

Studio

Report servizi qualificati anno 2016

A cura di:

Luca Bonuccelli, Luca Cipriani, Graziano Scaffai

Settore Infrastrutture e Tecnologie per lo sviluppo della Società dell'Informazione
Settore Sistema Informativo di Supporto alle Decisioni. Ufficio Regionale di Statistica

Regione Toscana

Indice

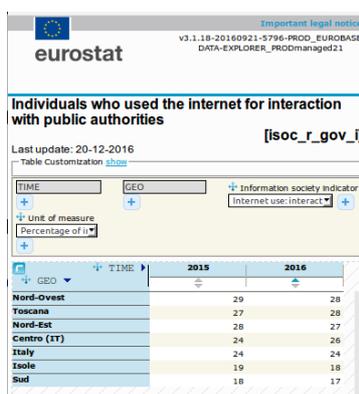
Perchè un report e analisi statistiche?.....	3
Un report a supporto delle azioni di e-gov.....	3
Un report per cambiare la modalità di realizzazione e mantenimento dei servizi della PA.....	3
Un report per ottenere Open data a supporto della trasparenza e di nuove opportunità di crescita.....	4
La sintesi punto per punto.....	4
Analisi di un paniere di servizi.....	4
Serie storiche.....	5
Caratteristiche della popolazione utente.....	6
Distribuzione territoriale degli utenti.....	9
Distribuzione degli utenti per numero di accessi.....	12
Distribuzione settimanale e oraria.....	13
Analisi dei servizi offerti.....	14
Correlazione d'uso dei servizi.....	16
Focus SPID.....	16
Nota metodologica.....	17

Perchè un report e analisi statistiche?

Un report a supporto delle azioni di e-gov

Secondo il DESI (indice di digitalizzazione dell'economia e della società) l'Italia ha fatto parte, nell'anno 2015, del gruppo di Paesi che stanno recuperando il ritardo nella digitalizzazione e sebbene i risultati italiani siano ancora inferiori alla UE nel suo insieme, il Paese sta registrando rapidi progressi.

L'indicatore "Servizi digitali pubblici" riferito all'anno 2015 posiziona l'Italia nella media europea; purtroppo uno dei sottoindicatori pone una criticità: anche se i servizi della pubblica amministrazione esistono, gli utilizzatori reali sono circa la metà rispetto alla media degli altri Paesi.¹



The screenshot shows the Eurostat interface for the indicator "Individuals who used the internet for interaction with public authorities". The table displays data for 2015 and 2016 across various geographical regions. The values are percentages of the population.

GEO	2015	2016
Nord-Ovest	29	28
Toscana	27	28
Nord-Est	28	27
Centro (IT)	24	25
Italia	24	24
Isole	19	18
Sud	18	17

In altre parole *"In Italia i servizi sono fruibili, ma non sono fruiti come nel resto dell'Europa"*.

Secondo l'anticipazione Eurostat (dati al 20/12/2016) il territorio toscano si posiziona sopra la media italiana e sopra la media dei territori del centro Italia, collocandosi alla pari dei territori del nord est.

Nel 2016 il 28% dei cittadini toscani che utilizzano internet ha utilizzato servizi on line per interagire con la pubblica amministrazione (l'indice riferito al centro Italia ha subito un aumento di 2 punti percentuali rispetto all'anno precedente contro-bilanciando, a livello nazionale, il decremento delle altre

macrozone).

Un primo intento di questo studio è pertanto quello di arricchire, attraverso l'elaborazione dei dati già disponibili, la conoscenza utile a monitorare i fenomeni e a supportare le azioni opportune per migliorare ulteriormente questi indicatori.

Un report per cambiare la modalità di realizzazione e mantenimento dei servizi della PA

L'analisi dei comportamenti degli utenti è coerente con il paradigma **"Il cittadino deve essere messo al centro, e tutto deve partire dai suoi bisogni"**

Misurando come l'azione di digitalizzazione stia avvenendo, quali sono i servizi **realmente utilizzati** e, soprattutto, **quale porzione di popolazione non utilizza i servizi on line**, si possono ottenere indicazioni utili su come e dove concentrare eventuali azioni correttive (tecnologiche, formative, nonché orientate alla comunicazione e alla promozione).

¹<https://ec.europa.eu/digital-single-market/en/egovernment-users-desi-indicator-5a1>

Un report per ottenere Open data a supporto della trasparenza e di nuove opportunità di crescita

La strategia che sottintende alla pubblicazione di dati in possesso alla pubblica amministrazione è anche finalizzata a promuovere ricchezza e sviluppo economico, incentivando la crescita di imprese innovative.

Inoltre il recente decreto legislativo “trasparenza” introduce la libertà di accesso di chiunque alle informazioni concernenti l’attività della pubblica amministrazione, anche tramite la pubblicazione dei dati.

In tal senso l’attività di ricerca e analisi che ha prodotto come esito il presente studio, offre come risultato “collaterale” l’individuazione di nuovi dataset da pubblicare quali opendata.

La sintesi punto per punto

Di seguito una sintesi delle principali evidenze statistiche rilevate, in termini descrittivi:

- Nell’anno 2016 si registra, rispetto all’anno precedente, un aumento rilevante degli utenti coinvolti (+32,2%), accompagnato da un incremento (+20,6%) del numero di utenti che utilizzano più frequentemente i servizi offerti.
- Gli utenti della fascia d’età 40-60 sono maggiormente uomini, ma le donne utilizzano più frequentemente i servizi, mentre tra i giovani (under 35) non vi è differenza di genere né per numero di utenti né per frequenza di utilizzo dei servizi
- I servizi on line abbattano i confini: circa il 15% degli accessi è avvenuto da fuori Toscana.
- L’ampiezza demografica e la conformazione del territorio influiscono sull’uso dei servizi: la montagna risulta un po’ svantaggiata.
- I servizi orientati ai professionisti hanno mediamente maggiore frequenza d’uso rispetto a un servizio orientato all’interazione del “normale” cittadino con la PA.
- I servizi on line non hanno orari: quasi un terzo degli accessi avviene di sabato o domenica oppure fuori l’orario 9-18, mentre un accesso ogni venti avviene tra le 20 e le 6 del mattino.
- L’utilizzo di SPID, quale strumento alternativo a CNS e CIE per l’identificazione degli utenti, sta crescendo in modo più che lineare.

Analisi di un paniere di servizi

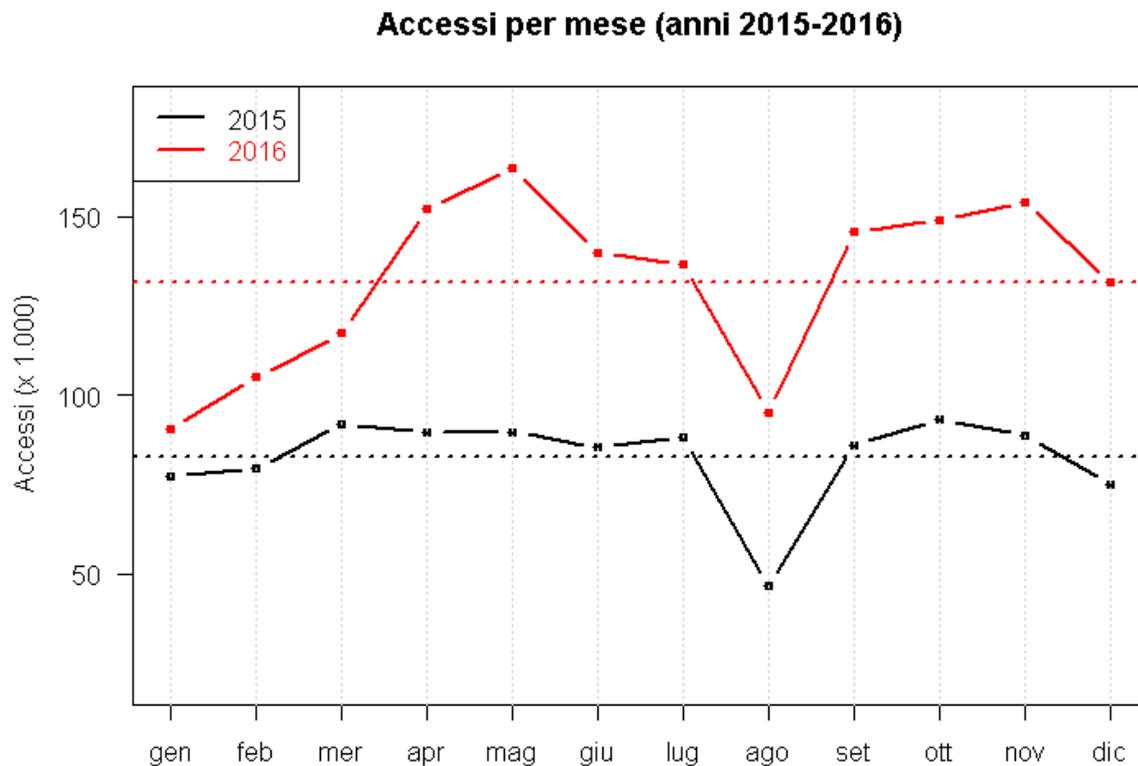
L’analisi è stata svolta per gli anni 2015 e 2016, con riferimento ad un paniere di servizi digitali qualificati (ossia per cui è necessaria l’autenticazione forte dell’utente tramite CNS CIE o, dal 15 marzo 2016, SPID) messi a disposizione da Regione Toscana per il cittadino. Per semplicità, la selezione dei servizi inclusi in tale paniere è avvenuta sulla base dell’infrastruttura tecnologica attraverso la quale vengono erogati (nella fattispecie ARPA -

Infrastruttura per l'autenticazione, autorizzazione e accesso ai servizi di Regione Toscana), tuttavia l'analisi può essere estesa a tutti i servizi indipendentemente.

Le grandezze prese in esame riguardano il numero di **utenti** (o visitatori) distinti che hanno fatto uso di almeno un servizio nel periodo considerato e il numero giorni (indicati come **accessi**) in cui tali utenti hanno utilizzato detti servizi.

Serie storiche

La seguente immagine riporta l'evoluzione mensile degli accessi nei due anni considerati 2015 e 2016:



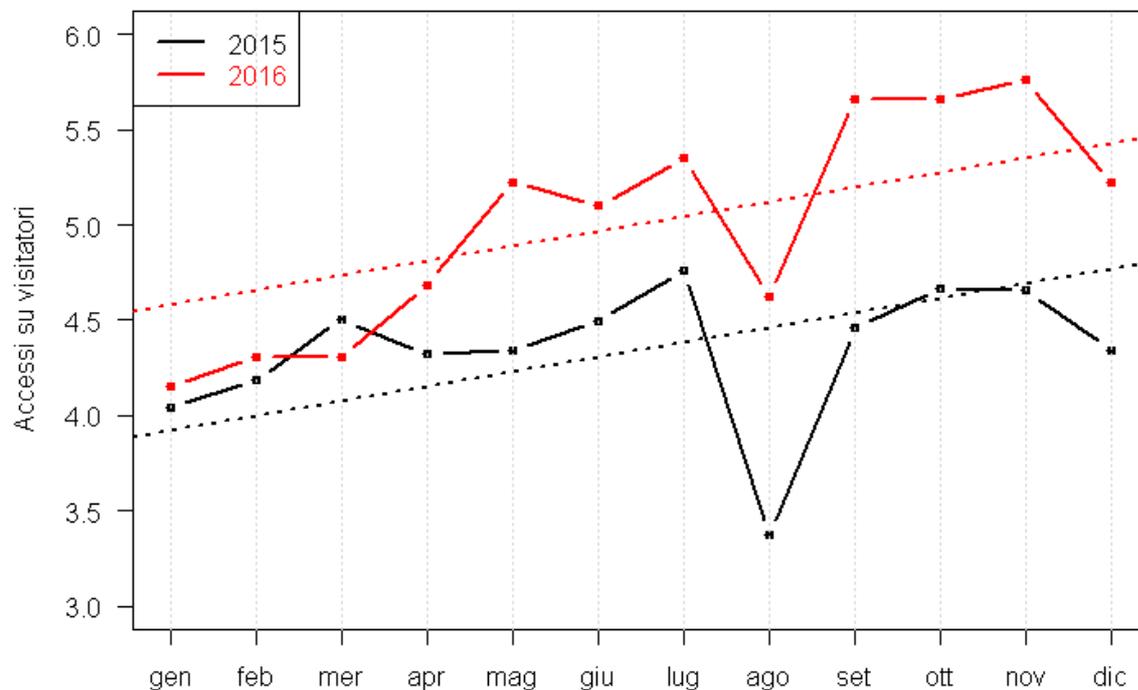
Si osserva un andamento stagionale degli accessi, con un punto di minimo ben marcato nel mese di agosto e si registra tra i due anni una differenza significativa (*“effetto anno”*), evidenziata nel grafico dalla distanza, pari a 49.169 accessi medi mensili, tra le due rette orizzontali (in termini percentuali, **la variazione fra i due anni è stata pari a +59.5%**).

Gli accessi avvenuti nel 2016 sono stati effettuati da 97.440 utenti (73.695 nel 2015) con un **incremento su base annua del 32.2%**.

I dati evidenziano quindi che **la variazione relativa dei visitatori è stata inferiore a quella degli accessi**: se ne desume un maggiore utilizzo dei servizi qualificati da parte degli utenti rispetto all'anno precedente, circostanza che può essere legata sia all'utilizzo da parte degli stessi utenti di **un numero di servizi maggiore**, sia ad una **maggiore frequenza** con la quale gli stessi utenti visitano i servizi di loro interesse.

Esula dagli obiettivi di questa analisi approfondire il peso relativo dei due fattori suddetti sulla spiegazione del fenomeno osservato, fenomeno che d'altronde può essere apprezzato esaminando l'evoluzione temporale del rapporto "Accessi su Visitatori", riportato di seguito:

Accessi su Visitatori unici per mese (anni 2015-2016)



Il rapporto mensile tra accessi e visitatori rappresenta in media il numero di giorni al mese nei quali gli utenti utilizzano almeno un servizio del paniere preso in oggetto:

Il gap tra 2015 e 2016, che misura la distanza media di questo rapporto tra i due anni, **ammonta in percentuale al +15.1%**.

Se gli accessi e i visitatori fossero cresciuti con lo stesso tasso di crescita nel tempo, il rapporto "accessi su visitatori" sarebbe rimasto costante (+0.0%), invece, al netto delle oscillazioni stagionali e dei fattori accidentali, è significativamente cresciuto evidenziando **una maggiore e crescente assiduità d'uso dei servizi**.

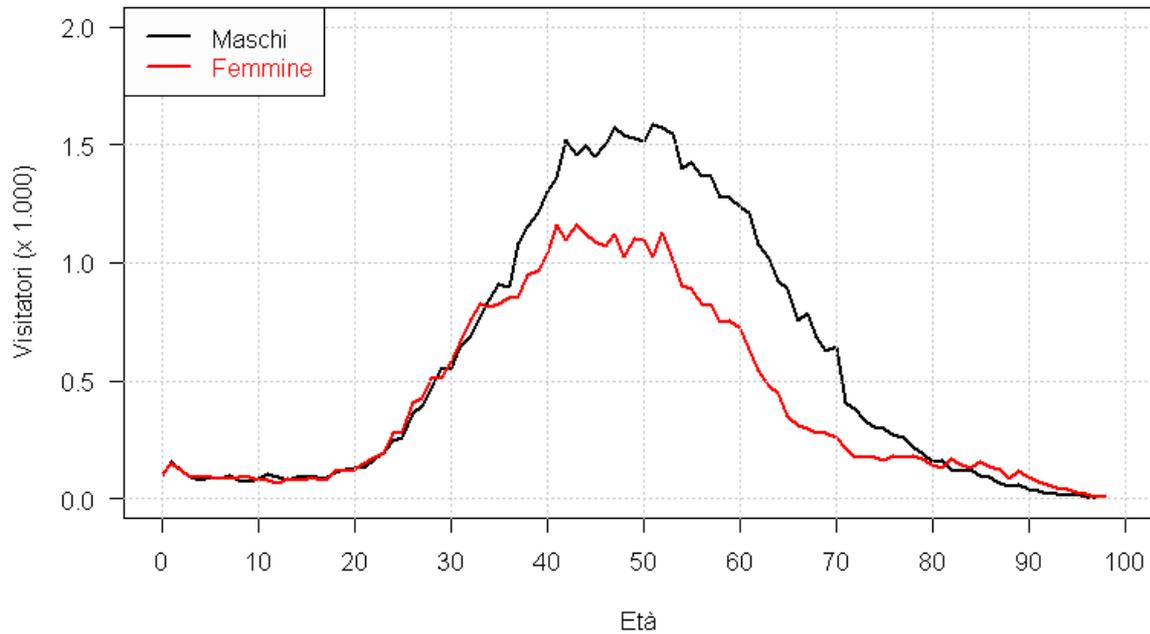
E' interessante osservare tra l'altro che oltre ad un "effetto anno", apprezzabile dalla distanza tra le due rette presenti nel grafico, la pendenza di tali rette mette in evidenza un "effetto mese": **il numero di accessi rapportato ai visitatori non solo è cambiato tra i due anni, ma anche tra i vari mesi all'interno dei due anni** (e tale effetto mese sembra essere analogo nei due anni, essendo analoga la pendenza delle due rette).

Caratteristiche della popolazione utente

L'analisi degli indicatori rispetto alle caratteristiche anagrafiche degli utenti offre spunti interessanti in un'ottica di marketing, così come il punto di vista territoriale.

In figura è rappresentata la distribuzione per sesso ed età degli utenti che hanno effettuato almeno un accesso ad un qualsiasi servizio ricompreso nel paniere nel corso del 2016:

Visitatori unici per età, anno 2016



In termini assoluti, quindi non rapportati alla popolazione, la distribuzione indica nel complesso una **prevalenza di uomini tra gli utenti dei servizi**.

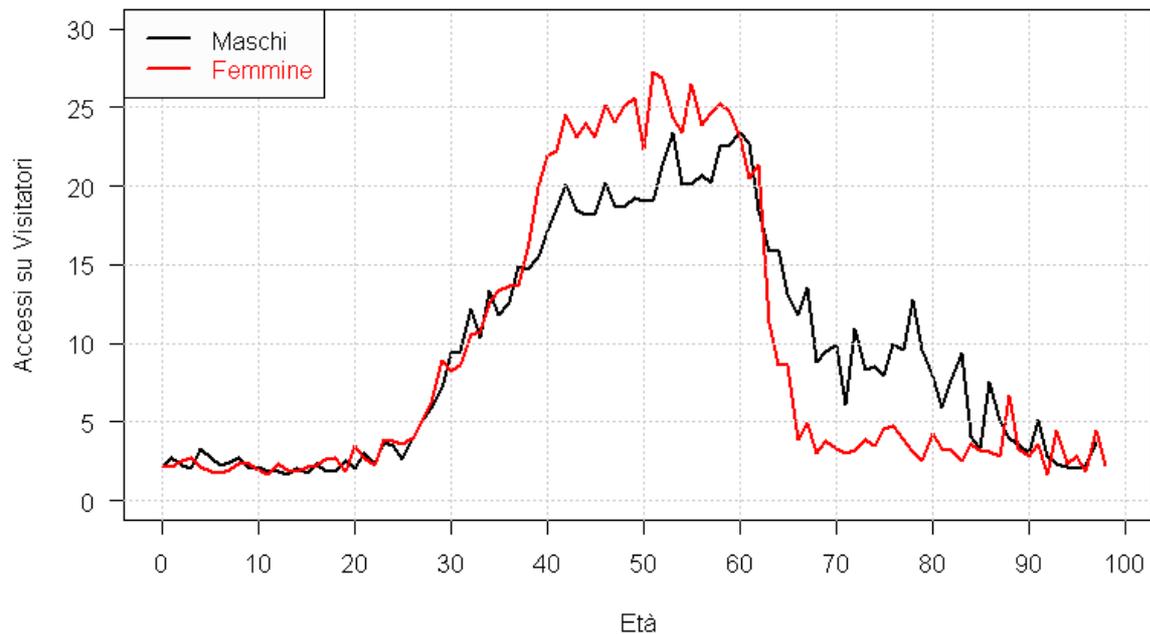
Nella fascia di età che va **da 0 a 35 anni la distribuzione per genere non presenta differenze significative**.

La maggior parte degli utenti che utilizzano i servizi è **concentrata nella fascia di età da 40 a 60 anni**, sia per le femmine che per i maschi.

La forma che la distribuzione assume nella fascia di età che va da 0 a 2 anni (gli accessi riferiti ad identità di soggetti minorenni sono eseguiti da chi esercita la potestà genitoriale) indica come il genitore acceda ai dati dei figli in via telematica principalmente nei primi anni di vita (tale effetto è più marcato per i servizi sociosanitari).

Si riporta ora una rappresentazione grafica dello stesso tipo, costruita però a partire dall'indicatore accessi su visitatori:

Accessi su Visitatori unici per età, anno 2016



Tale rappresentazione mette in evidenza una **maggior tendenza del genere femminile nella fascia di età da 40 a 60 anni** ad utilizzare i servizi del paniere con maggior frequenza rispetto ai maschi.

In altri termini, la popolazione di utenti del paniere di servizi considerato è contraddistinta per la fascia di età 40-60 da una **minor presenza di femmine**, ma da una **maggior assiduità con cui queste accedono ai servizi** offerti.

Questa conclusione non vale ad esempio per la fascia 60-85, nella quale sia la presenza di maschi che la loro frequenza di utilizzo prevalgono su quelle delle femmine.

Possiamo infine apprezzare con una tabella le variazioni che hanno subito tra il 2015 e 2016 gli indicatori di cui si è parlato, raggruppando le età in classi:

Classi età	Variazioni percentuali 2016 / 2015		
	Femmine	Maschi	Totale

VISITATORI			
[0,20]	51,56	43,77	47,52
(20,40]	19,65	33,72	26,50
(40,60]	34,73	29,57	31,62
(60,100]	49,56	33,26	38,79
Totale	32,83	31,78	32,22

ACCESSI SU VISITATORI			
[0,20]	21,54	14,97	18,09
(20,40]	24,90	10,69	17,56
(40,60]	42,48	17,52	27,70
(60,100]	20,55	-4,50	-0,94
Totale	35,70	11,18	20,63

Osservando i totali ed interpretando il rapporto "Accessi su Visitatori" in via approssimata come indicatore di continuità d'utilizzo o di fidelizzazione, si osserva un **significativo aumento (+20,63%) del numero di utenti che utilizzano meno sporadicamente i servizi offerti, accompagnato da un aumento ancor più rilevante degli utenti coinvolti (+32,22%)**.

Tuttavia, mentre la popolazione degli utenti è aumentata nel 2016 di una misura sostanzialmente equivalente tra i due generi, **la frequenza di utilizzo del servizio da parte del genere femminile ha registrato un aumento di oltre il 35% quindi quasi triplo rispetto all'aumento registrato per i maschi (+11,18%)**.

I dati in tabella mostrano inoltre una notevole diversità di comportamento tra gli utenti, sia complessivi che distinti per genere, nell'ambito delle differenti fasce di età.

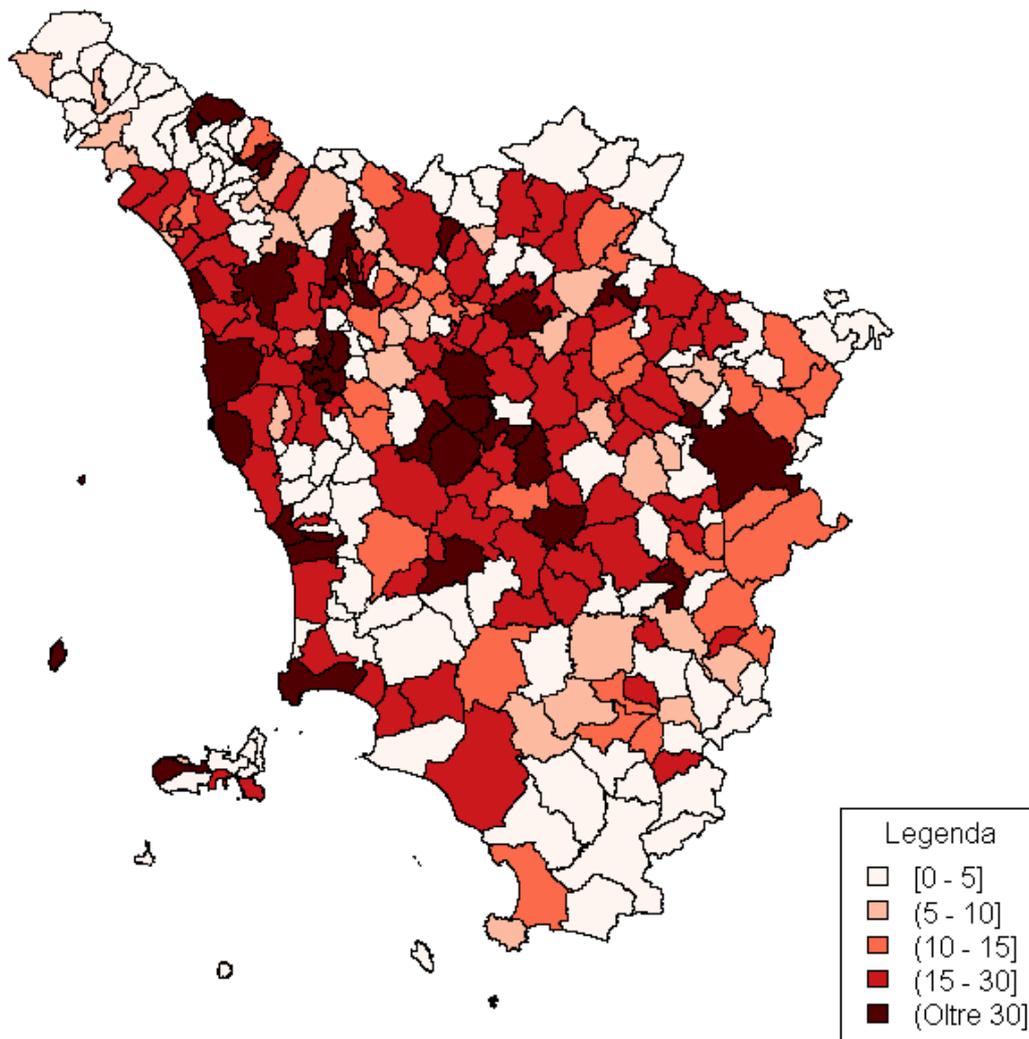
Distribuzione territoriale degli utenti

Si è provveduto a geo-localizzare a livello comunale gli accessi al paniere avvenuti nel 2016, con una percentuale di successo di circa il 60%.

Su tale base (e nell'ipotesi in cui questo campione non statistico di accessi sia rappresentativo della totalità), l'ammontare degli accessi avvenuti da fuori Toscana è quantificabile nel 15% circa del totale (le prime dieci regioni più rappresentate sono in ordine decrescente Lazio, Liguria, Lombardia, Emilia-Romagna, Piemonte, Puglia, Sicilia, Campania, Veneto, Abruzzo).

La distribuzione geografica per Comune degli utenti che hanno eseguito il restante 85% degli accessi, ponderata rispetto alla popolazione residente nel Comune, è espressa nel cartogramma seguente:

- VISITATORI DISTINTI SU 1000 RESIDENTI (COMUNI)-



Ulteriori approfondimenti dimostrano la presenza di alcuni fattori che determinano la distribuzione osservabile dal cartogramma.

Un primo fattore è costituito dalla tipologia del territorio in termini di zone socio-sanitarie, mentre un secondo fattore è costituito dalla zona altimetrica.

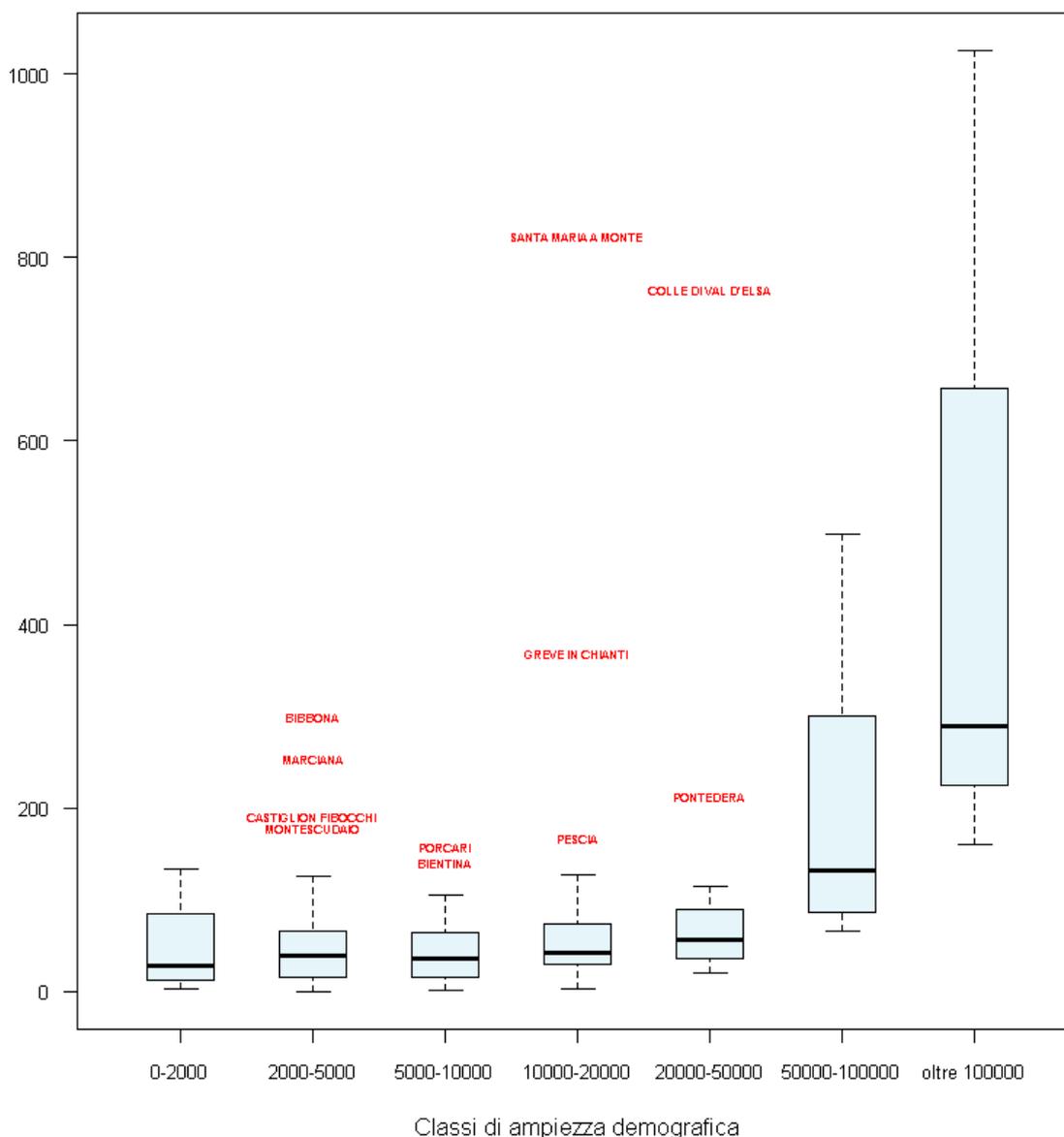
Tra le zone socio-sanitarie, quelle contraddistinte da un valore più alto del rapporto Visitatori su Residenti sono in ordine decrescente quelle di Firenze, dell'Alta Val d'Elsa, Pisana, Livornese, Aretina, della Piana di Lucca, della Val d'Era, Grossetana, Pratese, della Bassa Val di Cecina.

Tenendo conto invece della classificazione del territorio in cinque zone altimetriche (pianura, collina interna, collina litoranea, montagna interna e montagna litoranea), quella contraddistinta da un maggior rapporto tra Visitatori e Residenti è la pianura, mentre l'indicatore assume il valore più basso in corrispondenza della montagna interna (la montagna litoranea e le zone collinari si collocano in mezzo alla graduatoria e non sono tra loro significativamente differenti).

Un altro fattore esplicativo emerge se si considera per ciascun Comune il rapporto degli accessi rispetto alla popolazione residente.

Si nota che nei Comuni con ampiezza demografica più bassa, tale dimensione demografica non produce effetti significativi sui valori che assume l'indicatore e sulla loro distribuzione, mentre **per i Comuni più grandi l'effetto demografico diviene assai rilevante**, come evidenziano i seguenti box-plot:

Distribuzione degli accessi su 1000 residenti per classi di ampiezza demografica dei comuni



Come mostra il grafico, **al crescere della dimensione demografica del Comune, in particolare sopra a 20.000 abitanti, cresce il tasso di accesso e si può dimostrare come tale crescita non sia accidentale, ma statisticamente significativa.**

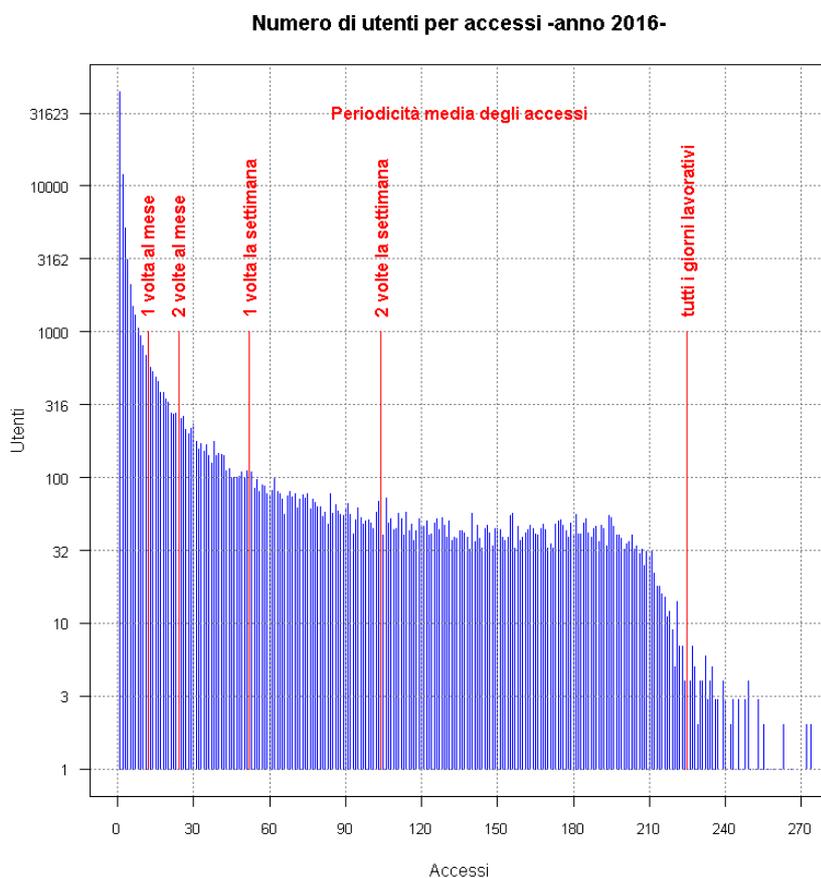
I Comuni più grandi (dal punto di vista demografico) sono dunque caratterizzati da un tasso di accesso molto più alto rispetto agli altri: una possibile spiegazione di questo fenomeno, corroborata dalle analisi presentate di seguito, è legata alla caratterizzazione dei servizi in termini di utenza obiettivo (professionale, non professionale) poiché alla dimensione del Comune corrisponde normalmente una maggiore presenza di strutture che offrono servizi professionali.

Nel diagramma a box-plot sono inoltre evidenziati 11 Comuni per i quali si registra un valore anomalo, rispetto alla classe di ampiezza di cui fanno parte, del rapporto accessi su residenti: tali circostanze possono essere legate a specifiche condizioni che hanno favorito in tali Comuni un maggior accesso ai servizi (si pensi ad esempio a iniziative locali o sperimentazioni).

Distribuzione degli utenti per numero di accessi

La distribuzione degli utenti per numero di accessi, calcolata a livello aggregato sull'intero paniere, ci informa di come si caratterizzano i servizi ricompresi nel paniere in termini di frequenza con cui vengono utilizzati.

Occorre tener conto che tale frequenza non rappresenta necessariamente fidelizzazione, in quanto anche correlata alla eterogeneità dei servizi presenti nel paniere, con particolare riferimento alla tipologia di utenti a cui sono rivolti (i servizi orientati ai professionisti sono caratterizzati ad esempio per loro natura da una maggiore frequenza d'uso rispetto agli altri). Tenendo conto di queste premesse, per l'anno 2016 si osserva la seguente distribuzione:

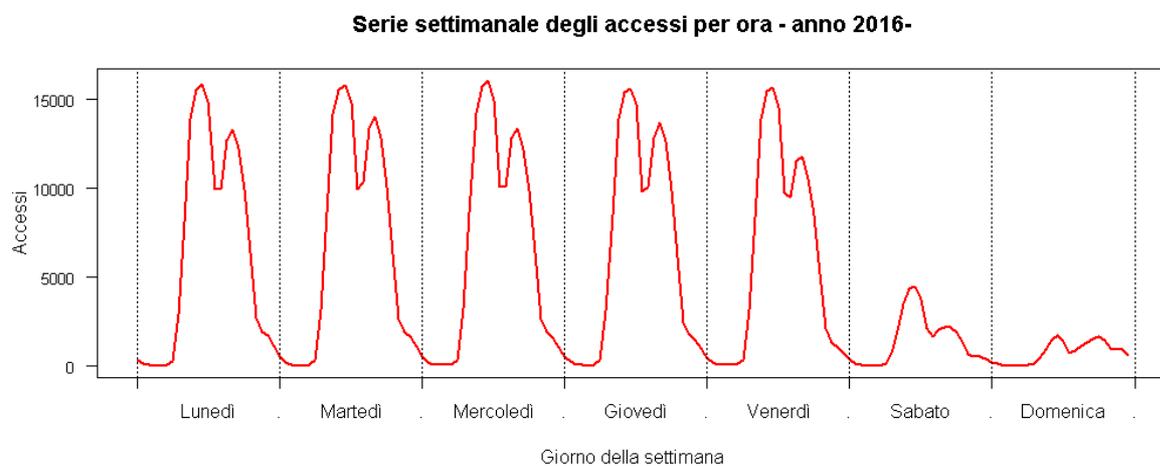


Calcolando il numero di utenti per classi di frequenza di accesso, si osserva che quasi la metà (48.75%) ha acceduto esattamente 1 volta in tutto il 2016, circa il 5% da una a due volte al mese e meno del 4% da una a due volte alla settimana.

Dal grafico emerge inoltre un accenno di moda intorno a 190-200 giorni: questo evidenzia come assieme a molti utenti occasionali, responsabili del generale andamento decrescente, vi siano utenti "sistematici" che accedono quasi tutti i giorni: un approfondimento dell'analisi per tipologia di servizio potrebbe confermare l'ipotesi che i servizi caratterizzati da questa sistematicità siano proprio quelli orientati ai professionisti, come premesso inizialmente.

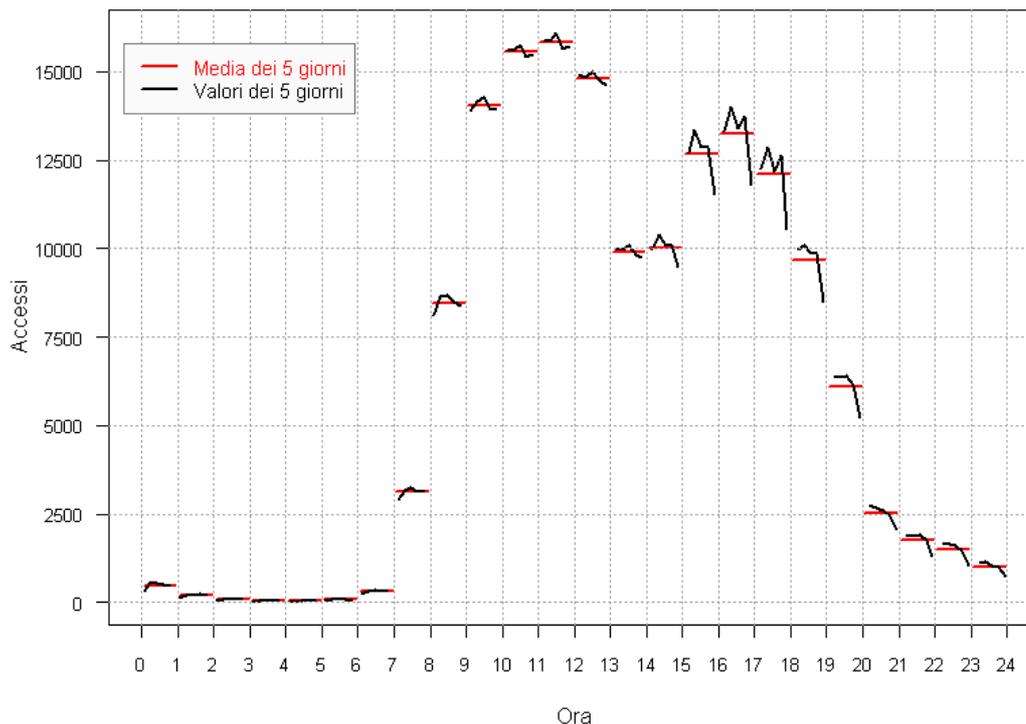
Distribuzione settimanale e oraria

Ad esclusione del sabato e della domenica, l'andamento degli accessi è molto simile nei diversi giorni della settimana, evidenziando tra l'altro un picco durante la mattina dalle 9 alle 11 e uno nel pomeriggio intorno alle 16:



Il seguente grafico, che mette a confronto i giorni dal lunedì al venerdì per le diverse fasce orarie di accesso, evidenzia come il comportamento degli utenti sia poco variabile tra i vari giorni per le fasce orarie antimeridiane e sia invece caratterizzato da maggiore variabilità nelle fasce orarie pomeridiane e serali, soprattutto a causa di una marcata differenza di comportamento nel venerdì:

Accessi nelle singole ore dei giorni: lunedì-venerdì, anno 2016



Si evince dai due grafici che i servizi on line vengono acceduti in misura rilevante (il 28%) anche nei giorni e nelle fasce orarie di chiusura degli sportelli fisici (sabato, domenica e negli altri giorni dalle 18 alle 9 del mattino successivo) e che una quota del 5,7% è costituita da accessi addirittura notturni (dalle 20 alle 7 del mattino successivo).

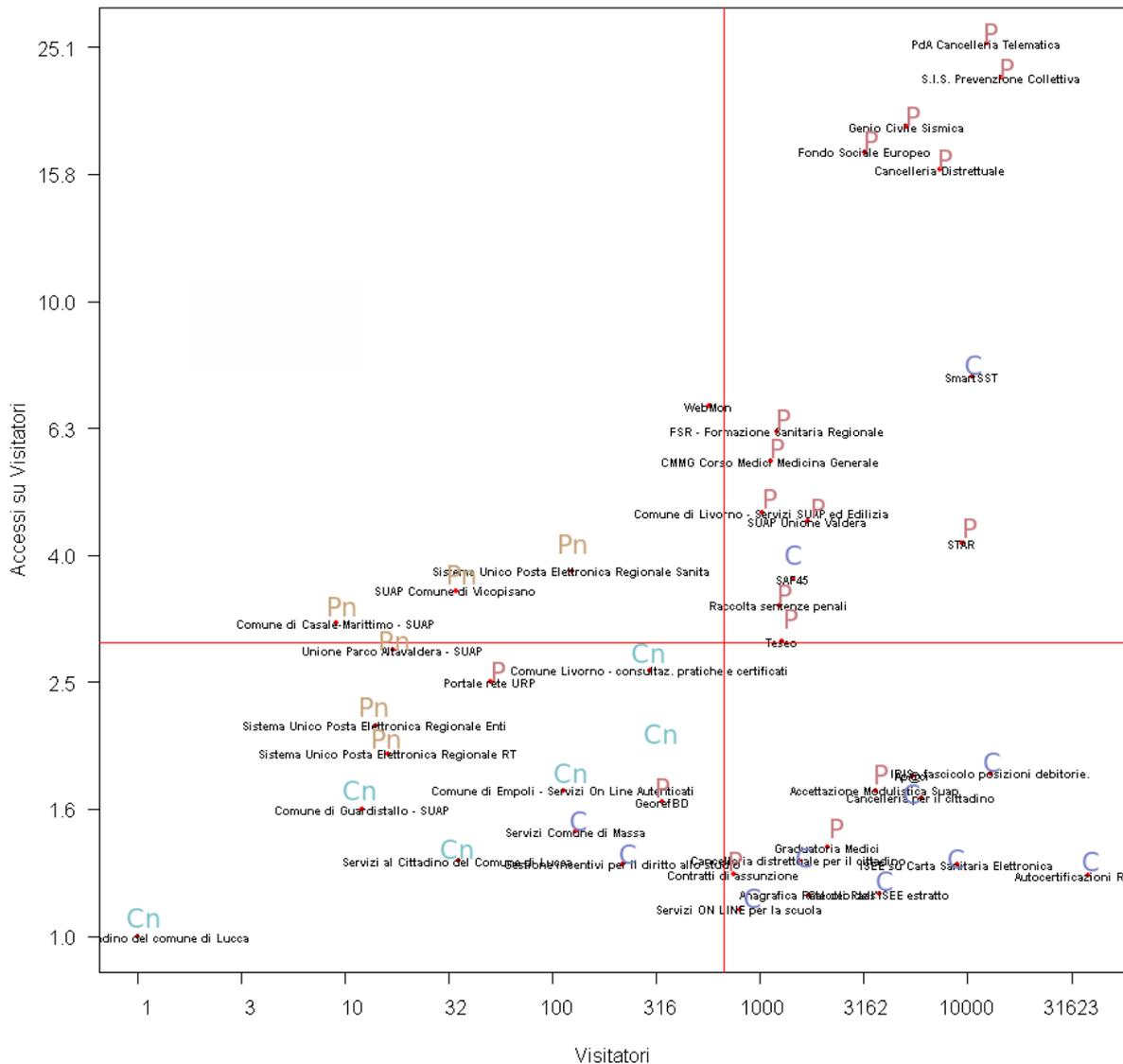
Analisi dei servizi offerti

Dopo aver analizzato i dati di accesso al paniere dei servizi nel loro complesso, si esaminano adesso i singoli servizi, con l'obiettivo di individuare tra loro gruppi omogenei e correlazioni d'uso.

Gli indicatori rispetto ai quali è risultato più significativo effettuare le analisi sono ancora il numero di utenti (o visitatori) e il rapporto tra accessi e visitatori.

Il grafico posiziona ciascuno dei servizi che compongono il paniere in funzione del numero di visitatori e del rapporto accessi su visitatori che lo caratterizza, con riferimento a tutto l'anno 2016:

Visitatori e accessi su visitatori alle risorse - anno 2016



I servizi sono stati categorizzati in servizi orientati principalmente ai cittadini (indicati con la lettera C) e servizi orientati principalmente a professionisti (indicati con la lettera P).

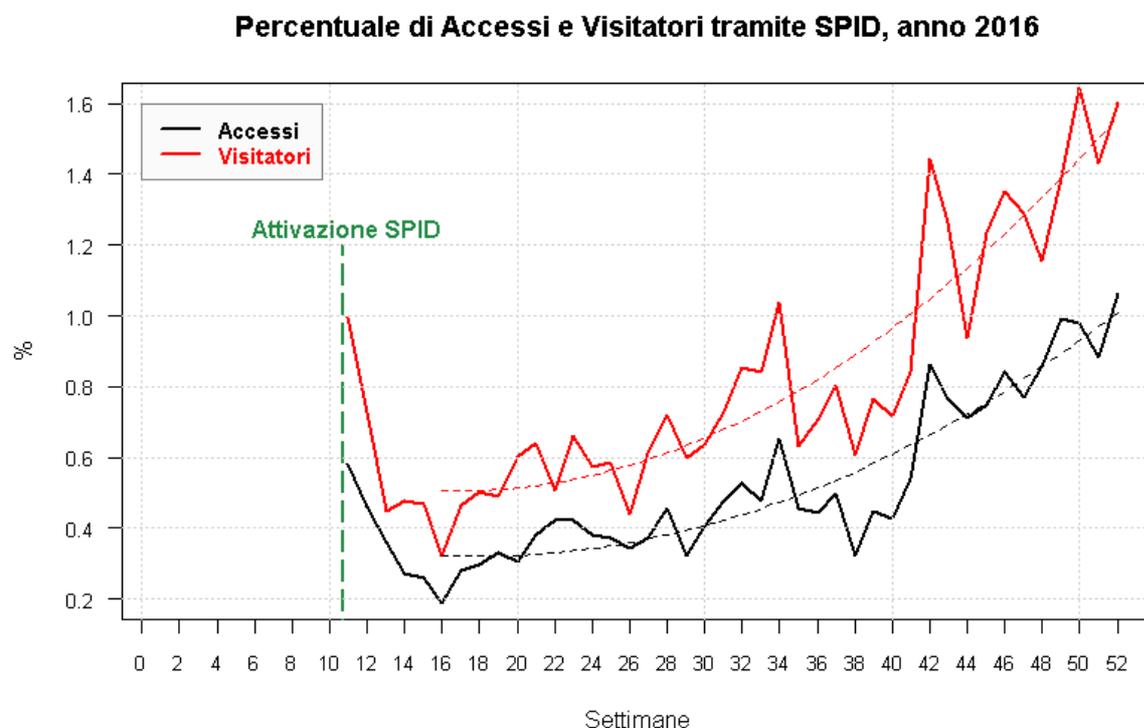
Ai servizi di recente rilascio è stata aggiunta la lettera n.

I dati evidenziano come i servizi per i professionisti (sia di recente rilascio che non) siano utilizzati molto spesso (proprio perchè il servizio diviene parte della suite di lavoro quotidiano per l'utente), mentre i servizi per il cittadino, pur interessando un'ampia fetta della popolazione sono per loro stessa natura utilizzati sporadicamente in funzione delle singole esigenze.

Riprendendo un'ipotesi alla quale eravamo giunti con l'analisi territoriale, l'impatto del fattore demografico sul tasso di accesso ai servizi può essere determinato quindi dalla più alta densità d'uso "professionale" nei Comuni con maggiore popolazione e maggior presenza di attività economiche.

sviluppare nuovi servizi o nuove modalità di erogazione degli stessi basati sulle peculiarità della neutralità tecnologica e della facilità di utilizzo delle credenziali.

Il seguente grafico rappresenta per ciascuna settimana del 2016 la percentuale dei visitatori e degli accessi tramite SPID rispetto al totale di visitatori e di accessi:



Anche se queste percentuali sono al momento basse, al netto del picco nelle prime settimane di attivazione, l'evoluzione sembra indicare una crescita più che lineare e maggiormente significativa per i visitatori che per gli accessi.

Nota metodologica

Il paniere di servizi è stato individuato a partire da quelli disponibili per il cittadino nell'infrastruttura per l'autenticazione, autorizzazione e accesso ai servizi di Regione Toscana (ARPA). La fonte dati è pertanto costituita dai log di accesso ai singoli servizi dell'infrastruttura, anonimizzati a meno del sesso e dell'anno di nascita dell'utente e georeferenziati per Comune, a partire dall'indirizzo ip che ha effettuato l'accesso (la georeferenziazione automatica è avvenuta, come riportato nel testo, per circa il 60% degli accessi).

E' possibile e di indubbio interesse futuro ampliare le analisi a tutti i servizi offerti al cittadino, indipendentemente dall'infrastruttura di sicurezza che utilizzano.

L'indicatore "utenti", nel testo chiamato anche "visitatori" o "visitatori distinti", è calcolato come conteggio distinto dei codici fiscali anonimizzati che hanno eseguito almeno un accesso nell'unità di tempo e luogo considerata in ciascuna analisi.

L'indicatore "accessi" è invece ottenuto conteggiando per ciascun utente il numero di giorni nei quali tale utente ha eseguito almeno una visita a servizi disponibili nel paniere (gli accessi non sono condizionati pertanto dal numero di "clic" o dal numero di volte con cui in un giorno l'utente ha visitato i servizi).

Le analisi sono state svolte in un'ottica prevalentemente esplorativa con l'obiettivo di "far parlare i dati". Per questo scopo abbiamo fatto largo uso di grafici statistici che, nel mostrare i dati individuali, consentano l'emersione di eventuali "patterns" dai quali cogliere strutture sottostanti e in tal modo fornire una informazione sulla struttura. Laddove queste esplorazioni hanno evidenziato particolari differenze abbiamo applicato anche qualche tecnica "confermativa" effettuando test statistici per verificare se tali differenze erano compatibili con fattori del tutto accidentali o meno. Verifiche più puntuali degli eventuali fattori esplicativi degli indicatori, richiedono la considerazione congiunta del set di fattori che si assumono come esplicativi, ma questa impostazione esulava dall'obiettivo di questo rapporto.

Riteniamo che i risultati ottenuti consentano anche di individuare aspetti da approfondire. Ad esempio, osserviamo un rapporto accessi/popolazione elevato nelle grandi città e modesto nei piccoli comuni, in particolare della montagna, e questo potrebbe dipendere non tanto dal volume della popolazione quanto dalla presenza nei grandi centri urbani di più alti livelli di istruzione e professionali. Altro aspetto è l'andamento dell'utilizzo che, pur presente in tutto l'arco della giornata, sembra seguire in prevalenza l'orario di ufficio: anche questo potrebbe essere connesso a usi professionali dei servizi. Alcune evidenze propongono inoltre approfondimenti puntuali (ad esempio, l'elevato numero di accessi in alcuni Comuni, quali Santa Maria a Monte, Colle Val d'Elsa, Bibbona, ...).

Infine, un'analisi che abbia l'obiettivo di individuare i fattori che favoriscono l'accesso richiede non solo l'utilizzo congiunto di tutte le informazioni recuperabili dai log di accesso ai servizi, ma anche di quelle relative alla distribuzione della popolazione per territorio, per utilizzo delle tecnologie, per titolo di studio e professione, che probabilmente sono caratteristiche non trascurabili per spiegare i comportamenti nei confronti degli accessi i servizi.

Dal punto di vista tecnico, i dati sono stati raccolti, resi anonimi e geolocalizzati con Logstash, collezionati in Elasticsearch (vers. 5) e successivamente elaborati in ambiente Hadoop (distribuzione Big Top vers. 1.1.0) attraverso l'utilizzo di Hive.

La fase di geolocalizzazione ha utilizzato dati GeoLite2 creati da Maxmind. Le elaborazioni statistiche e le rappresentazioni grafiche sono state ottenute con il software R (vers. 3.2.5).