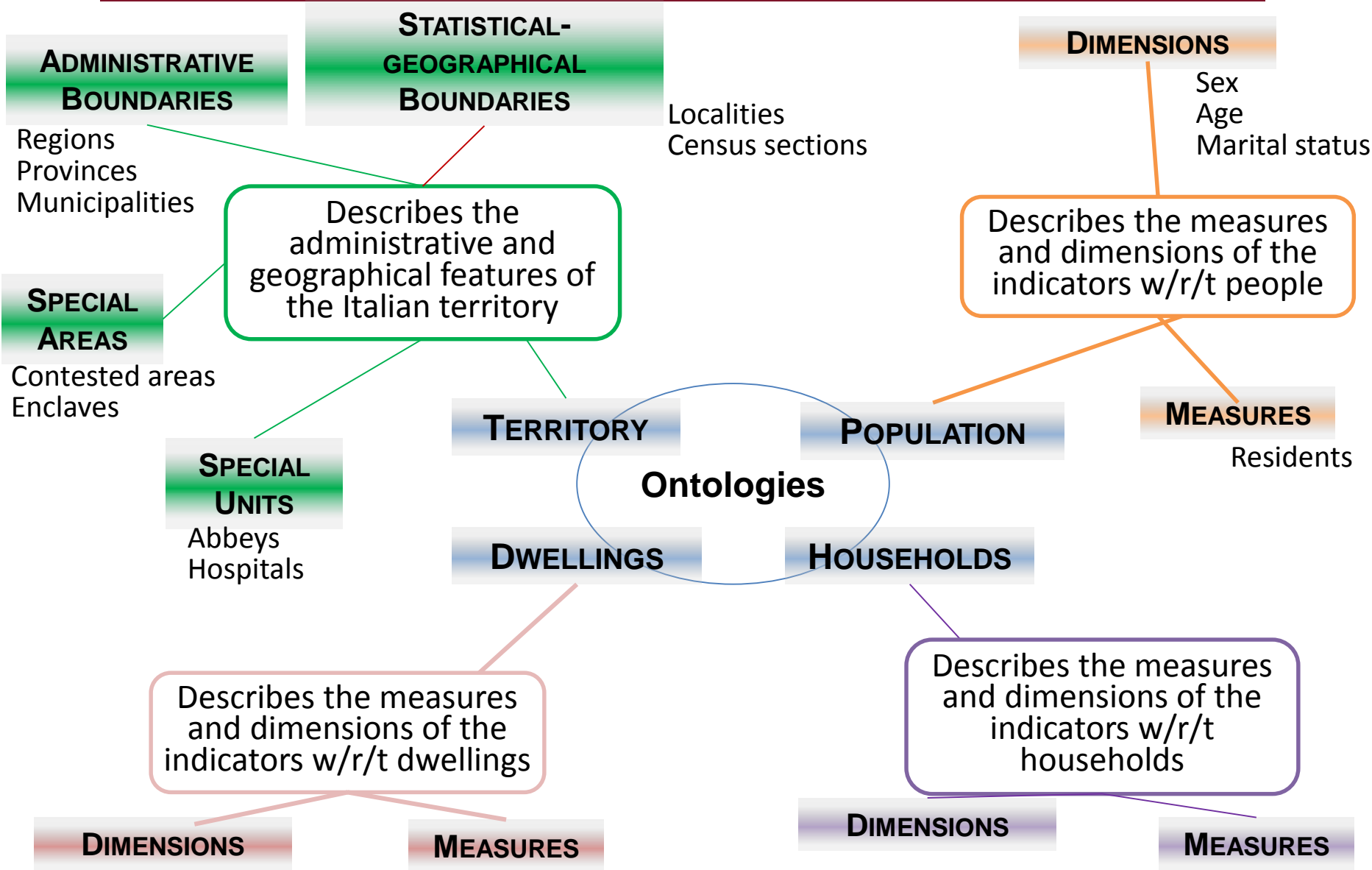


# STEP 3

## Enrich the data with a semantic layer

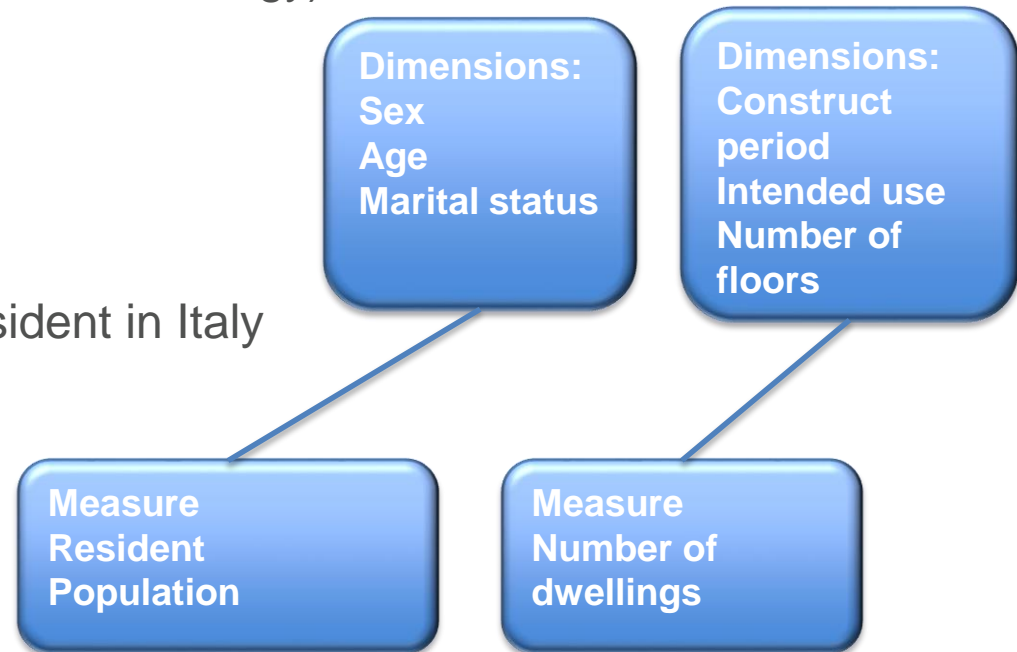
- ha appartenenza
- haCodiceSLL
- eContieneDa
- confinaCon
- eInAreaSpeciale
- eAssegnataA
- subClassOf





# Census Data Ontology

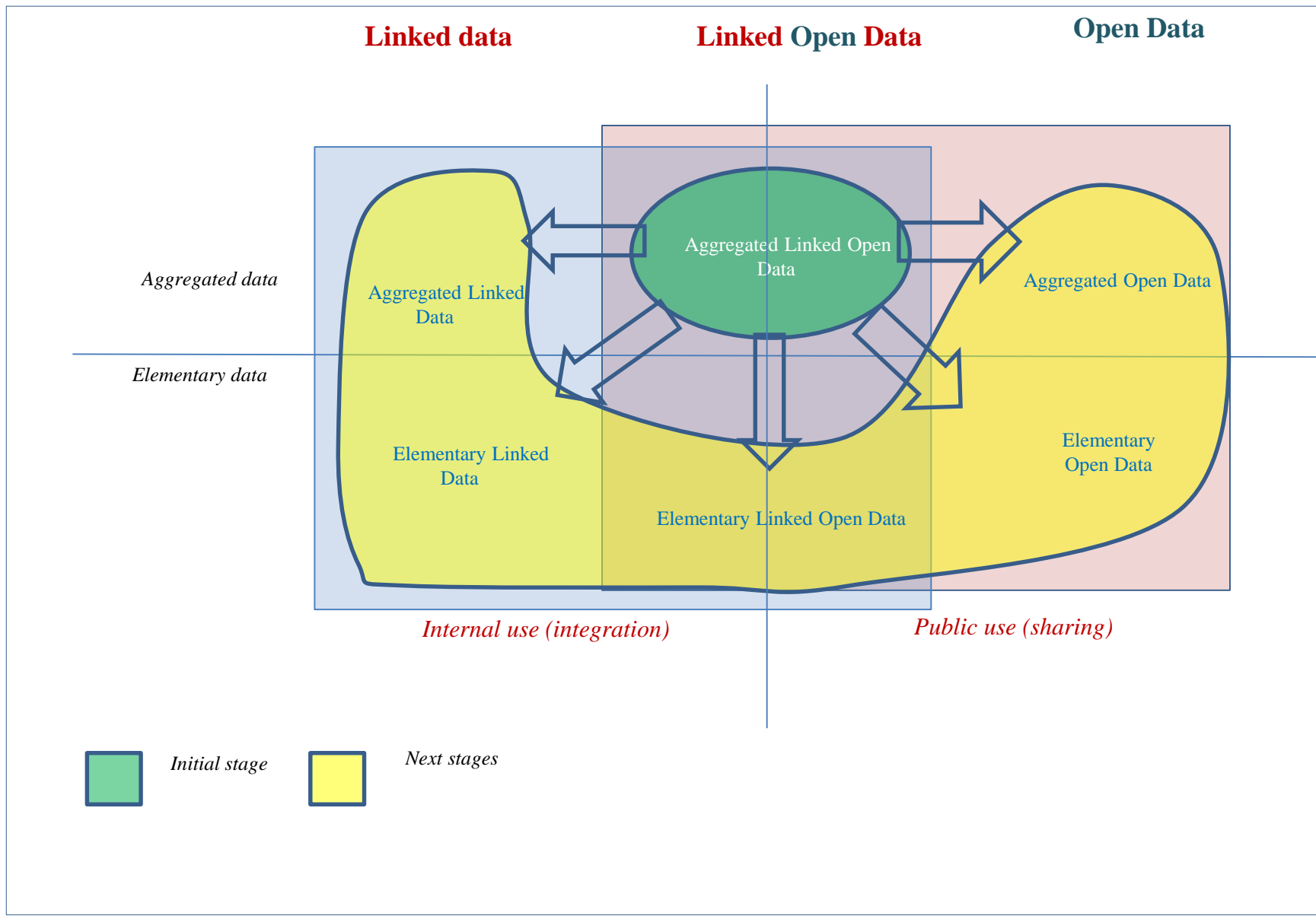
- Use of RDF Data Cube Vocabulary that allows to publish multi-dimensional data
- Dimensions:
  - Sex
  - Age class
  - Citizenship
  - Territory (defined in the Territorial Ontology)
  - Construction Period
  - Number of floors
  - .....
- Measures:
  - Number of residents
  - Foreigners and stateless resident in Italy
  - Number of Housing
  - .....



The future plans for the platform are targeted to the implementation of an integrated environment of data, including:

- Open data for public uses (also non-linked)
- Linked data for internal integrated uses (through the use of ad-hoc designed ontologies)

Elementary and aggregate Linked Open Data for internal and public use, available for human and machine users, related to the most part of statistics domain and possibly enriched by geographical representation



- Macroscale vs microscale modeling
  - Pseudo-Einstein (as simple as possible but not simpler)
  - Von Neumann (agent-based modeling)
- Technological constraint → enabling technology
- It widens the space of what is feasible:
  - In production: our experience with SBS.Frame
  - In analysis and research...
- A paradigm shift?
  - Statistical mechanics vs Agent-based modeling
  - Just because you can doesn't mean you should
- Back to open data
  - From dissemination to release (“data liberation” at StatCan)
  - The regulators have need to new rules in line with the new scenarios