

Impegno D8 - Campagna permanente di comunicazione/sensibilizzazione organizzata dalla Community OGP Italia

“Sviluppo collaborativo di uno schema di dati aperti sull’imprenditoria femminile e giovanile in Italia al fine di contribuire a creare un modello dati rappresentativo ed equo sul fenomeno delle disuguaglianze generazionali e di genere”

Indice

1. Finalità	3
2. Schema dati e loro fonti	4
3. Formato dataset e metadati	5
4. Accesso ai dataset, licenza e data visualization	10
5. Riferimenti normativi e linee guida	11

1. Finalità

Il presente documento ha come scopo quello di fornire un'ipotesi di ontologia dati per realizzare dei dataset riferiti alle imprese femminili e giovanili: dataset in formato aperto che siano machine readable, disponibili liberamente e quindi riusabili, secondo il presupposto di interoperabilità delle basi dati della PA, e rispondenti il più possibile ai principi del Governo Aperto.

L'obiettivo, all'interno del team di lavoro *"Promuovere la consapevolezza su disuguaglianza di genere e intergenerazionale in Italia - Campagna permanente di comunicazione/sensibilizzazione organizzata dalla Community OGP Italia"* obiettivo *"D Promuovere la consapevolezza su disuguaglianza di genere e intergenerazionale in Italia"* del VI NAP, è quello di inquadrare in modo equo e rappresentativo il fenomeno delle disuguaglianze generazionali e di genere per orientare meglio servizi ed iniziative in tal senso sia da parte della PA (eventualmente anche delle OSC) attraverso, anche, la l'eventuale creazione di un catalogo permanente e aggiornato nel tempo con i dati aperti sulle imprese femminili e giovanili. In questo modo alla comunità verrebbe restituito un patrimonio informativo utile anche a moltiplicare le opportunità di studio e analisi del fenomeno.

Nella progettazione si è tenuto conto, oltre che delle regole e delle prassi rispetto la creazione e condivisione dei dati aperti, anche delle *"Linee Guida recanti regole tecniche per l'apertura dei dati e il riutilizzo dell'informazione del settore pubblico"* 2 pubblicate dall'AGID ad agosto 2023.

In questa prima stesura abbiamo individuato come fonte di approvvigionamento dei dati la banca dati del Registro Imprese appartenente alle Camere di commercio, ma si può considerare anche il mashup con altri dataset provenienti da più fonti del Catalogo Nazionale dei Dati Aperti (dati.gov.it) e dei partecipanti alla Community OGP. Per questo motivo l'ipotesi di ontologia presentata in questo documento è aperta alle eventuali proposte di integrazione da parte delle OSC e di altre istituzioni/gruppi di interesse presenti nel team di lavoro e nella Community.

2. Schema dati e loro fonti

1. [https://open.gov.it/governo-aperto/piano-nazionale/6 nap](https://open.gov.it/governo-aperto/piano-nazionale/6%20nap)
2. <https://www.agid.gov.it/it/agenzia/stampa-e-comunicazione/notizie/2023/08/03/open-data-agid-adotta-linee-guida>

Ad oggi la principale fonte dati sull'imprenditoria femminile e giovanile¹ è rappresentata dal Registro Imprese. Le informazioni acquisibili da Registro Imprese sono principalmente di 2 tipi: **dati di stock** (numero di imprese con determinate caratteristiche riferite a un determinato periodo e territorio) ed **elenchi di imprese** (questi ultimi non sono disponibili gratuitamente e in quanto elenchi richiedono una ulteriore elaborazione per ottenere dei dati riferiti a variabili quale la forma giuridica d'impresa, l'area geografica, il settore, ecc.).

I **dati di stock** sono reperibili anche sotto forma di dataset in formato aperto e liberi di essere riutilizzati, in particolar modo per le imprese femminili, grazie ad alcune Camere di Commercio che da anni pubblicano open data. Questi dataset sono scaricabili liberamente e gratuitamente e possono essere mixati con altri dataset per arricchirne gli spunti interpretativi:

- Imprese Femminili Attive in Italia per Territorio, Settore Ateco e Tempo (frequenza trimestrale)
- Imprese Femminili Attive in Italia per Territorio, Sottocategoria Ateco (6 cifre ATECO) e Tempo (frequenza trimestrale)

Per quanto riguarda le imprese giovanili, al momento non è nota la disponibilità di dataset in formato aperto riferiti al territorio "Italia" e con le stesse caratteristiche sopra indicate per le imprese femminili. Resta la possibilità di richiedere appositamente l'estrazione di questi elenchi (potrebbe essere previsto un costo):

- Imprese Giovanili Attive in Italia per Territorio, Settore Ateco e Tempo (frequenza trimestrale)
- Imprese Giovanili Attive in Italia per Territorio, Sottocategoria Ateco (6 cifre ATECO) e Tempo (frequenza trimestrale)

¹ Si definiscono "Imprese femminili" le imprese la cui partecipazione di genere risulta complessivamente superiore al 50% mediando le composizioni di quote di partecipazione e cariche attribuite. Per le Società di persone e cooperative: si definisce impresa femminile quella con oltre il 50% di "Soci" donne. Per le Ditte individuali: si definisce impresa femminile quella con il titolare donna. Per le altre forme giuridiche: si definisce impresa femminile quella con oltre il 50% di "Amministratori" donna. Si definiscono "Imprese giovanili" le imprese la cui partecipazione del controllo e della proprietà è detenuta in prevalenza da persone di età inferiore ai 35 anni.

3. Formato dataset e metadati

Secondo il profilo di metadatazione (DCAT-AP_IT) i dataset devono avere una serie di metadati predefiniti, con determinate caratteristiche e codifiche. Alcuni campi sono obbligatori, altri raccomandati per favorire l'interoperabilità.

Nella tabella seguente è indicata la successione dei metadati che devono corredare ogni dataset, con evidenza dello status di obbligatorietà e un primo riferimento (campo "note") alla loro codifica:

Elementi	Descrizione	Status	Note
Identificativo del dataset	L'identificativo univoco del Dataset assegnato dal catalogo nazionale	Obbligatorio	L'identificativo univoco del dataset è assegnato dal catalogo nazionale ed è costituito da <i>codiceIPA:D.numeroProgressivo. Esempio—> agid:D.1</i>
Titolo del dataset	Il titolo del Dataset	Obbligatorio	dct:title
Descrizione del dataset	La descrizione del Dataset che indica cosa contiene il dataset.	Obbligatorio	Si raccomanda l'utilizzo di un linguaggio semplice che possa aiutare qualsiasi utente a identificare il contenuto del dataset (dct:description)
Editore	Editore dataset. La proprietà lega l'oggetto (dominio) ad un oggetto di tipo Agent (un soggetto/organizzazione che gioca un certo ruolo sui dataset) tramite URI (specificato mediante un URI - Uniform Resource Identifier)	Obbligatorio	Un'organizzazione (o pubblica amministrazione) responsabile di rendere disponibile (pubblicare) il catalogo. Si raccomanda di evitare l'inserimento di nomi di singole persone.
Titolare- Codice Ipa/P. Iva	Titolare codice Ipa. Nel caso di Pubbliche Amministrazioni l'identificativo è rappresentato dal codice IPA (Indice della Pubblica Amministrazione). Per organizzazioni private, l'identificativo è rappresentato dalla Partita IVA.	Obbligatorio	Specificare il nome ufficiale della pubblica amministrazione così come riportato nell'Indice della Pubblica Amministrazione. Nel caso di organizzazione privata, specificare il nome ufficiale della stessa così come riportato nel Registro Imprese.
Autore	Il creatore del Dataset. La proprietà lega l'oggetto (dominio) `Dataset` <dataset.html> a uno o più oggetti	Obbligatorio	Un'organizzazione (o pubblica amministrazione) che ha materialmente prodotto il dataset. Si raccomanda di evitare l'inserimento di nomi di singole persone.

Elementi	Descrizione	Status	Note
	(codominio) di tipo Agent (Soggetto) (specificato mediante un URI - Uniform Resource Identifier)		
Ultima modifica del dataset	La data di ultima modifica del Dataset	Obbligatorio	Si raccomanda di usare il formato ISO 8601, i.e., yyyy-mm-dd (dct:modified)
Titolare del dataset	La proprietà lega l'oggetto (dominio) Dataset a un oggetto (codominio) di tipo Agent (Soggetto) (specificato mediante un URI - Uniform Resource Identifier)	Obbligatorio	Il metadato assume come valore un URI (NON una stringa). Esso rappresenta un'organizzazione (o pubblica amministrazione) responsabile della gestione complessiva del dataset in virtù dei propri compiti istituzionali. dct:rightsHolder
Frequenza aggiornamento dataset	La frequenza di aggiornamento del Dataset. La proprietà lega l'oggetto (dominio) Dataset a un oggetto (codominio) di tipo dct:Frequency (specificato mediante un URI - Uniform Resource Identifier)	Obbligatorio	La frequenza con cui il dataset viene aggiornato assume necessariamente uno dei valori definiti nel vocabolario Europeo sulle frequenze. Nel caso la frequenza di aggiornamento non sia disponibile è possibile indicare una frequenza sconosciuta utilizzando il seguente URI: http://publications.europa.eu/resource/authority/frequency/UNKNOWN
Distribuzione del dataset	La distribuzione del Dataset. La proprietà lega l'oggetto (dominio) Dataset a un oggetto (codominio) di tipo Distribuzione (specificato mediante un URI - Uniform Resource Identifier)	Obbligatorio nel caso di dati aperti, Opzionale negli altri casi	Una distribuzione è una forma attraverso cui il dataset è disponibile. Ogni dataset può essere disponibile in diverse forme, come per esempio diversi formati o differenti endpoint (e.g., SPARQL endpoint). <i>dcat:distribution</i>
Temi del dataset	I temi attraverso cui classificare il Dataset. La proprietà lega l'oggetto (dominio) Dataset a uno o più oggetti (codominio) di tipo skos:Concept (specificato mediante un URI - Uniform Resource	Obbligatorio	Il metadato assume come valore un URI (NON una stringa con l'URL del tema) che deve essere necessariamente uno tra quelli definiti nel vocabolario Europeo sui Temi per i dati : https://op.europa.eu/en/web/eu-vocabularies .

Elementi	Descrizione	Status	Note
	Identifier)		
Licenza	La licenza della distribuzione del dataset.	Obbligatorio	La licenza è definita mediante la specifica della classe <i>dcatapit:LicenseDocument</i> identificata univocamente da un URI (Uniform Resource Identifier). Consigliata la licenza CC-BY 4.0*
Sottotema dataset	La sottocategoria in cui può essere classificato il Dataset. La proprietà lega l'oggetto (dominio) Dataset a uno o più oggetti (codominio) di tipo skos:Concept (specificato mediante un URI - Uniform Resource Identifier)	Raccomandato	In corrispondenza del tema scelto, si possono specificare più categorie per indicare parole chiave specifiche con cui il dataset può essere ulteriormente classificato.
Data ultimo aggiornamento	data di ultima modifica del Dataset		Si raccomanda di usare il formato ISO 8601, i.e., yyyy-mm-dd
Data di rilascio dataset	data di rilascio del dataset	Obbligatorio	E" la data in cui il dataset è reso disponibile.
Lingua del dataset	La lingua del Dataset. La proprietà lega l'oggetto (dominio) Dataset a uno o più oggetti (codominio) di tipo dct:LinguisticSystem (specificato mediante un URI - Uniform Resource Identifier)	Opzionale	Indicare una o più lingue utilizzate nel catalogo. La scelta della lingua è regolata dall'uso obbligatorio del vocabolario definito a livello Europeo sulle lingue. https://op.europa.eu/en/web/eu-vocabularies
Copertura geografica	La copertura geografica del dataset.	Opzionale	definita mediante la specifica della classe Location (Localizzazione) identificata univocamente da un URI
Estensione Temporale	La copertura (o estensione) temporale del dataset è definita mediante la specifica della classe PeriodOfTime (Periodo di Tempo) identificata univocamente da un URI (Uniform Resource Identifier).	Obbligatorio	definisce il periodo di tempo ovvero una data di inizio dcatapit:startDate e una data di fine dcatapit: endDate. Qualora si indichi una copertura temporale per il dataset è obbligatorio specificare almeno la data di inizio. Nel caso di periodi temporali ancora aperti alla data corrente, si può evitare di specificare la data di fine.

L'adozione, qualora possibile e/o previsto, di una licenza diversa dalla CC-BY 4.0 o CC0 DEVE essere formalmente motivata, anche alla luce dei principi espressi dalla Direttiva, salvo che sia stata adottata una licenza altrettanto compatibile come la CDLA 2.0 permissive ovvero qualsiasi altra licenza aperta equivalente o meno restrittiva, che consenta il riutilizzo salvo obbligo di attribuzione, dando credito al concedente. Per le serie di dati di elevato valore vale quanto indicato nel Regolamento (UE) di esecuzione n. 2023/138. (Art. 8, cc. 1 e 2) SI RACCOMANDA di non utilizzare le licenze Creative Commons precedenti alla 4.0, in cui i diritti sui generis non erano citati/previsti (2.5), o erano richiamati come. meramente rinunciati (3.0).

Dati relativi alle imprese e alla proprietà delle imprese:

Le serie di dati sono messe a disposizione per il riutilizzo:

- [senza indebito ritardo dopo l'ultimo aggiornamento];
- alle condizioni della licenza Creative Commons BY 4.0 o di una licenza aperta equivalente o meno restrittiva, con condizioni supplementari relative al riutilizzo dei dati personali, se del caso;
- in un formato aperto leggibile meccanicamente e riconosciuto dall'Unione o a livello internazionale (XHTML per i documenti che rientrano nell'ambito di applicazione del Regolamento Delegato (UE) 2018/815 della Commissione; altri formati, se e ove prescritto dal diritto dell'Unione applicabile), con metadati completi ([per i documenti che rientrano nell'ambito di applicazione del Regolamento Delegato (UE) 2016/1437 della Commissione, i metadati specificati in tale regolamento, se del caso;] per gli altri documenti, i metadati prescritti dal diritto dell'UE applicabile, se del caso); la leggibilità meccanica non è imposta per i dati conservati in formati non leggibili meccanicamente (ad es. documenti e campi di dati non strutturati/non leggibili meccanicamente inclusi come parte di documenti leggibili meccanicamente);
- tramite API e download in blocco;
- [a livello di singola azienda]

Linee Guida Open Data https://www.agid.gov.it/sites/default/files/repository_files/ig-open-data_v.1.0_1.pdf

POSSIBILE ESEMPIO DI APPLICAZIONE DEI METADATI AD UN DATASET

Campo	Valore
Titolo del dataset	Imprese Femminili Attive in Italia per Settore Ateco
Descrizione del dataset	Imprese Femminili Attive in Italia per Settore di Attività Economica (Ateco 2007). Sono definite "femminili" le imprese nelle quali la partecipazione femminile sia superiore al 50%, con riferimento alla natura giuridica, all'eventuale quota di capitale sociale detenuta da ciascun socio donna e alla percentuale di donne presenti tra gli amministratori, titolari o soci dell'impresa. Nel 2014 è stata introdotta una modifica all'algoritmo utilizzato per la determinazione del grado di partecipazione femminile delle società di persone. Ciò determina l'impossibilità di comparare la serie dei dati relativi a periodi successivi al 1° gennaio 2014, con quelli precedenti.
Editore	Da definire in base alla PA o OSC che decide di pubblicare il dataset
Ultima modifica del dataset	GG/MM/AAAA
Titolare del dataset	Da definire in base alla PA o OSC che decide di pubblicare il dataset, può essere differente dal valore "Editore"
Frequenza aggiornamento dataset	Trimestrale
Distribuzione del dataset	Allegato o allegati in formato aperto
Temi del	Economia e Finanze

Campo	Valore
dataset	
Licenza	CC BY
Sottotema dataset	Analisi economica, Struttura economica
Ultima modifica Dataset	GG/MM/AAAA
Data di rilascio dataset	GG/MM/AAAA
Lingua del dataset	Italiano
Copertura geografica	Italia
Estensione Temporale	da/a

4. Accesso ai dataset, licenza e data visualization

In coerenza con le “Linee Guida recanti regole tecniche per l’apertura dei dati e il riutilizzo dell’informazione del settore pubblico” i dataset devono essere resi disponibili secondo le seguenti caratteristiche:

- essere leggibili meccanicamente (almeno 3 stelle nella classificazione “Data.europa.eu - Data Quality Guidelines” cfr. tabella che segue)
- il formato dei file contenenti i dati deve essere aperto ovvero tale da consentire la consultazione senza l’utilizzo di software protetti da licenze
- ogni dataset deve essere provvisto dei relativi metadati
- il catalogo e i relativi dataset devono essere disponibili gratuitamente
- la licenza da applicare ai dataset deve essere di tipo “standard” ovvero senza vincoli di riutilizzo, anche commerciale, dei contenuti (sugerita CC BY 4.0)
- il catalogo e i dataset devono essere raggiungibili e consultabili attraverso le tecnologie dell’informazione e della comunicazione (quindi disponibili online)

Una modalità iniziale di offerta dei dataset potrebbe essere il formato .CSV affiancato, a tendere, dal formato RDF.

Data visualization

Nella pubblicazione dei dataset è possibile proporre agli utenti i file da scaricare con annessa tabella/schema dei metadati e le principali informazioni sui dataset, la loro provenienza, la loro finalità, in modo da avere un quadro del riutilizzo dei dati stessi.

5. Riferimenti normativi e linee guida

- Per la composizione del catalogo e dei dataset e i metadati di questi ultimi cfr. DCAT-AP_IT ([Linee guida per i cataloghi dati | 1. Catalogo](#))
- Per le modalità di pubblicazione e condivisione dei dati cfr. Linee Guida recanti regole tecniche per l'apertura dei dati e il riutilizzo dell'informazione del settore pubblico ([Open Data: AgID adotta le Linee Guida | Agenzia per l'Italia digitale](#))
- Per la classificazione ATECO, dal 1° Aprile 2022 è operativa la nuova classificazione ISTAT delle attività economiche per accogliere le evoluzioni del tessuto imprenditoriale e i cambiamenti sociali intervenuti. ([Classificazione delle attività economiche Ateco versione 2007, aggiornamento 2022](#))