



*Sistema Pubblico di
Connettività e Cooperazione*

COMMISSIONE DI COORDINAMENTO SPC

MATERIALE RACCOLTO SU ALCUNE ESPERIENZE ITALIANE DI OPEN DATA

(Appendice alle Linee guida per l'interoperabilità semantica attraverso i Linked Open Data)



Agenzia per l'Italia Digitale

INDICE

1. PREFAZIONE.....	3
2. ESPERIENZA CNR.....	4
3. ESPERIENZA COMUNE DI FIRENZE.....	5
4. ESPERIENZA INPS	9
5. ESPERIENZA MIUR.....	19
6. ESPERIENZA COMUNE DI NAPOLI	21
7. ESPERIENZA COMUNE DI PADOVA.....	22
8. ESPERIENZA TRENTO: IL GEOCATALOGO SEMANTICO	26



1. PRAFAZIONE

Questa appendice raccoglie il materiale fornito da alcuni partecipanti del Gruppo di Lavoro della Commissione di Coordinamento SPC che si è occupato, nell'ambito della definizione delle infrastrutture condivise, di definire le "linee guida per l'interoperabilità semantica attraverso i Linked Open Data". Il materiale descrive parti delle esperienze Open Data intraprese da tali partecipanti.



2. ESPERIENZA CNR

data.cnr.it è un portale realizzato secondo le best practices del W3C per i linked data e permette livelli avanzati di interazione mediante modalità diverse, che includono anche piattaforme di ricerca esplorativa e interazione innovative a livello internazionale, come il Semantic Scout.

I dati del CNR sono stati estratti dai database CNR con tecniche automatiche e riproducibili, in funzione dei casi d'uso richiesti, realizzando un'ontologia del CNR in OWL e RDF (gli standard W3C per il Web Semantico), che esprime chiaramente i tipi di entità e le loro relazioni, ed è allineata alle ontologie più usate per i linked data. Ulteriori dati sono stati estratti dai testi presenti nei dati CNR con tecniche di trattamento del linguaggio naturale, per esempio titoli e abstract di articoli, programmi di ricerca di progetti e istituti, etc. Infine, nuovi dati sono stati ottenuti applicando tecniche di inferenza automatica. Il risultato è quindi un arricchimento dei dati disponibili da database, non solo la loro disponibilità in formati interoperabili come OWL e RDF.

I dati sono accessibili mediante uno SPARQL endpoint per l'interrogazione diretta e per l'accesso programmatico, ma anche mediante un'interfaccia Web e soprattutto mediante il Semantic Scout.

Il Semantic Scout permette una ricerca a parole chiave che rintraccia non i documenti, ma le entità del CNR: ricercatori, istituti, progetti, etc. A partire da queste entità, si accede poi a una navigazione esplorativa con tecniche grafiche innovative e personalizzabili, che permettono anche la creazione di report di viaggio che riassumono tutte le informazioni incontrate e selezionate durante l'esplorazione dell'utente.

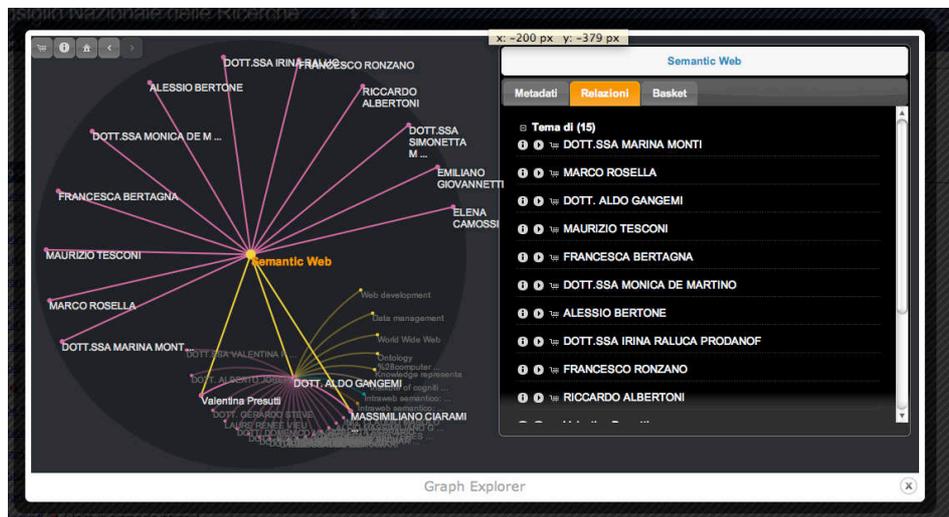


Figura 1: Esempio di navigazione dei dati offerta da data.cnr.it

3. ESPERIENZA COMUNE DI FIRENZE

Gli aspetti organizzativi del processo di pubblicazione dei dati sono persino prevalenti su quelli tecnologici relativi alla individuazione di una piattaforma di catalogazione, alla sua personalizzazione ed alla sua alimentazione.

E' infatti indispensabile creare all'interno dell'Ente gestore, in modo pervasivo rispetto a tutta la sua struttura organizzativa, una diffusa cultura sugli scopi del progetto di pubblicazione dei dati della Pubblica Amministrazione e sulla necessità, perché il progetto abbia successo, di assoluta condivisione trasversale dei dati visti non più come supporto alle quotidiane attività lavorative delle singole strutture ma come veicolo essenziale a produrre livelli informativi integrati.

Ogni Ente elabora i propri dati utilizzando applicativi dedicati alle funzionalità in carico alle singole strutture. Senza una adeguata architettura di gestione informatizzata delle attività dell'Ente è velleitario qualsiasi progetto di raccolta e pubblicazione di dati che risulterebbero avulsi da contesti strutturati. Nasce quindi l'esigenza di una infrastruttura centrale dedicata all'estrazione dei dati soggetti ad una gestione decentrata ed alla loro raccolta in un data warehouse classicamente strutturato in dimensioni, le entità principali, ed atti, riferibili alle stesse dimensioni, gestiti in maniera storica.

Integrazione e federazione di dati e servizi richiedono competenze informatiche di alto livello come quelle che normalmente si trovano all'interno delle strutture che hanno in carico il sistema informativo dell'Ente. E' quindi opportuno che all'interno di queste strutture siano progettate e realizzate le infrastrutture tecnologiche in grado di monitorare gli aggiornamenti, bonificare i dati riferibili al patrimonio informativo dell'Ente Pubblico ed integrarli a livello centrale.

Una volta realizzate tali infrastrutture devono essere poste in atto campagne di bonifica che riconducano ad un modello comune le relazioni tra le diverse entità in modo da poter realizzare viste integrate delle stesse.

Indispensabile al raggiungimento del necessario livello di qualità dei dati è l'organizzazione di una rete trasversale di referenti tecnici in grado di monitorare i processi di bonifica di ambito periferico. Tale attività deve essere pianificata a livello di Ente con la predisposizione e condivisione di piani dettagliati di obiettivi per ogni struttura organizzativa coinvolta, per ciascuno dei quali vengono definiti i risultati attesi e gli indicatori per la valutazione qualitativa del lavoro svolto.

Tra i risultati attesi deve essere considerata anche l'autorizzazione alla pubblicazione rilasciata dai responsabili al trattamento dei diversi dataset nei formati più idonei al loro utilizzo. Ugualmente deve essere condivisa ed approvata la progettazione, secondo criteri di estesa accessibilità e di facilità d'uso, del sito di pubblicazione dei dati.

Ottenuta la qualità del dato prefissata occorre infine confezionare, in maniera leggibile e coerente, gli archivi così inventariati e pubblicabili verificando la loro congruenza in tutte le situazioni che richiedano la loro messa in relazione.

In questa fase ha un ruolo rilevante l'attività di metadattazione dei dataset propedeutica alla loro catalogazione ed in definitiva alla loro ricerca all'interno del sito predisposto alla pubblicazione. Gli standard di metadattazione dovranno essere scelti tra quelli certificati come ad esempio il Dublin Core per i dataset alfanumerici ed il 19115 per quelli geografici.



Il coordinamento di tutte le fasi deve essere svolto a livello centralizzato perché tempi, modalità e risultati siano omogenei.

Infine, una volta portata a termine la pubblicazione dei dati, è importante condividere con gli attori, tecnici e referenti di strutture, i risultati ottenuti anche in termini di realizzazione di nuovi servizi utili ai cittadini fatta all'esterno delle Amministrazioni da detentori di interessi specifici sui tematismi resi disponibili.

Il crescente interesse verso gli open data e la presa di coscienza delle potenzialità insite nell'utilizzo congiunto di informazioni coerenti, che siano provenienti da sorgenti diverse, ha messo in evidenza la necessità di sviluppare strumenti adeguati per l'interoperabilità. In quest'ottica, l'esigenza che nasce spontanea riguarda il "significato" che viene attribuito ai singoli dati, ovvero il loro livello "semantico". Volendo sintetizzare il problema, la questione consiste nella necessità di poter definire in modo chiaro e flessibile le equivalenze informative e le interconnessioni logiche che esistono fra gruppi di dati più o meno omogenei. Questo problema, da un punto di vista di un utilizzatore umano è meno sentito, in quanto le persone tendono naturalmente ad interpretare i dati, nel senso che tendono istintivamente ad associare un "significato" ai valori sulla base delle proprie conoscenze ed esperienze. Ad esempio, per un essere umano è semplice capire un campo nominato "CAP" e uno chiamato "c.a.p." all'interno di una rubrica telefonica sono equivalenti, ma questa associazione può non essere né semplice né tantomeno automatizzabile tramite un software. Discorso analogo vale per le possibili associazioni fra i le informazioni. Per questo motivo, esiste la necessità di associare ai dati anche un contenuto aggiuntivo che ne descriva gli aspetti semantici e le interconnessioni.

La tecnica che negli anni si è imposta su questo fronte deriva direttamente dai concetti sviluppati originariamente attorno al "web semantico" e consiste in ultima analisi nel produrre una descrizione dei dati mediante un grafo di nodi interconnessi, in cui sia i nodi che i collegamenti possono possedere svariati gruppi di proprietà. Il linguaggio che è assorto a standard de facto per la descrizione di queste informazioni è RDF, di cui esistono alcune varianti sintatticamente diverse, come XML-RDF e N3. I concetti alla base di questa tecnologia sono sostanzialmente due e consistono nel descrivere ogni entità o proprietà mediante proposizioni nella forma "soggetto-predicato-oggetti" e nell'identificare ogni "soggetto/predicato/oggetti" mediante valori espliciti oppure una URI, ovvero una risorsa univocamente definita tramite la sua collocazione spaziale nel World Wide Web. Attraverso questo paradigma è possibile descrivere collezioni di dati, sia dal punto di vista del loro effettivo contenuto (valorizzando gli opportuni campi o collegando le risorse specifiche) che della loro struttura, ma al contempo diventa assolutamente necessario poter rendere accessibili dal web le singole risorse quando la struttura delle informazioni richiede l'accesso ad un dato più complesso di quello semplicemente alfanumerico. Di contro, poiché questo aspetto di referenziazione mediante URI è intrinseco nella tecnologia, diventa particolarmente semplice creare "reti" di grafi interconnessi, raggiungendo così in modo naturale quella condizione di riferimenti incrociati che stanno alla base di un sistema di interoperabilità.

Più nello specifico, nella realizzazione di un insieme di open data semanticizzati (o "linked open data") è particolarmente importante la fase di progetto, quando viene definita la struttura dei dati, "ontologia", ed il "dizionario" ad essi associato. Queste due attività, intimamente connesse tra di loro, sono critiche non solo per l'effettiva trasformazione dei dati originali in un dataset semantico, ma anche per rendere possibile il collegamento a strutture dati esterne. In generale, l'approccio semplicistico a questa problematica consiste nel prendere come dato di fatto la struttura dell'informazione così com'è alla



sorgente dei dati e, quindi, trasferirla senza trasformazioni nel dizionario, lasciando come gradi di libertà nella compilazione di quest'ultimo solo la scelta su cosa debba essere entità o proprietà. Operando a questo modo, però, si corre il rischio di modellare nel grafo delle relazioni anche l'eredità che proviene dai sistemi che originano i dati, nel senso che il risultato finale sarà una proiezione della logica che sta dietro alle risorse dati. Una corretta progettazione del "dizionario ontologico", invece, deve tenere in considerazione l'essenza dei dati e la struttura logica dell'informazione che rappresentano, prendendo in esame vari aspetti fra di loro concorrenti. Da un lato, è importante che i dati prelevati dalla sorgente siano poi organizzati secondo uno schema a granularità molto fine, per consentire al documento che descrive la loro logica di aggregazione di essere sufficientemente puntale e specifico; questo, anche al fine di avere un numero il più elevato possibile di "punti di aggancio", che potrebbero essere utilizzati come base per la realizzazione delle eventuali interconnessioni con l'esterno. Dall'altro lato, invece, il dizionario deve catturare l'essenza dell'informazione attraverso entità e proprietà semplici, generali e ampiamente riutilizzabili, in modo da semplificare l'interpretazione dei dati e la definizione di equivalenze con dizionari esterni. Si crea quindi una situazione dove tanto più fine è la struttura atomica dell'informazione, tanto più è difficile evitare che il dizionario sia patologicamente adattato ai dettagli specifici dell'informazione in oggetto, con notevoli conseguenze negative in termini di flessibilità e riuso. Dall'altro, un dizionario troppo semplice e poco strutturato impoverisce l'informazione originale, producendo un documento di descrizione della struttura logica, che infine risulta insufficiente ed inutilizzabile.

La metodologia di riferimento in questo senso deve basarsi su alcune semplici linee guida che possono essere sintetizzate brevemente nelle indicazioni che seguono.

. Entità ricorrenti.

E' buona prassi definire come entità tutte quelle informazioni che appaiono con una certa frequenza all'interno dei dataset distribuiti. Questo permette di riutilizzare i "concetti" legati all'ontologia e agevola la stesura di un dizionario conciso.

. Modularità dell'ontologia.

Tutti i concetti che potrebbero avere natura diversa presso un altro Ente, oppure che potrebbero subire modifiche sostanziali nel corso del tempo, devono essere concepiti come entità individuali definite mediante un grafo RDF separato, ma interconnesso all'ontologia originale. Questo permette in modo semplice di estendere, modificare e sostituire sezioni intere dell'ontologia qualora ciò si renda necessario. Da un punto di vista tecnico, questo significa distribuire il dizionario su più "namespace", ovvero suddividere la definizione dei concetti all'interno di più domini di definizione separati.

. Riutilizzo dei dizionari.

Quando un concetto è già stato definito all'interno di un altro dizionario, questo dovrebbe essere importato senza ridefinirlo nuovamente. Ciò semplifica la scrittura del dizionario e aumenta le possibilità di interconnessione con altri linked open data.

. Creazione delle equivalenze.

Quando non è possibile riutilizzare un concetto già definito altrove, pur essendo questo coerente con quanto si vuole rappresentare, è almeno opportuno stabilire una relazione di equivalenza o di similarità fra i due. Questo è particolarmente importante quando due valori semanticamente equivalenti sono



definiti mediante "formati" diversi, come ad esempio se si ha a che fare con due campi Codice di Avviamento Postale uno scritto in formato stringa e l'altro come numero puro.



4. ESPERIENZA INPS

Il punto di partenza per la realizzazione di un processo di apertura dei dati è la definizione di una metodologia condivisa dagli enti. Tale metodologia deve, da una parte, dare indicazioni di massima sulle procedure da seguire, dall'altra, consentire l'adattamento delle fasi del processo alle esigenze specifiche dell'ente produttore.

In questo senso è stato pubblicato da alcuni mesi le "Linee guida OpenData per la PA", realizzato ed esposto da Dati.Gov.it con l'intenzione di offrire un vocabolario generale sul tema e di fornire un supporto pratico per le fasi di realizzazione.

Partendo dai punti indicati nelle linee guida – rispettivamente 1)Identificazione dei dati, 2)Analisi dei dati, 3)Pubblicazione dei dati – verranno ampliate e descritte con maggiore ricchezza tutte le tappe necessarie alla pubblicazione di dataset, esponendo le fasi del processo di apertura dei dati INPS nella loro realizzazione applicativa e nella loro importanza per lo sviluppo dei progetti opendata.

Fasi di lavoro

Il processo di individuazione e analisi dei dati è sicuramente il punto di partenza condiviso dal quale avviare il processo di apertura. Tuttavia non è sempre chiaro che cosa si intenda per analisi, come condurla e quali risultati aspettarsi. Si tratta in sostanza di capire come realizzare dataset di qualità, in modo che possano essere riutilizzati in modo libero ed adeguato, evidenziando i vantaggi per gli utenti e per la PA.

1- Analisi di dominio e individuazione dei dati

L'obiettivo è quello di identificare i contenuti già esposti dall'INPS all'interno delle relative aree di pubblicazione. In questa prima fase di lavoro, infatti, dove è molto comune trovarsi di fronte ad un gran volume di dati da gestire, è necessario organizzare il lavoro di selezione dei dati per diverse ragioni. Conoscere il dominio permette di:

- 1- Selezionare i dati rilevanti alla pubblicazione.*
- 2- Individuare i contenuti in relazione alle aree in cui sono esposti.*

E' infatti utile specificare, all'interno della struttura dei dataset ultimati, da quale area di riferimento provengono. In questo modo si contribuisce a mantenere l'autenticità della fonte, specificandone in modo dettagliato l'appartenenza.

- 3- Individuare le macroaree semantiche.*

Queste saranno poi utili alla definizione di quei metadati che rappresentano le informazioni sul contenuto del dato – i metadati semantici.



4- Individuare subito i dati che possono essere aperti.

Questo permette di distinguere quei dataset che necessitano di data refinement – cioè di una procedura di pulizia e normalizzazione della struttura del dato - da quelli che presentano una qualità già sufficiente alla loro pubblicazione. Questo approccio presenta dei vantaggi: è infatti possibile rendere parallele le filiere di lavoro sui dati, riadattando le fasi complessive del processo di liberazione alla specifica situazione dell'ente produttore.

L'analisi del dominio INPS, per quanto mostrato fino ad ora, ha mostrato l'evidente vantaggio di affrontare fin dall'inizio tutti i punti critici dell'intero processo fin dalla sua prima fase. In questo modo è possibile già da subito:

1- valutare costi e tempi dell'operazione complessiva.

2- Individuare eventuali punti critici del processo e pianificare immediatamente la loro risoluzione.

3- Individuare le esigenze specifiche dell'ente produttore e modificare di conseguenza la realizzazione del processo di apertura.

Ad esempio, all'interno del dominio INPS sono state individuate le seguenti aree in cui erano già esposti dati:

- Relazione Annuali
- Bilanci Sociali
- Osservatorio Statistico
- Statistiche della previdenza
- Rapporti sulla coesione sociale

Per ogni area è stata verificata la qualità dei dati presenti e il formato in cui erano esposti. In questo modo è stato possibile organizzare un flusso di lavoro che avesse come primo obiettivo quello di affrontare i punti più deboli – come i dati esposti in formato .pdf – e risolverli – convertendoli da .pdf al formato superiore .xls.

Sono necessarie competenze di Knowledge Management per affrontare questa prima fase in modo efficace. E' altresì necessario che tali competenze siano affiancate da uno o più esperti di dominio, in modo tale che i dati mantengano, in ogni loro fase di lavorazione, tutte le caratteristiche che lo rendono appartenente alla fonte originale.

2- Analisi tipologica dei dati

Una volta individuati i dati, la loro fonte e la loro qualità si procede a identificare la tipologia. Questa fase presenta due livelli distinti.

Il primo livello di analisi individua il tipo in base al modo in cui è stato prodotto. E in questo caso utile distinguere tra:



- 1- *Dati grezzi*: (da non confondere con la definizione “Raw” trad. *Grezzo* relativa al formato) questi hanno la caratteristica di essere raccolte di dati semplici, nei quali di solito è presente una sola chiave con un valore associato. Ad esempio, la tabella “Nuclei familiari con entrambi i genitori e almeno un figlio minore” individua l’importo dell’ANF in relazione al reddito familiare.

NUCLEI FAMILIARI CON ENTRAMBI I GENITORI E ALMENO UN FIGLIO MINORE IN CUI NON SIANO PRESENTI COMPONENTI INABILI												
Importo complessivo mensile dell'assegno per livello di reddito e numero componenti il nucleo												
Reddito familiare annuo di riferimento valido dal 1° luglio 2010												
Reddito familiare annuo (euro)	Importo dell'assegno per numero dei componenti il nucleo familiare											
	1	2	3	4	5	6	7	8	9	10	11	12
fino a 13.211,14			137,50	258,33	375,00	500,00	625,00	773,75	922,50	1.071,25	1.220,00	1.368,75
13.211,15 - 13.316,82			136,73	257,25	374,04	499,58	624,38	773,04	921,69	1.070,35	1.219,01	1.367,67
13.316,83 - 13.422,51			135,95	256,17	373,08	499,17	623,75	772,31	920,88	1.069,44	1.218,00	1.366,56
13.422,52 - 13.528,20			135,18	255,08	372,13	498,75	623,13	771,60	920,07	1.068,54	1.217,01	1.365,48
13.528,21 - 13.633,89			134,40	254,00	371,17	498,33	622,50	770,88	919,25	1.067,63	1.216,00	1.364,38
13.633,90 - 13.739,58			133,63	252,92	370,21	497,92	621,88	770,16	918,44	1.066,73	1.215,01	1.363,29
13.739,59 - 13.845,27			132,85	251,83	369,25	497,50	621,25	769,44	917,63	1.065,81	1.214,00	1.362,19
13.845,28 - 13.950,96			132,08	250,75	368,29	497,08	620,63	768,72	916,82	1.064,91	1.213,01	1.361,10
13.950,97 - 14.056,65			131,30	249,67	367,33	496,67	620,00	768,00	916,00	1.064,00	1.212,00	1.360,00

- 2- *Dati statistici*: sono i dati già manipolati, spesso prodotti nella gestione e nel controllo dell’attività di terzi e dell’istituto stesso. Generalmente questi dati presentano una struttura elaborata più complessa rispetto al primo caso individuato. Ad esempio, la tabella “Riepilogo attività di vigilanza dei cari enti” individua in riferimento agli anni 2000/2001/2002 il numero di aziende divise per tipologia (regolari, non regolari, totalmente irregolari) in relazione alla regione di appartenenza.

Tab. 1 - Riepilogo attività di vigilanza dei vari enti. Province del Veneto.
Anni 2000-2001-2002

	2000				2001			
	Regolari	Non regolari	Totale	Irreg./tot. (a)	Regolari	Non regolari	Totale	Irreg./tot. (b)
Belluno								
Congiunta Speciale*	12	11	23	47,8%	4	1	5	20,0%
Ministero del lavoro	2	3	5	60,0%	48	8	56	14,3%
Inps	108	109	217	50,2%	200	225	425	52,9%
Inail	65	145	210	69,0%	383	242	625	38,7%
Altri enti	100	69	169	40,8%	87	83	170	48,8%
Altri enti	1	0	1	0,0%	0	3	3	100,0%
Totale**	274	323	597	54,1%	670	553	1.223	45,2%
Padova								
Congiunta Speciale*	0	3	3	100,0%	35	45	80	56,3%
Ministero del lavoro	1	13	14	92,9%	26	34	60	56,7%
Inps	270	331	601	55,1%	376	461	837	55,1%
Inail	1.026	938	1.964	47,8%	1.490	1.081	2.571	42,0%
Altri enti	343	346	689	50,2%	184	350	534	65,5%
Altri enti	18	13	31	41,9%	0	17	17	100,0%
Totale**	1.657	1.628	3.285	49,6%	2.050	1.909	3.959	48,2%
Rovigo								

Questa operazione, oltre a tracciare una distinzione in base alla tipologia di produzione del dato,



consente di verificare il grado di complessità strutturale dei dataset già presenti. All'interno del dominio INPS questa fase si è rivelata determinante per progettare l'implementazione dei formati in cui i dati sono pubblicati all'interno del portale OpenData INPS. L'obiettivo di questa verifica è stato quello di identificare il grado di annidamento dei dati nei tipi di dataset presenti. E' stata effettuata una distinzione a tre livelli:

I° livello: semplice

	Maschi	Femmine
	Numero	Numero
2005	71.207	43.344
2006	59.640	40.884
2007	54.268	40.338
2008	54.417	40.263
2009	75.855	49.804

II° livello: medio

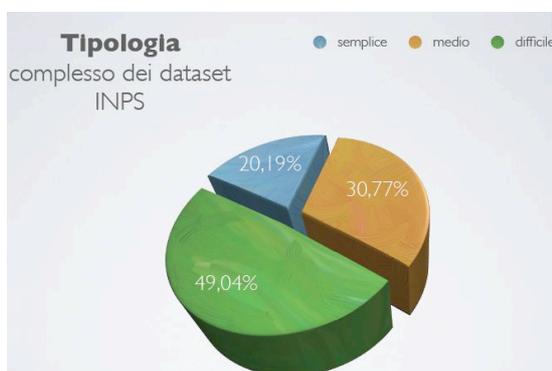
Anno	ARTIGIANI-COMMERCianti			
	Titolari di qualunque età e coadiuvanti /coadiutori di età superiore ai 21 anni		Coadiuvanti /coadiutori di età inferiore ai 21	
	I fascia di reddito	II fascia di reddito	I fascia di reddito	II fascia di reddito
2011	20,00%	21,00%	17,00%	18,00%
2012	21,30%	22,30%	18,30%	19,30%
2013	21,75%	22,75%	18,75%	19,75%
2014	22,20%	23,20%	19,20%	20,20%
2015	22,65%	23,65%	19,65%	20,65%
2016	23,10%	24,10%	20,10%	21,10%
2017	23,55%	24,55%	20,55%	21,55%
2018	24,00%	25,00%	21,00%	22,00%

III° livello: difficile

Tavola III.3.1.2 - Beneficiari di disoccupazione non agricola requisiti ordinari e disoccupazione speciale mensile 2007-2009; 1° semestre 2010 (valori assoluti)

MESI	Nord			Centro		
	Numero Beneficiari	Entrate	Uscite	Numero Beneficiari	Entrate	Uscite
	2007					
gennaio	86.272	-	15.785	37.866	-	6.128
febbraio	80.645	10.158	15.533	36.532	4.794	5.817
marzo	75.625	10.513	20.788	35.355	4.640	8.172
aprile	68.898	14.061	15.803	31.553	4.370	7.841
maggio	62.917	9.822	15.554	27.481	3.769	6.852
giugno	58.650	11.287	13.167	25.353	4.724	4.880
luglio	74.009	28.526	10.113	33.159	12.686	4.150
agosto	76.158	12.262	19.247	34.484	5.475	9.076
settembre	71.388	14.477	23.965	30.584	5.176	9.487
ottobre	73.481	26.058	11.188	31.049	9.952	5.132
novembre	87.194	24.899	13.056	36.117	10.200	5.094
dicembre	94.310	20.172	16.503	38.213	7.190	4.543
Media annua beneficiari	75.796	-	-	33.146	-	-

E' stato così possibile identificare il complesso dei dataset per tipologia all'interno del dominio INPS.



Il secondo livello di analisi permette di distinguere i dati che contribuiscono alla trasparenza, e che quindi sono legati maggiormente all'Open Government, da quelli di generale interesse pubblico. Rispettivamente:

- 1- (*Open Government*) data: tutte le tipologie di dati che contribuiscono alla trasparenza della amministrazioni. Ad esempio, dati sui CV dei dirigenti o dati sugli appalti pubblici.
- 2- *Open (Government data)*: tutte le tipologie di dati di interesse pubblico. Ad esempio la posizione degli uffici amministrativi o l'importo medio delle pensioni.

Questa distinzione permette da una parte di rendere evidenti i dati sulla trasparenza nella PA, dall'altra di isolare tutti quei dati di interesse pubblico a fini applicativi (come le applicazioni di mashup sviluppati a partire da OpenData).

3- Preparazione dei dati e definizione dei formati

Questa fase del processo è cruciale per la qualità dei dati da pubblicare. Poiché molte delle fonti presentano le informazioni in maniera non strutturata è opportuno effettuare un'operazione di pulizia sulla struttura e sul contenuto delle tabelle. La qualità dei dati non è solo un criterio estetico. Creare i primi formati (.xls per lo più) in modo preciso e puntuale consente di:

- 1- Implementare in modo più veloce i formati successivi (XML, RDF fino ai Linked Open Data).
- 2- Offrire dati più facilmente leggibili sia dagli utenti che dalle macchine. In questo modo si favorisce sia la comunicazione e la trasparenza con i cittadini, sia un riuso basato su servizi e applicazioni ad integrazione di dataset (mashup, openservice).

Ad esempio, una buona formattazione .XLS consente una più facile conversione XML tramite XSD (XLSschema) e CSV. Un esempio di tabella strutturata in modo coerente e pulito.

REGIONE	NOME ALBERGO	n° stelle	Cat. tariff. Terme	VIA	CAP	LOCALITA'	Provincia
Abruzzo	MAIELLA E DELLE TERME	3	1S	Via Roma, 29	65023	Caramanico terme	PE
Abruzzo	VINCENZELLA	3	1S	Via Roma, 21	65023	Caramanico Terme	PE
Abruzzo	VIOLA	3	1S	Via della libertà, 9	65023	Caramanico Terme	PE
Basilicata	TERME LUCANE S.r.l. - Hotel Monte Alpi	3	1S	Contrada Calda	85043	Latronico	PZ

La stessa tabella prima del refinement:



CURE FORME VIE RESPIRATORIE						
ELENCO ALBERGHI CONVENZIONATI PER LA STAGIONE TERMALE 2011 (V. N.B. IN FONDO)						
LOCALITA' e Provincia	NOME ALBERGO	n° stelle	Cat. tariff. Terme	Indirizzo albergo n° telef. e n° fax	Distanza in metri dalla terme e mezzo di collegamento (solo alberghi non dotati di terme interne)	Trattamenti particolari (non termali) praticati gratuitamente
Abruzzo						
CARAMANICO TERME (PE)	MAIELLA E DELLE TERME	3	IS	Via Roma, 29 - 65023 Caramanico Terme (PE) tel. 085/92201 fax 085/923070 e-mail:info@albergomaella.it	INTERNE	Albergo annesso allo Stabilimento termale; parcheggio riservato non custodito; camere con servizi privati; tv color sat e telefono diretto; piccola colazione a buffet; frutta a buffet; menù a scelta; dessert (due volte alla settimana); animazione; pianobar
CARAMANICO TERME (PE)	VINCENZELLA	3	IS	Via Roma, 21 - 65023 Caramanico Terme (PE) tel. 085/922104	30 metri	Dolce a pranzo la domenica e il giovedì, escursioni naturalistiche
CARAMANICO TERME (PE)	VIOLA	3	IS	Via della libertà, 9 - 65023 Caramanico Terme (PE) Tel. 085/922292 - fax: 085/922189	50 metri	Una serata danzante settimanale
CARAMANICO TERME (PE)	IACOBUCCI	3	IS	Via della Libertà 7 - 65023 Caramanico Terme (PE) Tel. 085-922103	15 metri	
Basilicata						
LATRONICO (POTENZA)	TERME LUCANE S.r.l. - Hotel Monte Alpi	3	a	Contrada Calda 85043 - Latronico (PZ) telef. 0973/859238 fax 0973/859856	250 metri - bus navetta	Sconto del 30% sulle tariffe in vigore
Calabria						
ACQUAPPESA-GUARDIA PIEMONTESE(CS)	GRAND HOTEL	4	1a SUPER	LOCALITA' TERME LUGIANE GUARDIA PIEMONTESE N.TELEFONO 0982-94052, N.FAX 0982.94478	TERME INTERNE	SERVIZIO DI ANIMAZIONE-SERVIZIO NAVETTA DA E PER LA SPIAGGIA-INGRESSO PARCO TERMALE IN OCCASIONE DELLE MANIFESTAZIONI DELL'ESTATE CULTURALE-SCONTO DEL 50% SUL PREZZO DI INGRESSO AL PARCO TERMALE.
ACQUAPPESA-GUARDIA PIEMONTESE(CS)	ALBERGO TERMALE MODERNO	3	1a SUPER	LOCALITA' TERME LUGIANE GUARDIA PIEMONTESE N.TELEFONO 0982-94048, N.FAX 0982.94478	250 metri di strada carrabile e pedonale,collegamento con servizio di navetta dedicato.	SERVIZIO DI ANIMAZIONE-SERVIZIO NAVETTA DA E PER LA SPIAGGIA-INGRESSO PARCO TERMALE IN OCCASIONE DELLE MANIFESTAZIONI DELL'ESTATE CULTURALE-SCONTO DEL 50% SUL PREZZO DI INGRESSO AL PARCO TERMALE.

In questo modo è possibile produrre XML e CSV di qualità senza dover ripetere ad ogni conversione l'operazione di pulizia e controllo. Un esempio di XML INPS.

```
<data>
  <datum Anno="2007" Regione_Area_Geografica="Piemonte">
    <Maschi>2595</Maschi>
    <Femmine>1674</Femmine>
    <Maschi_e_Femmine>4269</Maschi_e_Femmine>
  </datum>
  <datum Anno="2007" Regione_Area_Geografica="Valle d'Aosta / Vallée d'Aoste">
    <Maschi>826</Maschi>
    <Femmine>161</Femmine>
    <Maschi_e_Femmine>987</Maschi_e_Femmine>
  </datum>
  <datum Anno="2007" Regione_Area_Geografica="Lombardia">
    <Maschi>3898</Maschi>
    <Femmine>1447</Femmine>
    <Maschi_e_Femmine>5345</Maschi_e_Femmine>
  </datum>
</data>
```

4- Definizione dei metadati

Lo step successivo nella creazione degli OpenData INPS è stata la definizione dei metadati, cioè informazioni che descrivono uno o più insieme di dati. Tali informazioni rappresentano delle proprietà più significative fissate e consentono la ricerca e il recupero dei dati stessi con lo scopo di migliorarne la visibilità e facilitarne l'accesso.

Esistono alcuni importanti standard metadata, tra i più importanti sicuramente gli ISO 15836 Dublin Core. Da poco il W3C ha rilasciato uno standard metadata per gli OpenData (ADMS – Asset Description Metadata Standard) per favorire l'interoperabilità al livello europeo. Tuttavia i vocabolari dei metadati sono molto ampi e richiedono una selezione di quelli che effettivamente possono



arricchire di informazioni i dati in questione. Anche qui è quindi necessario adattare le specificità di ciascun ente agli standard trovati.

La creazione e l'implementazione dei metadati richiede attenzione alle risorse informative che tali metadati devono riflettere. Molta attenzione deve essere data alla qualità, perché un pessimo metadato è peggio dell'assenza di metadati. Per l'INPS è stata realizzato un set che include gli standard Dublin Core ma li arricchisce con alcuni tipi di metadati proprio del dominio INPS. Questi sono di due tipi:

- *Metadati sintattici*: descrivono il dato attraverso le sue caratteristiche strutturali, in cui sono comprese le informazioni sul processo di creazione, organizzazione e pubblicazione. Tali metadati non aggiungono informazioni sul contenuto del dataset.

Titolo: titolo dei dati

Data di pubblicazione: Questa informazione riguarda la data di pubblicazione del dato all'interno del portale/minisito OpenData Inps.

Formato: l'estensione del file (.pdf, .xls; .rdf; .doc)

Licenza: il tipo di licenza a cui è legato il dato

Dimensione in byte: la dimensione fisica del dato

Identificativo: numero progressivo univoco del dato

Link: (URL-permanent link-, link annidati, short URL)

Fonte: opzionale, chi ha prodotto il dato (Direzione, altro soggetto)

- *Metadati semantici*: descrivono il dato attraverso le sue caratteristiche semantiche, ovvero le informazioni sul contenuto. Tali aspetti consentono di individuare univocamente le risorse.

Copertura temporale: rappresenta l'estensione temporale del dato (es. ante 2000, 2001-2006-2011, 2012-)

Livello di lavorazione del dato: grezzo, statistica, bilancio.

Tags: parole chiave dell'argomento principale

Area di pubblicazione: a sezione in cui viene pubblicato all'interno del portale Inps.it.

Descrizione: un breve abstract del dato

Data di aggiornamento: mostra ogni quanto il dato viene tenuto aggiornato.

Argomento trattato: il soggetto semantico del dato



Pertanto ogni dataset INPS è identificato in questo modo, ed è possibile trovarlo sul portale INPS Open Data.



FASCE DI REDDITO FAMILIARE PER NUCLEI SENZA FIGLI	
titolo	Fasce di reddito familiare per nuclei senza figli
descrizione	Il documento riporta le fasce di reddito entro le quali calcolare l'importo dell'assegno per il nucleo familiare senza figli a carico.
Tag	ANF; prestazioni; famiglia; reddito
Argomento trattato	Prestazioni sociali/ANF
Data di aggiornamento	27 giugno 2004
Formato	.xls
licenza	IODL
dimensione	358 kb
identificativo	ID12
URL	http://www.inps.it/bussola/visualizzadoc.aspx?sVirtuAIURL=/circolari/./circolariZip/Circolare numero 70 del 30-5-2005_Allegato n 1.xls
Livello di lavorazione	statistico
Copertura spaziale	nazionale
Copertura temporale	1° luglio 2003 - 30 giugno 2005

Una volta descritta la risorsa attraverso opportuni metadati è necessario rappresentare il livello di descrizione anche nei successivi formati di pubblicazione.

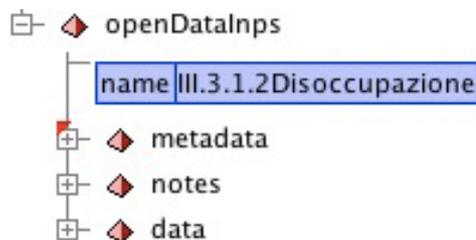
Dal formato XML in poi è utile implementare l'interoperabilità dei dati. E' stato ritenuto importante, quindi, fare nuovamente un match con gli standard metadata Dublin Core al fine di ritornare una lista metadati:

- 1- coerente con quelle internazionali.
- 2- Adeguata agli standard di interoperabilità europea.
- 3- specifica per INPS.

Metadati INPS	Adeguamento Dublin Core
Titolo	dc: title
Data di Pubblicazione	dcTerms: issued
Licenza	dcTerms: license
Link	dc: Identifier
Fonte	dc: Publisher
Copertura Temporale	dcTerms: Temporal
Area	dc: Source
Descrizione	dc: Abstract
Argomento trattato	dc: Subject

Ecco quindi l'esempio di una struttura XML per un dataset INPS, completa di metadati, note ai dati e dati. Qui la struttura complessiva dell'XML:





Note ai dati:

```
<metadata>
<notes>
  <note>(*) Soggetti che hanno presentato la domanda entro il mese di marzo dell'anno
  <note>Fonte: INPS - Coordinamento Generale Statistico Attuariale</note>
</notes>
<data>
```

Metadati:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<openDataInps xmlns="http://www.inps.it/OpenData/1.0" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcterms="">
  <metadata>
    <dc:publisher>INPS</dc:publisher>
    <dc:title>III.3.1.6 - Beneficiari di indennità di disoccupazione non agricola (*) con requisiti ridotti
    <dcterms:abstract>Rapporto sulla coesione sociale - Dati sui beneficiari di indennità di disoccupazione n
    <dcterms:issued>17/01/2012</dcterms:issued>
    <dcterms:license>IODL 2.0</dcterms:license>
    <dcterms:LicenseDocument>http://www.dati.gov.it/i odl/2.0/</dcterms:LicenseDocument>
    <dc:identifier>http://www.inps.it/docallegati/Mig/OpenData/III.3.1.6_Disoccupazione.xls</dc:identifier>
    <facets>
      <argomento>
        <dc:subject>Lavoro.LavoroDipendente</dc:subject>
        <dc:subject>Lavoro.AltriTipiDiLavoro</dc:subject>
        <dc:subject>PrestazioniAssistenziali.Disoccupazione</dc:subject>
      </argomento>
      <fonte>
        <dcterms:source>Rapporti</dcterms:source>
      </fonte>
      <periodo>
        <dcterms:temporal>2006-2011</dcterms:temporal>
      </periodo>
    </facets>
  </metadata>
</openDataInps>
```



Dati:

```
<metadata>
<notes>
<data>
  <datum Anno="2007" Regione_Area_Geografica="Piemonte">
    <Maschi>7266</Maschi>
    <Femmine>13701</Femmine>
    <Maschi_e_Femmine>20967</Maschi_e_Femmine>
  </datum>
  <datum Anno="2007" Regione_Area_Geografica="Valle d'Aosta / Vallée d'Aoste">
    <Maschi>788</Maschi>
    <Femmine>1073</Femmine>
    <Maschi_e_Femmine>1861</Maschi_e_Femmine>
  </datum>
  <datum Anno="2007" Regione_Area_Geografica="Lombardia">
    <Maschi>10364</Maschi>
    <Femmine>25541</Femmine>
    <Maschi_e_Femmine>35905</Maschi_e_Femmine>
  </datum>
  <datum Anno="2007" Regione_Area_Geografica="Trentino-Alto Adige">
    <Maschi>3129</Maschi>
    <Femmine>5431</Femmine>
    <Maschi_e_Femmine>8560</Maschi_e_Femmine>
  </datum>
  <datum Anno="2007" Regione_Area_Geografica="Bolzano-Bozen">
    <Maschi>1000</Maschi>
    <Femmine>1000</Femmine>
    <Maschi_e_Femmine>2000</Maschi_e_Femmine>
  </datum>
</data>
</notes>
</metadata>
```

5- *Attribuzione della licenza*

L'INPS ha scelto di pubblicare la propria sezione aggiornata della licenza IODL 2.0 (Italian Open Data License). Tale licenza prevede che l'utente possa liberamente:

- 1- Consultare, estrarre, copiare e pubblicare i dati.
- 2- Creare un lavoro derivato integrando diversi dataset.

Tale licenza non prevede, al contrario della precedente versione 1.0, l'obbligo dell'utente di pubblicare o condividere i lavori con la stessa licenza. Richiedendo semplicemente di indicare la fonte gli utenti sono incentivati al riutilizzo attivo dei dati.

Gli Open Data INPS offrono all'utente un ricco set di informazioni intorno al dato stesso che ne migliorano l'esperienza di uso e di riuso, e garantiscono, essendo parte integrante del dataset, l'autenticità della fonte.



5. ESPERIENZA MIUR

Recentemente, in merito al tema degli Open Data, il MIUR ha avviato un'attività di apertura di dati pubblici.

L'iniziativa è nata per volontà del Ministro dell'Istruzione, Università e Ricerca, impegnato a valorizzare il proprio patrimonio informativo ed a favorire la trasparenza amministrativa nei processi di gestione e partecipazione al miglioramento del sistema scolastico. Ciò anche per favorire la nascita di una nuova generazione di servizi per studenti, insegnanti e famiglie.

Analizzando i dati a disposizione, è stato deciso di aprire in prima istanza quelli relativi agli alunni, al personale scuola e alle strutture scolastiche, in quanto subito utilizzabili da un ampio pubblico per aggregazioni e rilevazioni statistiche.

L'iniziativa trova la sua governance nella Direzione Generale per gli Studi, la Statistica ed i Sistemi Informativi.

L'iniziativa è stata avviata, nell'intento di sviluppare una maggiore sensibilità al tema oltre che la necessaria base di conoscenza teorica e degli strumenti necessari, formando quattro risorse interne presso l'Università La Sapienza – Università di Roma, organizzatrice del Master di II livello in Interoperabilità per la pubblica amministrazione e le imprese.

La scelta di puntare da subito sugli open data è stata rafforzata dagli indirizzi del Ministro di cui all'avviso pubblico nota n.11789 del 23/12/2011, in virtù del quale il MIUR si è dotato di competenze altamente specializzate in materia di Open Data - semplificazione e trasparenza delle pubbliche amministrazioni e accesso libero ai dati attraverso le tecnologie dell'informazione.

I dati scaricabili all'URL http://archivio.pubblica.istruzione.it/scuola_in_chiaro/open_data/index.html e resi in una forma organica nell'ambito del progetto "Scuola in chiaro", sono relativi a tutte le scuole italiane di ogni ordine e grado. "Scuola in chiaro", attivato sul sito del MIUR dal 12 gennaio 2012, fornisce tutti i dati posseduti dal sistema informativo e consente ad ogni Istituzione scolastica di pubblicare i propri, secondo un indice ed un format comune.

I dataset "made in Scuola in chiaro" sono costituiti, infatti, da dati già presenti presso il Ministero e rivolti sia alle scuole perché li possano aggiornare che alle famiglie perché colgano l'opportunità di cercare, trovare e valutare la scuola che meglio risponde alle esigenze formative dei propri figli e di procedere alla loro iscrizione on line, attraverso una semplice procedura guidata.

I dati comunque, possono essere utilizzati per ogni scopo, personale o commerciale, e senza vincoli, al fine di comprendere meglio il mondo della scuola e creare servizi innovativi.

Fasi di lavoro

La metodologia seguita consiste nelle fasi di seguito riportate:



Data Cleaning: molto spesso i dati in possesso non rispettano dei criteri di qualità che invece sono richiesti al fine di rendere i dati, una volta pubblicati, comprensibili a tutti. Alcuni di questi criteri riguardano la pulizia del dato, ovvero i dati possono essere “sporchi” se ad esempio non sono conformi al fenomeno che descrivono e quindi non sono confrontabili con altri, se utilizzano dei valori propri dell'applicazione che li utilizza come ad esempio dei caratteri di controllo. Pertanto una fase di data cleaning è stata avviata per far in modo che i dati fossero conformi a controlli di qualità. Il cleaning è stato svolto con la collaborazione del Fornitore del servizio di gestione del sistema informativo (di seguito Fornitore).

Modeling: dopo la fase descritta in precedenza, si è proceduto alla fase di modellazione con la scelta dei concetti e dei relativi dati da pubblicare. Questo ha portato alla creazione di molte istanze di classi, ognuna descritta dai dati interni al database del MIUR.

Mapping: al termine delle fasi precedenti si sono “puliti” i dati in maniera tale da essere trasformati secondo una modellazione ben definita. Questa modellazione ha portato alla creazione di file .xls. Gli strumenti da utilizzare sono stati scelti sulla base della loro facilità di utilizzo. Il formato utilizzato rende il dato open ma non machine-readable: idoneo pertanto solo ad elaborazione manuali dei dump e pertanto idoneo a consultazione da parte di stakeholders come giornalisti e sindacati.

Gestione ed evoluzione: le attività di gestione e aggiornamento del sistema di generazione degli open data sono portate avanti da funzionari statistici in servizio presso la Direzione Generale per gli Studi, la Statistica ed i Sistemi Informativi e la Direzione Generale per il Personale Scolastico.

I database da cui sono estratti i dati sono: Anagrafe degli Studenti, Osservatorio Tecnologico, Anagrafe delle Scuole ed il SIDI (Sistema Informativo dell'Istruzione).

L'aggiornamento è previsto a cadenza periodica e segue specifici eventi amministrativi, quali l'iscrizione degli studenti e la distribuzione dell'organico del personale docente.

Della manutenzione correttiva ed evolutiva si occupa il Fornitore

All'indirizzo dati@istruzione.it possono essere recapitate indicazioni, domande e suggerimenti da parte dei riutilizzatori. Esse costituiscono un elemento fondamentale per migliorare l'approccio del MIUR verso i linked data. I file sono in formato .csv mentre le licenze d'uso sono del tipo C.C. 3.0 (Creative Commons 3.0).

I passi successivi mirano sia a rendere open altri dataset che a rendere questi ultimi “linkabili”. La pubblicazione di questi dataset rappresenta il primo passo di una strategia che ha l'obiettivo di rendere disponibile e riutilizzabile il patrimonio informativo detenuto dagli uffici dell'Istruzione Pubblica nell'ottica di promuovere un modello di “governo aperto”, fondato su un rinnovato rapporto di fiducia e collaborazione tra pubblico e privato. L'evoluzione del progetto include ulteriori interventi volti a replicare lo stesso modello di condivisione su altre basi di dati a disposizione del MIUR e perfezionarne la qualità tecnica.

Il processo di riutilizzo è fondamentale per il MIUR che invita tutti coloro che hanno realizzato una soluzione, un servizio, una ricerca con questi dati a comunicarlo a dati@istruzione.it in modo da raccogliere i casi d'uso e offrire loro tutta la visibilità possibile.



6. ESPERIENZA COMUNE DI NAPOLI

La crescente richiesta avuta dall'ente per lo sviluppo di applicazioni mobili, soprattutto nell'ambito turistico ci ha spinto nella direzione dell'open data.

Ci siamo chiesti come ente perché sviluppare un APP in proprio? Perché non fornire tutti i dati in nostro possesso soprattutto in ambito turistico e consentire a sviluppatori, studenti, analisti sociologi di sviluppare applicativi sui nostri dati?

Da queste semplici domande, alcuni mesi fa è iniziato un processo di organizzazione e studio sull'Open Data.

In tal senso, si è partiti nel reperire i primi semplici dati già disponibili da pubblicare. Non a caso i primi dati che saranno resi pubblici sono quelli relativi a siti di particolare:

- a) interesse turistico, eventi organizzati all'interno delle strutture turistiche e culturali;
- b) rilevamento del distributore o distributori più economici, relativamente al prezzo del gasolio e della benzina in città.
- c) Probabile anche la disponibilità immediata dei dati relativi al posizionamento dei mezzi pubblici (bus, funicolari).

Fasi di lavoro

Al momento della disponibilità di tale documento il Comune di Napoli sta organizzando la struttura dati relativa alla gestione e pubblicazione web dei siti di particolare interesse turistico, essendo tale sito di fondamentale importanza sia per la georeferenziazione degli access point wifi previsti per i prossimi mesi, i quali visualizzeranno nelle captive portal eventi e siti turistici più vicini alla loro posizione sul territorio, sia per consentire lo sviluppo o l'integrazione in app mobili di tali dati.

Infine rendere pubblici dati attendibili e già disponibili sui distributori più economici di carburante e il posizionamento di mezzi pubblici.

Allo stato attuale sono impegnate sul progetto open data 3 risorse umane interne all'ente, ed è prevista consulenza gratuita da parte dell'università Federico II di Napoli, per gestione APP già sviluppate.

Al momento sono allo studio le mappe fornite da Google (foto aeree) e le mappe fornite da Microsoft (che consentono di visualizzare anche il singolo piano, avendo un angolo di circa 45%).



7. ESPERIENZA COMUNE DI PADOVA

Il Comune di Padova sta affrontando il progetto per la disponibilità al pubblico di propri dati. Relativamente ai dataset, intende avviare procedure operative istituzionali con prevalenza automatiche atte a garantire la produzione e l'aggiornamento costante dei dati individuati.

I dati di riferimento sono reperiti dai documenti relativi alla normale attività amministrativa. Un nuovo sistema tecnologico estrae dall'insieme del database-ente i dati appositamente marcati tenuto conto delle tutele previste dalla normative sull'accesso, sulla riservatezza e sulla privacy.

E' attivato un gruppo di lavoro atto a individuare e codificare i dati che verranno trattati da nuovi processi informatici orizzontali al fine di produrre i dataset stabiliti.

Uno dei compiti consiste nella individuazione delle tassonomie dei dataset entro le quali i dati, anche in formato linked verranno esposti. Un esempio della gerarchia in corso di adozione è la seguente:

Classe	Sottoclasse
Amministrazione	Archivio Territorio Toponomastica Associazionismo
Risorse Finanziarie e Patrimoniali	Bilanci Piano Esecutivo di gestione Rendiconto di gestione Beni Immobili Beni Mobili Concessionari ed altri incaricati Pubblicità
Pianificazione e Gestione del territorio	Urbanistica: piano regolatore generale e varianti Urbanistica: strumenti di attuazione del Piano regolatore generale



	<p>Edilizia privata</p> <p>Edilizia pubblica</p> <p>Opere pubbliche</p> <p>Catasto</p> <p>Viabilità</p> <p>Trasporti pubblici</p> <p>Gestione dei rifiuti</p> <p>Ambiente: autorizzazioni, monitoraggio e controllo</p> <p>Protezione civile ed emergenze</p> <p>Concessioni edilizie</p> <p>Concessioni di agibilità</p> <p>.....</p>
<p>Servizi alla persona</p>	<p>Asili nido e scuola materna</p> <p>Istituti culturali (Musei, biblioteche, teatri, Scuola comunale di musica</p> <p>Attività ed eventi culturali</p> <p>Attività ed eventi sportivi</p> <p>Assistenza diretta e indiretta, benefici economici</p> <p>Politiche per la casa</p> <p>Politiche per il sociale</p>
<p>Attività economiche</p>	



	<p>Agricoltura e pesca</p> <p>Artigianato</p> <p>Industria</p> <p>Commercio</p> <p>Fiere e mercati</p> <p>Esercizi turistici e strutture ricettive</p> <p>Promozione e servizi</p>
Polizia locale e sicurezza pubblica	<p>Prevenzione ed educazione stradale</p> <p>Autorizzazioni di pubblica sicurezza</p>
Tutela della salute	<p>Salute e igiene pubblica</p> <p>Farmacie</p> <p>Zooprofilassi veterinaria</p>
Servizi demografici	<p>Stato civile</p> <p>Censimenti</p>
Elezioni e iniziative popolari	<p>Elezioni</p> <p>Referendum</p> <p>Sezioni elettorali</p>

Relativamente al contesto della tutela in particolare la legge 241/90 e il Codice in materia di trattamento dei dati personali, il nuovo sistema tecnologico è in grado di individuare i dati con le caratteristiche necessarie per essere dichiarati appartenenti alla categoria Open Data. E' ritenuto vincolante la definizione chiara dei tempi oltre i quali, dati sottoposti a particolari restrizioni, possano assumere lo status di Open Data.



La nuova tecnologia, integrata con le altre tecnologie di trattamento dei dati e di gestione dei documenti e dei procedimenti amministrativi, è in grado di estrapolare automaticamente e con certezza di diritto i dati e di costituire e mantenere costantemente aggiornati i dataset stabiliti.



8. ESPERIENZA TRENTO: IL GEOCATALOGO SEMANTICO

L'ambito di applicazione del progetto è quello dei dati territoriali. I dati territoriali hanno a che fare la conoscenza del mondo reale e di fatto costituiscono l'elemento conoscitivo di riferimento per tutte le politiche connesse al governo e alla gestione territorio. Si tratta quindi di informazioni essenziali sia come strumento di sviluppo sia come supporto alle decisioni.

L'obiettivo del progetto è favorire l'interoperabilità e la collaborazione tra le strutture che si occupano di pianificazione e gestione economico-produttiva e ambientale del territorio trentino attraverso la pubblicazione ordinata e aperta di dati territoriali linkabili. Nello specifico, il progetto è volto ad ampliare il portafoglio di servizi del Portale Geocartografico Trentino realizzando un servizio che mette a disposizione Geocatalogo Semantico che permetta di classificare, descrivere e ricercare le informazioni relative a geo-dati e geo-servizi. I benefici attesi sono essenzialmente (1) il miglioramento dell'accuratezza dei risultati di ricerca delle geo-informazioni che dovrebbero quindi portare a decisioni più accurate e (2) l'avanzamento verso la società della conoscenza attraverso lo sviluppo della cultura dell'uso di geo-informazioni nella vita quotidiana.

L'impostazione del progetto di sviluppo del Geocatalogo Semantico¹ è un esempio di Linked Open Data (LOD) coerente dal punto di vista organizzativo e tecnico con la visione di sviluppo dell'Infrastruttura Dati Territoriali del territorio provinciale ed è conforme alle indicazioni direttiva INSPIRE a livello europeo, e delle direttive IntesaGIS e DigitPA-RNDT a livello nazionale. L'infrastruttura qui intesa come una struttura di politiche, accordi istituzionali, tecnologie, dati e persone che rendono possibile una condivisione efficiente ed efficace dei dati geo-referenziati.

In questo contesto l'interoperabilità semantica tra i geo-dati e geo-servizi è un requisito necessario per l'interazione a livello tecnologico e concettuale fra sistemi differenti e non omogenei. Nell'ambito del progetto questa problematica è stata risolta estendendo l'implementazione di riferimento del geocatalogo² con uno strato «semantico» che aggiunge conoscenza specifica di dominio migliorandone le prestazioni di searching e matching e abilitando di fatto l'interoperabilità. L'architettura modulare della soluzione individuata ha permesso di estendere la funzione di search prevista dal geo-catalogo con una funzione di ricerca semantica evoluta³ basata sul modello S-Match⁴ (semantic matching) e sull'ontologia di dominio GeoWordNet⁵. Nel modello S-Match, dati due grafi che modellano aspetti diversi delle

¹ P. Shvaiko, A. Ivanyukovich, L. Vaccari, V. Maltese, F. Farazi. A semantic geo-catalogue implementation for a regional SDI. In Proc. of the INPSIRE Conference, 2010.

² La soluzione di riferimento è costituita da GeoNetwork, un'implementazione open-source di un geo-catalogo conforme alla direttiva INSPIRE.

³ F. Farazi, V. Maltese, F. Giunchiglia, A. Ivanyukovich. A faceted ontology for a semantic geo-catalogue. Technical Report DISI-10-061, Department of Information Engineering and Computer Science, University of Trento. In the proceedings of the 8th Extended Semantic Web Conference (ESWC), 2011.

⁴ F. Giunchiglia, P. Shvaiko, M. Yatskevich. S-Match: an algorithm and an implementation of semantic matching. In Proc. of ESWS, 2004.

⁵ F. Giunchiglia, V. Maltese, F. Farazi, B. Dutta. GeoWordNet: a resource for geo-spatial applications. In the Proc. of ESWC, 2010.



informazioni, l'operatore di matching è in grado di fare un confronto ed individuare coppie di nodi simili dal punto di vista semantico individuando così relazione tra i dati non evidenziata dagli approcci tradizionali.

Il progetto è inteso essere funzionale ad un'innovazione di processo che prevede di introdurre cambiamenti strutturali necessari ad una crescita dell'efficienza nella fruizione dei dati e delle informazioni territoriali. Di particolare interesse in quest'ottica il fatto che le meta-informazioni relative ai dati pubblicati nel catalogo sono prodotte durante il processo di creazione del dato stesso a tutto vantaggio della sostenibilità e celerità della pubblicazione.

Da un punto di vista organizzativo il progetto si è sviluppato secondo fasi operative successive:

- si è partiti dalla raccolta dei dati d'interesse trasversale in un unico repository di dati (circa 400 tematismi);

- si sono definiti gli standard di riferimento per la compilazione dei metadati in linea con le direttive del Repertorio Nazionale Dati Territoriali;

- sono state analizzate le possibili licenze per l'apertura al pubblico e si è identificata la Creative Commons 0;

- sono stati identificati i dati candidati ad essere aperti con l'obiettivo di pubblicare il maggior numero di dati possibili a meno di quelli sensibili e/o coperti da qualsivoglia vincolo alla diffusione;

- le strutture provinciali hanno aggiornato i metadati dei propri dati ai fini della pubblicazione;

- il Geocatalogo del Portale Geocartografico Trentino (fino a quel momento visibile solo in intranet) è stato adeguato alle esigenze OGD da Informatica Trentina s.p.a.: (Fase 7) Delibera sugli Open Government Data n. 195 del 17 Febbraio 2012.

Il progetto ha avuto un sviluppo temporale di tre anni tra il 2009 ed il 2011 è stato realizzato a supporto del progetto di "Trasferimento di conoscenze e know-how tra centri di ricerca e imprese anche attraverso la mobilità di ricercatori e tecnici" - Progetto finanziato FSE (2007 - 2013) nell'ambito dell'asse 4 obiettivo: Creazione reti tra università, centri tecnologici di ricerca e mondo produttivo.

Ad oggi il Geocatalogo contiene dati linkati di oltre 161 tematismi tutti raggiungibili attraverso il Portale Cartografico Trentino in modalità aperta. Come detto, per facilitare il riutilizzo dei dati territoriali pubblicati nel geo-catalogo, si è optato per la pubblicazione degli stessi con licenza Creative Commons 0 che di fatto non pone vincoli di nessun tipo alla fruizione degli stessi.

