

Studio Report servizi qualificati Aggiornamento al primo semestre 2017

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

La necessità di un aggiornamento.....	3
La sintesi punto per punto.....	3
Analisi di un paniere di servizi.....	4
Serie storiche.....	4
Caratteristiche della popolazione utente.....	6
Distribuzione territoriale degli utenti.....	8
Distribuzione degli utenti per numero di accessi.....	10
Distribuzione settimanale e oraria.....	11
Analisi dei servizi offerti.....	12
Focus SPID.....	13
Nota metodologica.....	15

La necessità di un aggiornamento

“*La potenza è nulla senza controllo*”, recitava un famoso spot di qualche anno fa e lo slogan potrebbe ben adattarsi al mondo della digitalizzazione della pubblica amministrazione.

I servizi on-line della pubblica amministrazione sono sempre più numerosi e l'unione europea ritiene che siano in linea come qualità e quantità a quelli della media dei paesi membri. Purtroppo la percentuale di utenti che hanno uso ad internet e che utilizzano i servizi on-line della pubblica amministrazione ha subito un brusco calo.

Gran parte delle attività che riguardano il monitoring, che si tratti di tool automatici o azioni di persone fisiche, rimane nascosta. È un servizio invisibile e sempre in moto che silenziosamente opera per garantire il funzionamento ottimale del singolo servizio.

Però occorre osservare il quadro di insieme e veicolare le informazioni in tempi utili.

Quindi si è ritenuto necessario verificare se (e quanto) le conclusioni a cui i dati ci avevano portato 6 mesi fa sono ancora valide.

Nota: In questo aggiornamento sono evidenziate i tratti più salienti rispetto al report precedentemente pubblicato con sfondo grigio

Il report 2016 è disponibile all'url <http://www.regione.toscana.it/-/report-servizi-qualificati>

La sintesi punto per punto

Di seguito una sintesi delle principali evidenze statistiche rilevate, in termini descrittivi, per il primo semestre 2017:

- continuano ad aumentare sia il numero di utenti che il numero di accessi ai servizi digitali offerti, ma con un tasso di crescita inferiore;
- complessivamente tende a diminuire la frequenza con la quale gli utenti accedono ai servizi, misurata dal rapporto mensile tra numero di accessi e numero di visitatori;
- l'utilizzo di SPID, quale strumento alternativo a CNS e CIE per l'identificazione degli utenti, ha registrato una iniziale fase di crescita più che lineare, seguita nell'ultimo periodo da una fase di stabilizzazione

Si confermano inoltre alcune tendenze già emerse nell'anno precedente:

- 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
- 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.

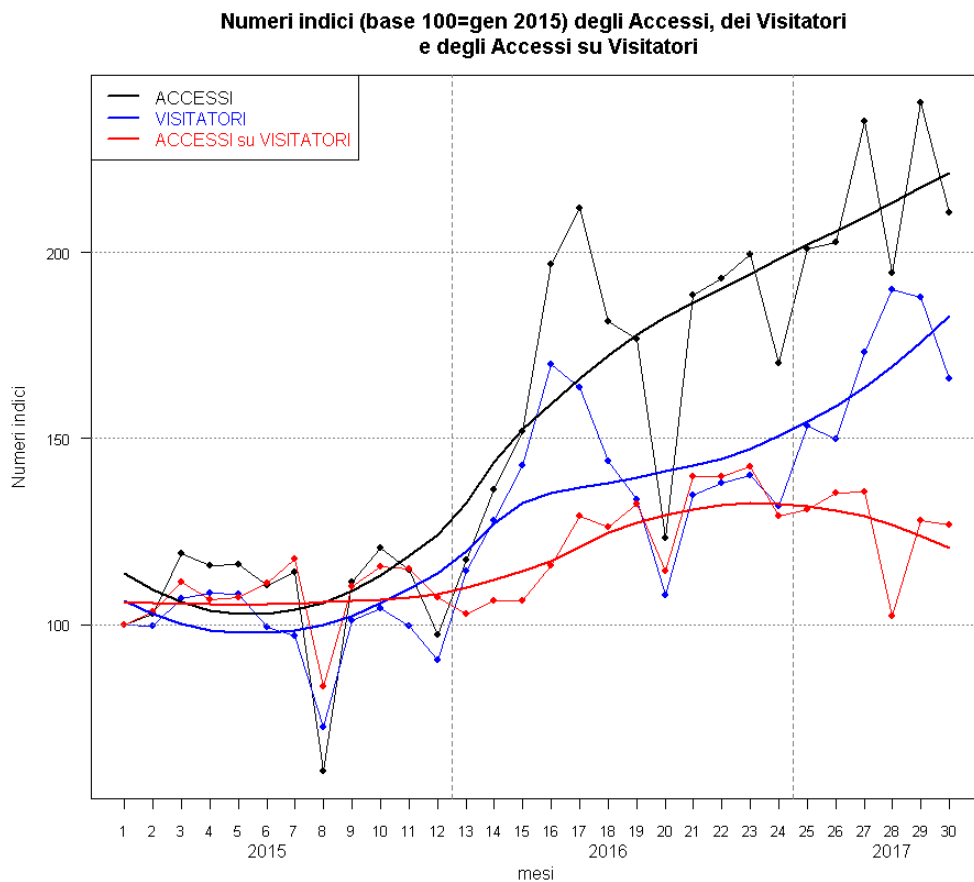
Analisi di un paniere di servizi

L'analisi è stata svolta per gli anni 2015, 2016 e primo semestre 2017 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 dall'infrastruttura utilizzata.

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 di giorni (indicati come **accessi**) in cui tali utenti hanno utilizzato detti servizi.

Serie storiche

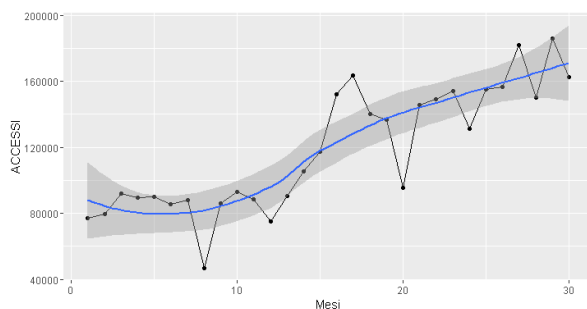
Per il periodo considerato, il grafico seguente riporta l'evoluzione mensile degli accessi, dei visitatori e del rapporto accessi su visitatori:



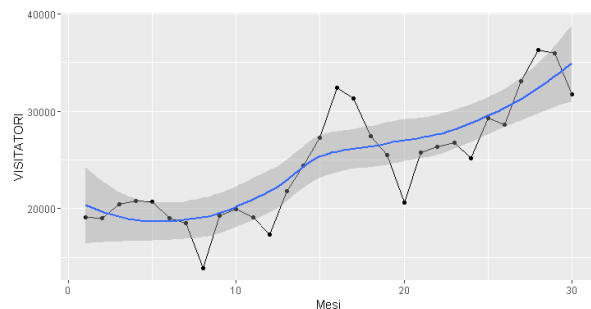
Per i 30 mesi presi in esame (gennaio 2015 – giugno 2017), vengono rappresentate sul grafico le tre serie storiche di numeri indici assumendo come base=100 il primo mese disponibile e interpolando ciascuna serie al fine di coglierne l'andamento complessivo (si è usato a tal fine un metodo di regressione locale).

Si osserva complessivamente un trend in crescita sia del numero di accessi che di visitatori, con il primo semestre 2017 caratterizzato da una crescita dei visitatori superiore a quella degli accessi (il rapporto accessi su visitatori tende infatti a diminuire): possiamo concludere qualcosa sul primo semestre 2017 tenendo conto al netto delle oscillazioni stagionali o accidentali che possono caratterizzare questi andamenti?

Un aiuto può provenire dalla rappresentazione delle bande di confidenza al 95% attorno alle tre curve interpolanti (di seguito le prime due, relative rispettivamente alla serie degli accessi e a quella dei visitatori):



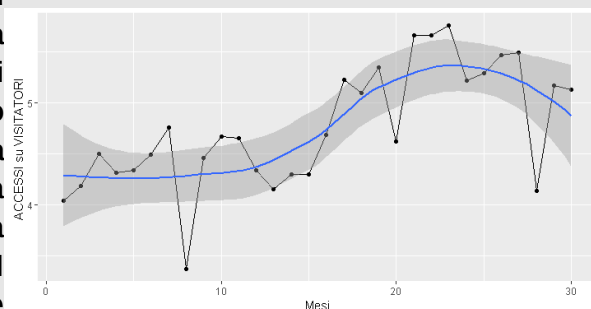
Stime della evoluzione degli accessi con bande di confidenza



Stime della evoluzione dei visitatori con bande di confidenza

Dall'osservazione di questi due grafici si desume una alta probabilità che, al netto di circostanze di natura accidentale, il trend al 30° mese (giugno 2017) sia superiore al valore dello stesso mese dell'anno precedente, sia riguardo agli accessi che ai visitatori.

Per quanto riguarda il rapporto accessi su visitatori, rappresentato di fianco, la valutazione differisce rispetto ai grafici precedenti poiché, tenendo conto dell'incertezza dovuta all'ampiezza della banda di confidenza, non vi è una sufficiente evidenza per affermare che la diminuzione del rapporto registrata nel recente periodo sia prova di una reale inversione del trend: tale diminuzione potrebbe essere del tutto accidentale, condizionata dal forte minimo registrato nel mese di aprile 2017.



Stime della evoluzione del rapporto accessi - visitatori con bande di confidenza

Con riferimento al picco negativo anomalo di aprile 2017, emerge da ulteriori analisi che la riduzione del rapporto accessi su visitatori coinvolge quasi tutti i servizi: solo 10 dei 56 servizi considerati (tra l'altro di peso modesto in termini di accessi) non

sono caratterizzati da questo andamento: un possibile fattore esplicativo di tale andamento anomalo è costituito dal minor numero di giorni lavorativi che ha caratterizzato il mese di aprile rispetto ai mesi contigui, per effetto del ponte Pasqua – 1° Maggio.

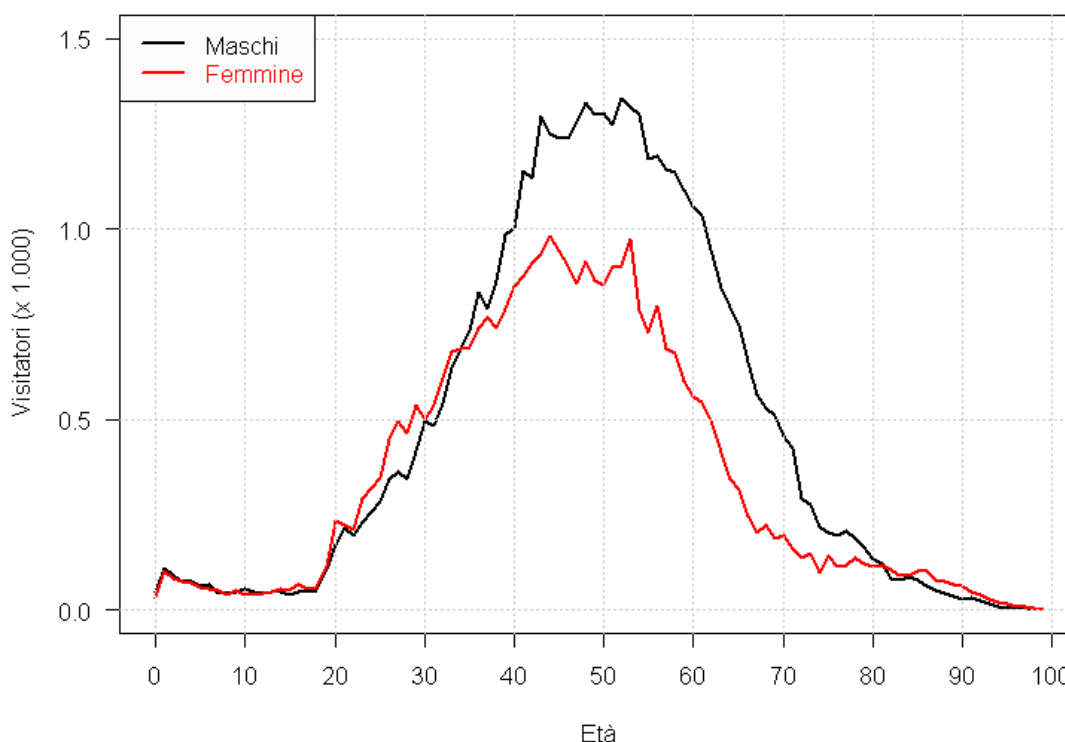
L'osservazione dell'andamento delle tre serie nel corso dei prossimi mesi fornirà ovviamente ulteriori elementi a riprova o meno delle ipotesi suddette.

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.

Nel grafico 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 primo semestre 2017:

Visitatori unici per età, anno 2017 (gen.-giu.)



Rispetto ai dati osservati per l'intero anno 2016 non sono emerse differenze significative.

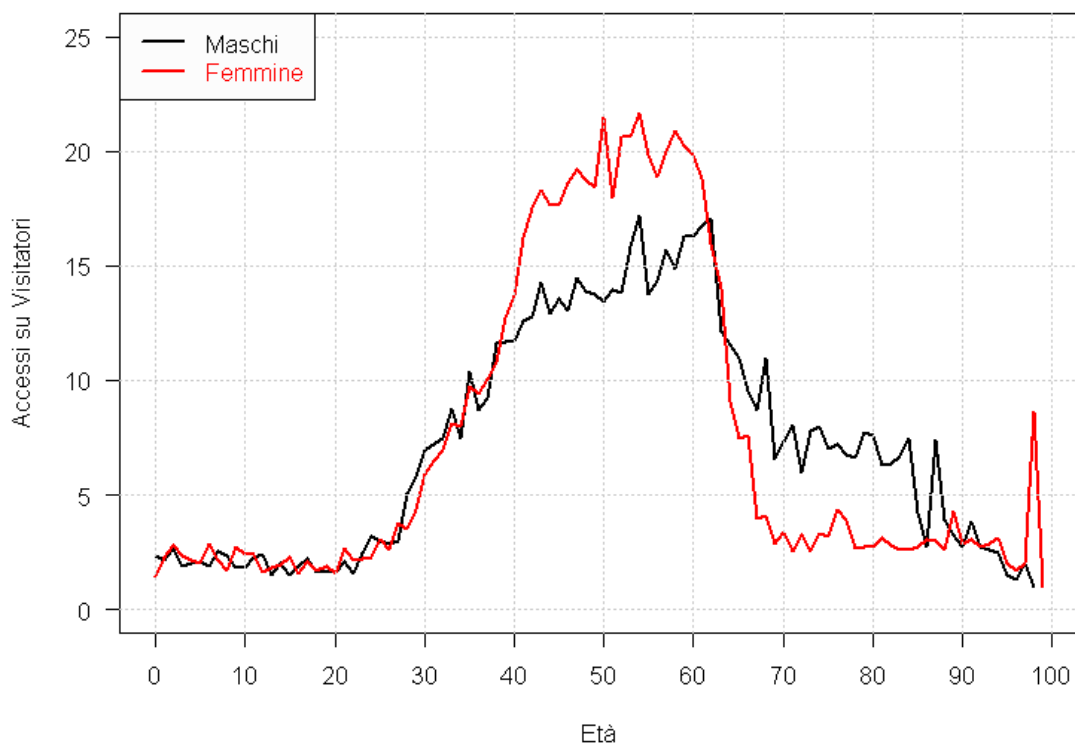
Anche per il primo semestre 2017 in termini assoluti, quindi non rapportati alla popolazione, la distribuzione indica nel complesso una prevalenza di uomini tra gli utenti dei servizi. Nella fascia di età che va **da 0 a 35 anni la distribuzione per genere non presenta differenze significative.**

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 2017 (gen.-giu.)



Tale rappresentazione mette in evidenza, analogamente a quella ottenibile per l'intero anno 2016, 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.

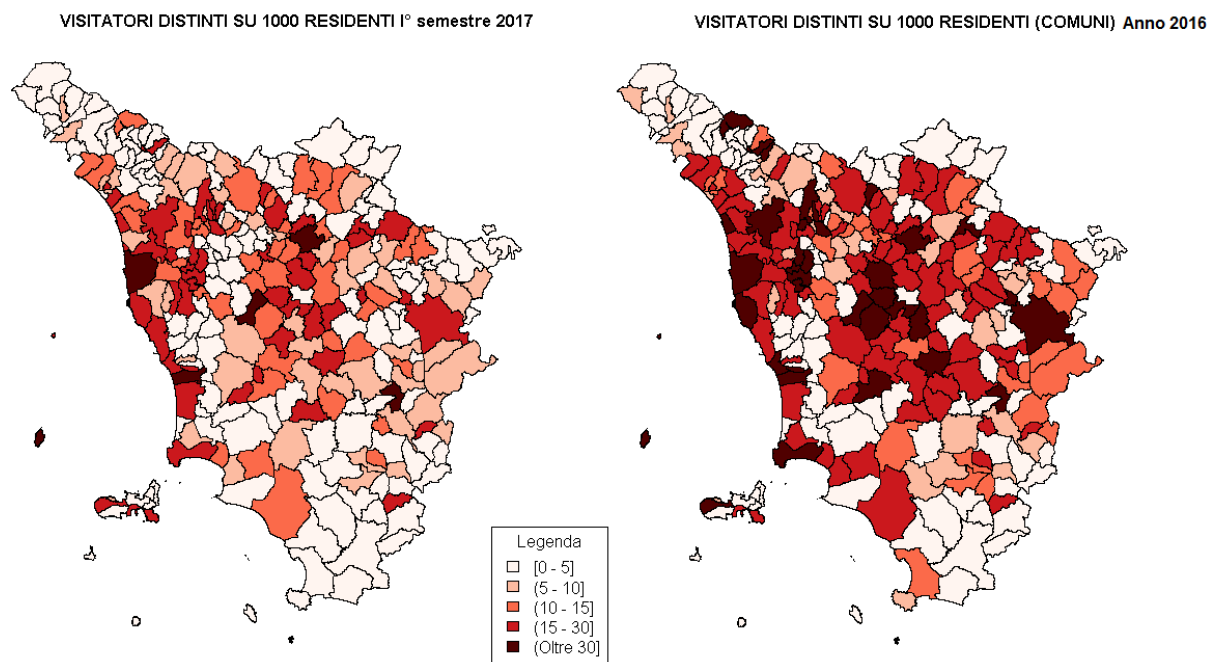
In altri termini, la popolazione di utenti del paniere di servizi considerato è contraddistinta per la fascia di età 40-60 da una **minor presenza di femmine**, ma da una **maggior assiduità con cui queste accedono ai servizi** offerti.

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.

Distribuzione territoriale degli utenti

Si è provveduto a geo-localizzare a livello comunale gli accessi al paniere avvenuti sia nell'anno 2016 che nel primo semestre 2017, 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), la distribuzione geografica per Comune degli utenti che hanno eseguito gli accessi dal territorio toscano, ponderata rispetto alla popolazione residente nel Comune, è espressa per ciascun anno nei seguenti cartogrammi:



Anche le analisi territoriali, analogamente a quelle per età e sesso, non mettono in evidenza differenze di distribuzione significative, mentre la differente intensità della colorazione è dovuta verosimilmente al fatto che si sta considerando per il 2017 il solo primo semestre.

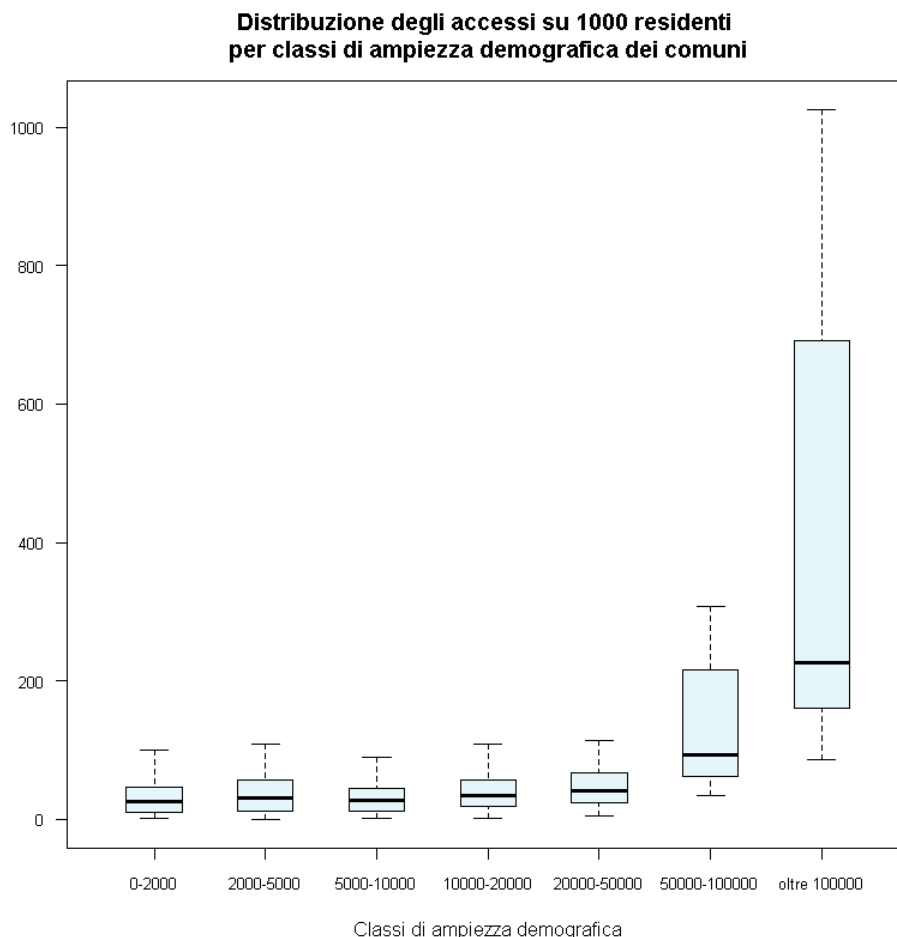
Si ricorda che un primo fattore esplicativo delle distribuzioni osservabili dai cartogrammi è costituito dalla tipologia del territorio in termini di zone socio-sanitarie, mentre un secondo fattore è costituito dalla zona altimetrica.

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, anche per il primo semestre 2017, 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:



Come mostra il grafico, **al crescere della dimensione demografica del Comune, in particolare sopra a 50.000 abitanti, cresce il tasso di accesso** e si può dimostrare come tale crescita non sia accidentale, ma statisticamente significativa

I Comuni più grandi (le città medio grandi della Toscana) (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 è nelle città che sono presenti strutture che offrono servizi professionali.

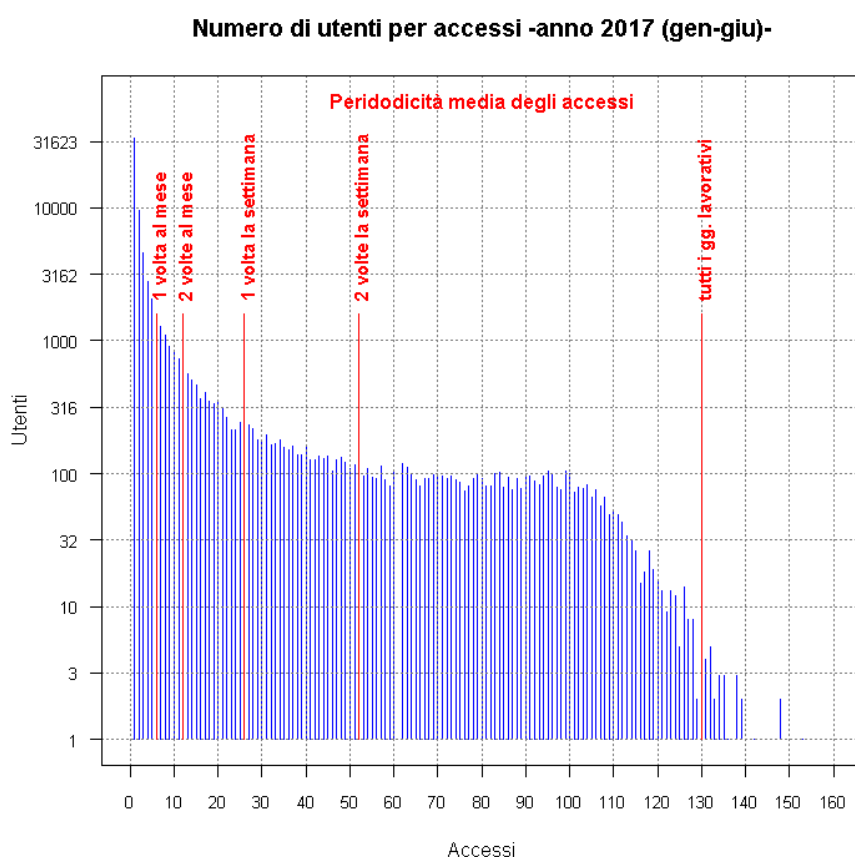
Si segnala infine che, a differenza di quanto registrato per l'anno 2016, non emergono nel primo semestre 2017 Comuni per i quali si registri una quantità di accessi anomalamente grande rispetto a quella che caratterizza la classe di ampiezza demografica di cui tali

Comuni fanno parte: questa evidenza conferma l'ipotesi che tali anomalie fossero dovute a fattori contingenti quali ad esempio sperimentazioni locali.

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.

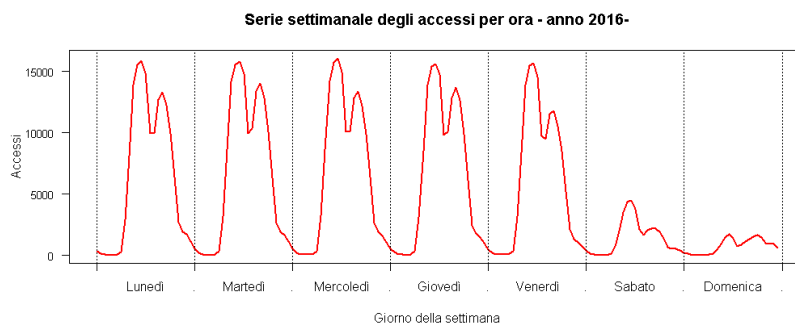
Occorre tener conto che tale frequenza non rappresenta necessariamente fidelizzazione, in quanto anche correlata alla eterogeneità dei servizi presenti nel paniere, con particolare riferimento alla tipologia di utenti a cui sono rivolti (i servizi orientati ai professionisti sono caratterizzati ad esempio per loro natura da una maggiore frequenza d'uso rispetto agli altri). Tenendo conto di queste premesse, si osserva per il primo semestre 2017 la seguente distribuzione:



Anche per questo semestre, come risultava per l'intero anno 2016, si evince come assieme a molti utenti occasionali, responsabili del generale andamento decrescente, vi siano utenti che accedono sistematicamente: un approfondimento dell'analisi per tipologia di servizio potrebbe confermare l'ipotesi che i servizi caratterizzati da questa sistematicità siano in prevalenza quelli orientati ai professionisti.

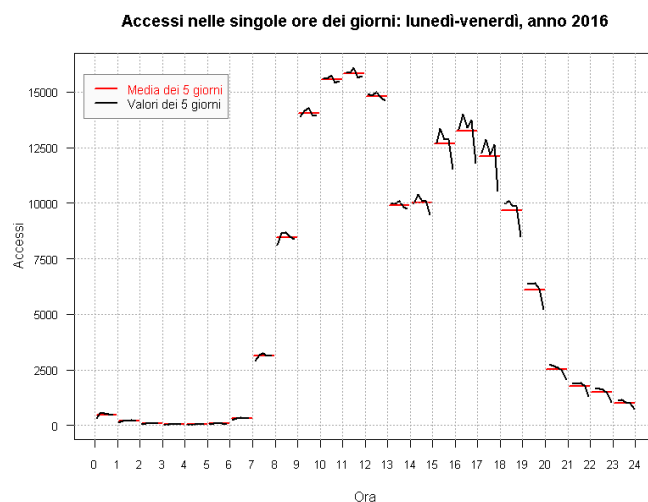
Distribuzione settimanale e oraria

Si conferma per il primo semestre 2017 un andamento nelle ore del giorno pressoché identico a quello registrato nell'anno precedente, che per praticità viene qui richiamato:



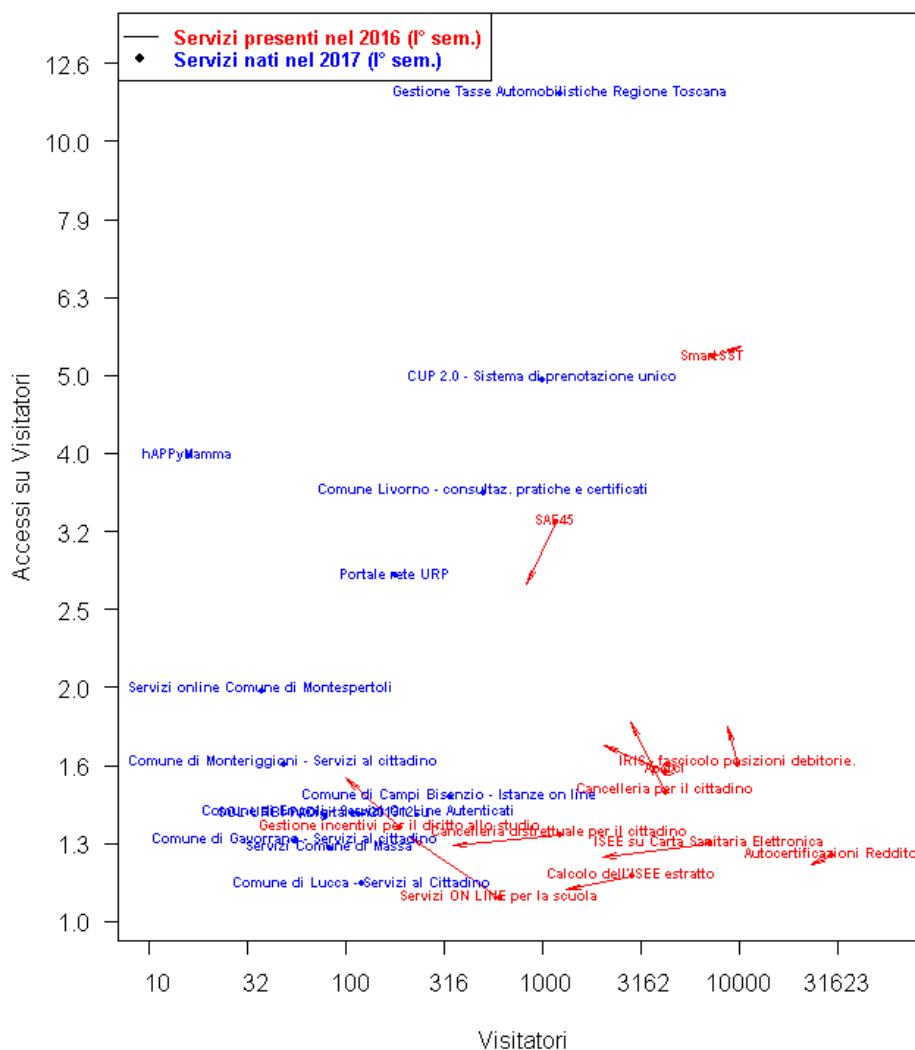
Ad esclusione del sabato e della domenica, l'andamento degli accessi è molto simile nei diversi giorni della settimana, evidenziando tra l'altro un picco durante la mattina dalle 9 alle 11 e uno nel pomeriggio intorno alle 16.

Il comportamento degli utenti è poco variabile tra i vari giorni nelle fasce orarie antimeridiane ed è invece caratterizzato da maggiore variabilità nelle fasce orarie pomeridiane e serali, dovuto soprattutto alla marcata differenza di comportamento nel venerdì pomeriggio (anche in questo caso si riporta di seguito la situazione 2016, essendo anche questo aspetto rimasto pressoché uguale nel primo semestre 2017.)



Si conferma quindi che i servizi on-line vengono acceduti in misura rilevante 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) o addirittura in orario notturno (dalle 20 alle 7 del mattino successivo).

Visitatori e accessi su visitatori alle risorse - evoluzione 2016 - 2017 (1° semestre),
SERVIZI PER IL CITTADINO



I dati confermano innanzi tutto come i servizi per i professionisti siano utilizzati molto spesso (proprio perché 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.

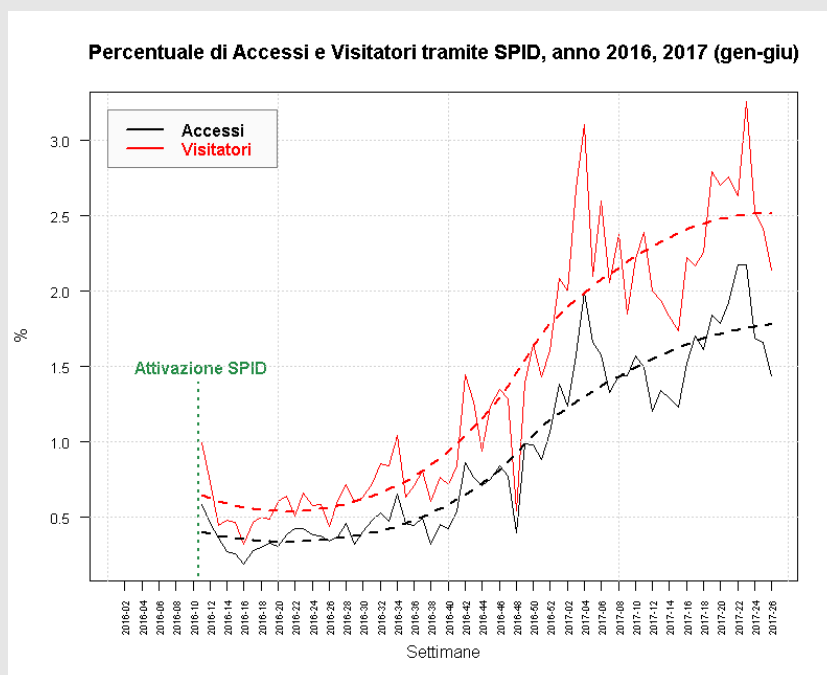
Con riferimento poi ai soli servizi presenti in entrambi gli anni, **nel caso dei servizi per il professionista si nota una generalizzata crescita del rapporto accessi su visitatori** (invece, il numero assoluto di visitatori in alcuni casi aumenta e in altri diminuisce), mentre **per i servizi orientati al cittadino una generalizzata riduzione dei visitatori** (invece, il rapporto accessi su visitatori in alcuni casi aumenta e in altri diminuisce).

Focus SPID

il 15 marzo del 2016 è divenuto operativo SPID (Sistema Pubblico di Identità Digitale) e gradualmente si sta diffondendo come strumento utile per l'accesso ai servizi qualificati. Trattandosi di una delle azioni prevista dall'Agenda Digitale italiana è opportuno analizzare quanto questo sia utilizzato per l'accesso ai servizi.

SPID rappresenta un fattore di espansione e di penetrazione nella popolazione su cui investire, sia per raggiungere soggetti non ancora coinvolti con lo strumento CNS sia per 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 e a partire dall'attivazione, la percentuale dei visitatori e degli accessi tramite SPID rispetto al totale di visitatori e di accessi:



L'evoluzione sembra indicare che a fronte di una crescita più che lineare che ha caratterizzato i primi mesi, **le quantità di accessi e di visitatori tendano a stabilizzarsi nell'ultimo periodo.**

Si può notare inoltre che la serie delle percentuali di utilizzo relativa ai visitatori è collocata sopra a quella degli accessi: questo fatto suggerisce la possibilità che l'utilizzo della credenziale di accesso SPID sia privilegiata da utenti che tendono ad utilizzare i servizi online più sporadicamente, mentre l'utilizzo della CNS o CIE sia preferita dagli utilizzatori più assidui (è possibile ad esempio che **i servizi con maggiore utilizzo in percentuale di SPID siano quelli caratterizzati da ampia diffusione, ma basso utilizzo, ossia principalmente i servizi orientati al cittadino.**)

Più precisamente, in termini numerici, tramite l'utilizzo dei canali CNS o CIE il numero degli accessi per visitatore è stato mediamente pari a 2.0, mentre nel caso di SPID questo stesso rapporto si assesta nel periodo analizzato su 1.4.

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 dati puntuali, cercano di far emergere 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.

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 ai 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 Hortonworks) 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).