Nota Tecnica

Relazioni critiche fra entità BIPEx

|  |  |
| --- | --- |
| **Doc\_ID** | IID5T-117200491-16300 |
| **Versione** | 2.1 |
| **Riassunto** | Questa nota tecnica descrive le relazioni critiche fra entità del protocollo BIPEx e i relativi impatti, in caso di relazioni incomplete sulle basi dati del Sistema Informativo Regionale dei Trasporti (SIRT) |
| **Numero di pagine** |  |

**Indice**

[1 Introduzione 3](#_Toc25315713)

[2 I vincoli dei flussi BIPEx 4](#_Toc25315714)

[2.1 TPL programmato 4](#_Toc25315715)

[2.1.1 Schema di alto livello delle entità 4](#_Toc25315716)

[2.1.2 Basi dati SIRT correlate 5](#_Toc25315717)

[2.1.3 Impatti 5](#_Toc25315718)

[2.2 TPL consuntivo 6](#_Toc25315719)

[2.2.1 Schema di alto livello delle entità 6](#_Toc25315720)

[2.2.2 Basi dati SIRT correlate 6](#_Toc25315721)

[2.2.3 Impatti 6](#_Toc25315722)

[2.3 SBE anagrafiche 7](#_Toc25315723)

[2.3.1 Schema di alto livello delle entità 7](#_Toc25315724)

[2.3.2 Basi dati SIRT correlate 8](#_Toc25315725)

[2.3.3 Impatti 8](#_Toc25315726)

[2.4 SBE consuntivi 9](#_Toc25315727)

[2.4.1 Schema di alto livello delle entità 9](#_Toc25315728)

[2.4.2 Basi dati SIRT correlate 10](#_Toc25315729)

[2.4.3 Impatti 10](#_Toc25315730)

[3 Controlli di coerenza e qualità del dato. 11](#_Toc25315731)

[3.1 Controlli pre-importazione 11](#_Toc25315732)

[3.2 Controlli post-importazione 11](#_Toc25315733)

[3.2.1 Verifiche sul numero di dati importati 11](#_Toc25315734)

[3.2.2 Verifiche di coerenza fra entità correlate di flussi diversi 12](#_Toc25315735)

[3.2.2.1 Abbonamenti 12](#_Toc25315736)

[3.2.2.2 Credito a scalare 13](#_Toc25315737)

[3.2.2.3 Controlli sul diario di viaggio 14](#_Toc25315738)

# Introduzione

I 4 flussi BIPEx (TPL programmato, TPL consuntivo, SBE anagrafiche e SBE consuntivi) utilizzati per popolare le 3 basi dati regionali del SIRT (TPL, Tariffazione e frequentazione) prevedono una serie di riferimenti incrociati di due tipologie:

* Interni al singolo flusso: questi garantiscono la coerenza e l’integrità dei dati del singolo flusso e in alcuni casi specifici, qualora risultino incompleti causano l’arresto della procedura di importazione
* Fra flussi BIPEx differenti: questi consentono a posteriori la ricostruzione di informazioni più complesse quali quelle legate a specifici indicatori di Business Intelligence.

Nei prossimi paragrafi si analizzeranno ad alto livello le relazioni più importanti nei 4 flussi BIPEx evidenziando gli impatti che una loro incoerenza può causare alle basi dati SIRT (Capitolo 2) e i controlli ad oggi effettuati sui dati sia nella fase pre-importazione che post- importazione (Capitolo 3).

Per i dettagli sul significato di ogni entità BIPEx e per l’elenco completo degli attributi si rimanda alla lettura delle “Linee guida alla compilazione del BIPEx”.

I vincoli dei flussi BIPEx

## TPL programmato

Questo flusso è dedicato alla descrizione delle dotazioni aziendali per l’esercizio del servizio TPL, alla descrizione della rete trasportistica e del relativo servizio TPL programmato.

### Schema di alto livello delle entità

Nel descrivere il flusso si evidenzieranno in rosso i riferimenti a entità appartenenti a flussi diversi da quello oggetto dell’analisi.

* Enti correlati al servizio TPL (ResourceFrame)
  + Operatori
  + Enti
  + CCA (con riferimento agli operatori che ne fanno parte)
* Dettagli contrattuali (ResourceFrame)

Anagrafica dei contratti di servizio con riferimenti a:

* + Enti Concedenti
  + Consorzi (operatori con ruolo Ente affidatario)
  + Operatori (operatori con ruolo Esercente)
* Dotazioni (ResourceFrame)

Anagrafica veicolare (con riferimento all’operatore a cui appartiene) descritta con

* + Tipi di veicolo
  + Veicoli (con riferimento al tipo di veicolo)
* Rete (ServiceFrame)
  + Fermate (con riferimento all’Area di appartenenza e alla Zona tariffaria – definita nel flusso SBE anagrafiche)
  + Aree
  + Archi fra fermate (con riferimento alle fermate agli estremi dell’arco)
  + Percorsi (con riferimento agli archi che lo compongono e alla linea di appartenenza)
  + Linee
* Calendari di validità (ServiceCalendarFrame)
  + Tipologia di giorno
  + Giorno di validità della corsa (con riferimento alla tipologia di giorno)
* Corse programmate (TimetableFrame) con riferimento a:
  + Tipo veicolo della corsa
  + Tipologia di giorno di validità
  + Linea
  + Percorso
  + Fermate
  + Contratto di servizio

### Basi dati SIRT correlate

La base dati correlata a questo flusso è quella **TPL programmato/esercito.**

Tuttavia siccome questi dati vengono analizzati in relazione anche alla base dati di Tariffazione, questo flusso influisce anche sulla base dati **Frequentazione**.

### Impatti

Qualora le relazioni indicate sopra fossero incomplete si avrebbero impatti su:

* Ricostruzione della correlazione fra servizio TPL pianificato aziendale e Programma di Esercizio Aziendale predisposto dall’Ente (PEA)
* Indicatori PEA relativi a prestazioni (teoriche): km totali, veicoli \* km ecc.
* Indicatori di offerta (servizio vs tipologie di veicoli): posti \* km

## TPL consuntivo

Questo flusso è dedicato al servizio monitorato e ai relativi giustificativi che consentono la consuntivazione del monitorato.

### Schema di alto livello delle entità

Nel descrivere il flusso si evidenzieranno in rosso i riferimenti a entità appartenenti a flussi diversi da quello oggetto dell’analisi.

* Giustificativi per corse effettuate con difformità rispetto a PEA
* Corse effettuate (DatedVehicleJourneyDelivery) con riferimento a:
  + Tipo veicolo della corsa (definito in TPL programmato)
  + Veicolo (definito in TPL programmato)
  + Tipologia di giorno di effettuazione (definito in TPL programmato)
  + Giorno di effettuazione
  + Percorso (definito in TPL programmato)
  + Corsa programmata (definito in TPL programmato)
  + Fermate (definito in TPL programmato)
  + Contratto di servizio (definito in TPL programmato)
  + Giustificati (per le sole corse con difformità)

### Basi dati SIRT correlate

La base dati correlata a questo flusso è quella **TPL programmato/esercito.**

Tuttavia siccome questi dati vengono richiamati anche dalla base dati di Tariffazione (nelle validazioni), questo flusso influisce anche sulla base dati **Frequentazione**.

### Impatti

Se le relazioni indicate sopra dovessero essere incomplete diverrebbe impossibile la verifica di:

* rispetto del PEA (correlazione fra TPL consuntivo e TPL programmato)
* indicatori relativi all’efficienza del servizio: km totali, veicoli \* km ecc.

## SBE anagrafiche

Questo flusso è dedicato all’anagrafica tariffaria e a tutte le dotazioni della rete di vendita e validazione del BIP.

### Schema di alto livello delle entità

Nel descrivere il flusso si evidenzieranno in rosso i riferimenti a entità appartenenti a flussi diversi da quello oggetto dell’analisi.

* Anagrafiche generali (FareFrame)
  + Tipo di ricarica del titolo
  + Tipologia di supporto (smartcard)
  + Tipologia di apparato di validazione / vendita
* Anagrafica tariffaria (FareFrame)
  + Zone tariffarie
  + Parametri di validità dei titoli di viaggio:
    - proprietà temporali di validità
    - proprietà spaziali di validità
    - tipologia di utente
    - operatore abilitato alla vendita (definito in TPL programmato)
  + Struttura tariffaria inteso come fattori spaziali e temporali
  + Prezzi con riferimento a
    - Parametri di validità
    - Struttura tariffaria
  + Elementi validabili con riferimento a:
    - Struttura tariffaria
    - Parametri di validità
    - Prezzi
    - Tipologia di ricarica del titolo
  + Prodotti tariffari sia abbonamenti che credito a scalare con riferimento a:
    - Parametri di validità
    - Prezzi
    - Tipologia di ricarica del titolo
  + Titoli vendibili con riferimento a:
    - Prodotti tariffari
    - Prezzi
    - Tipologia di supporto
* Rete di vendita (FareFrame)
  + Rivendite
  + Dispositivi con riferimento a:
    - Tipologia di dispositivo
    - Veicolo di installazione (se veicolare) (definito in TPL programmato)
    - Rivendita
    - Operatore (definito in TPL programmato)
  + SAM con riferimento a:
    - Operatore
    - Dispositivo associato

### Basi dati SIRT correlate

La base dati correlata a questo flusso è quella **Tariffazione**

Tuttavia, siccome questi dati vengono contengono riferimenti a TPL programmato e Anagrafiche tariffarie, questo flusso influisce anche sulla base dati **Frequentazione**.

### Impatti

Se le relazioni indicate sopra dovessero essere incomplete sarebbe impossibile capire quali titoli di viaggio siano venduti da specifici operatori.

## SBE consuntivi

Questo flusso è dedicato alla consuntivazione del sistema di bigliettazione economica.

### Schema di alto livello delle entità

Nel descrivere il flusso si evidenzieranno in rosso i riferimenti a entità appartenenti a flussi diversi da quello oggetto dell’analisi.

* Anagrafiche generali
  + Tipo di ricarica del titolo
  + Tipologia di supporto (smartcard)
  + Tipologia di apparato di validazione / vendita
* Vendite (SaleTransactionDelivery) con riferimento a:
  + Prodotto tariffario (definito in SBE anagrafiche)
  + Supporto su cui viene scritta la vendita
  + Parametri di validità
  + Operatore che effettua la vendita (definito in TPL programmato)
  + SAM che firma la vendita (definito in SBE anagrafiche)
  + Dispositivo usato nella vendita (definito in SBE anagrafiche)
* Supporti su cui avviene la vendita (TravelDocumentDelivery) con riferimento a:
  + Tipologia di supporto
  + Cliente
  + Operatore che effettua la vendita (definito in TPL programmato)
* Clienti (CustomerDelivery) con riferimento a:
  + Operatore che effettua la vendita (definito in TPL programmato)
  + Profilo utente (definito in SBE anagrafiche)
* Validazioni (ControlEntryDelivery) con riferimento a:
  + Operatore che effettua la corsa (definito in TPL programmato)
  + Corsa monitorata (definito in TPL consuntivo)
  + Dispositivo di validazione (definito in SBE anagrafiche)
  + Fermata di validazione (definito in TPL programmato)
  + Supporto
  + Veicolo (definito in TPL programmato)
* Liste nere Supporti e SAM (BlackListDelivery) con riferimento a:
  + Supporto
  + SAM (definito in SBE anagrafiche)

### Basi dati SIRT correlate

La base dati correlata a questo flusso è quella **Tariffazione.** Tuttavia, siccome questi dati vengono contengono riferimenti a TPL consuntivato e TPL programmato e Anagrafiche tariffarie, questo flusso influisce anche sulla base dati **Frequentazione**.

### Impatti

Se le relazioni indicate sopra dovessero essere incomplete sarebbe impossibile:

* correlare le validazioni al servizio effettuato e, di conseguenza, al programmato rendendo impossibile quindi la stima della domanda di trasporto pubblico: carichi, load factor
* ricostruire le matrici Origine/Destinazione di utilizzo del servizio pubblico
* ricostruire le vendite e quindi verificare il clearing
* completare con risultati affidabili qualunque sperimentazione tariffaria.

**Inoltre, in assenza di blacklist compilate correttamente la sicurezza dell’intero sistema sarebbe compromessa.**

Controlli di coerenza e qualità del dato.

## Controlli pre-importazione

Al momento dell’invio di un flusso BIPEx dal CCA al CSR-BIP e prima che il flusso venga importato nel database del CSR, vengono svolte verifiche su:

* Correttezza sintattica del flusso (per confronto con l’XSD BIPEx)
* Correttezza semantica delle entità
* Coerenza nei riferimenti fra entità differenti all’interno di ciascun flusso

Qualora il controllo sul singolo flusso non venga superato, il flusso stesso viene respinto.

Non vengono effettuati in questa fase controlli di coerenza fra i dati inviati nel singolo flusso e quelli inviati in altri flussi o già presenti nel DB del CSR (e quindi ricevuti in invii precedenti).

## Controlli post-importazione

Una volta inseriti i dati nelle apposite strutture, esistono procedure che copiano i dati in un DataWareHouse ottimizzato per interrogazioni rapide.

Su questo DataWareHouse vengono svolte verifiche più approfondite che hanno lo scopo di controllare la coerenza di dati importati in tempi anche molto diversi.

Di seguito vengono descritte le varie verifiche che vengono effettuate.

Questi controlli, ad oggi, sono effettuati manualmente e i relativi risultati vengono scambiati in formato elettronico con i CCA che ne fanno richiesta, ma verranno integrati in uno strumento gestionale specifico del CSR-BIP e potranno essere resi disponibili, automaticamente, via web, ai CCA.

### Verifiche sul numero di dati importati

Questo tipo di interrogazione prevede due modalità differenti:

* Verifica del numero di dati inviato in uno specifico flusso BIPEx, effettuabile su un qualunque flusso)
* Verifica del numero di dati relativi ad un giorno specifico: questo controllo è volto a verificare quanti dati sono ricollegabili ad un giorno ben determinato; è un confronto che è applicabile solo alle entità BIPEx correlate al tempo (e quindi, in particolare, ai consuntivi TPL e SBE).

Lo scopo di questa verifica è “contare” dati che a causa di ritardi fisiologici nello scaricamento dalla periferia (mezzo, validatore, rivendita) verso il CCA, non necessariamente arrivano nella loro completezza in un singolo invio.

Per questi dati viene prodotta una tabella che può essere condivisa con il CCA che ne faccia richiesta.

### Verifiche di coerenza fra entità correlate di flussi diversi

Lo scopo di questa verifica è controllare che tutti i riferimenti fra flussi diversi, inviati eventualmente in tempi differenti (evidenziati in rosso nel capitolo 2) siano stati valorizzati correttamente e quindi le relative relazioni possano essere completamente ricostruite.

Di seguito si riportano i controlli più critici già previsti; essendo stati creati per verificare la coerenza della validazione che è l’ultimo tassello del sistema BIP, consentono di ripercorrere e verificare l’intera struttura di dati ottenendo riscontri anche su:

* completezza della descrizione della rete
* coerenza dei dati relativi al servizio esercito per confronto con il servizio programmato
* coerenza dei dati relativi alle validazioni per confronto con servizio esercito e programmato.

I controlli effettuati, ad oggi, si differenziano in funzione della tipologia di titolo validato dagli utenti (abbonamenti o credito a scalare). I file di log che tracciano gli errori sono già in parte realizzati e in parte ancora in fase di realizzazione e saranno tutti integrati in un apposito strumento di verifica.

#### Abbonamenti

Gli utenti abbonati validano il titolo di viaggio solo in salita (ad ogni validazione corrisponderà nel database un record con tipo di validazione “*validityCheck”*), per cui la fermata in cui avviene la discesa dal mezzo deve essere stimata attraverso un algoritmo di stima del “check out” virtuale.

Di seguito elenchiamo i controlli preliminari da effettuare sulle validazioni in salita.

Presenza di record multipli

Ogni validazione è univocamente identificata mediante il codice della tessera, la data e l’ora, il codice della corsa (a consuntivo) e della fermata e il tipo di validazione. A parità di questi attributi due validazioni vengono considerate identiche, quindi ai fini del calcolo del check-out virtuale è necessario rimuovere tali record multipli erroneamente generati. Per tutte le altre applicazioni, per cui questo errore non risulti critico, ci si limita a produrre un log che elenca gli errori.

Corrispondenza tra corsa a consuntivo e corsa programmata

La validazione riporta il codice della corsa a consuntivo, mentre la descrizione dell’esercizio (database di riferimento per la stima della fermata di discesa) fa riferimento al codice della corsa programmata, quindi è necessario verificare che il codice a consuntivo corrisponda ad un codice programmato.

Ai fini del calcolo del check-out virtuale, in caso non si trovasse una corrispondenza, la validazione e la catena di spostamenti ad essa associata (indicata dalla chiave tessera-data) vengono eliminate e l’anomalia viene segnalata tramite un log di errore. In caso contrario, alla validazione viene associato il codice corsa programmata individuato. Per tutte le altre applicazioni, per cui questo errore non risulti critico, ci si limita a produrre un log che elenca gli errori.

Esistenza di fermata e corsa

I controlli verificano che:

1. la fermata di validazione esista nel database relativo alla Rete
2. la corsa su cui la tessera ha validato esista nel database Esercizio
3. la fermata si trovi nell’elenco delle fermate della corsa.

Ai fini del calcolo del check-out virtuale, le validazioni che non superano questo controllo e le catene di spostamenti ad esse associate (indicate dalla chiave tessera-data) vengono eliminate, e l’anomalia viene segnalata tramite un log di errore. Per tutte le altre applicazioni, per cui questo errore non risulti critico, ci si limita a produrre un log che elenca gli errori.

Coerenza con i calendari di validità

Avendo a disposizione il calendario di validità di ogni corsa del servizio programmato, cioè l’elenco dei giorni in cui viene effettuata, occorre controllare che la data della validazione avvenuta su una certa corsa (a consuntivo) sia effettivamente presente nel calendario di validità della corrispondente corsa programmata.

Ai fini del calcolo del check-out virtuale, si eliminano i record che presentano tale anomalia e si produce un log che segnala l’anomalia stessa. Per tutte le altre applicazioni, per cui questo errore non risulti critico, ci si limita a produrre un log che elenca gli errori.

Terminati i controlli preliminari sulle validazioni degli abbonamenti è possibile elaborare i dati in vista della costruzione del diario di viaggio, ricorrendo all’algoritmo per la stima del check out virtuale.

#### Credito a scalare

Gli utenti che viaggiano con il credito trasporti, a differenza degli abbonati, validano il titolo in salita e in discesa, in modo da consentire al sistema di calcolare e scalare dal credito l’importo corrispondente alla tratta percorsa.

I record che vengono generati dal credito trasporti nel database Validazioni, quindi, sono identificati dal tipo di validazione checkIn se corrispondono ad una salita e checkOut se corrispondono ad una discesa.

In entrambi i casi i controlli preliminari da effettuare sono gli stessi dettagliati nella sezione precedente (a cui si rimanda per i dettagli), ovvero:

* Presenza di record multipli
* Corrispondenza tra corsa a consuntivo e corsa programmata
* Esistenza della fermata e della corsa e appartenenza della fermata alla corsa
* Coerenza con i calendari di validità delle corse

L’elaborazione dei dati che hanno superato i controlli ai punti precedenti prevede, inoltre, l’abbinamento del “check in” con il corrispondente “check out” sulla medesima corsa, che insieme descrivono un singolo spostamento associato alla tessera. Per fare ciò, si considerano i “check in” e i “check out” avvenuti nello stesso giorno sulla stessa corsa e si selezionano (in base all’orario) il primo “check in” e l’ultimo “check out”.

Anche in questo caso, terminati i controlli preliminari sulle validazioni è possibile elaborare con la costruzione del diario di viaggio.

#### Controlli sul diario di viaggio

Il diario di viaggio è la struttura che raccoglie le informazioni dettagliate degli spostamenti associati alle tessere: per ogni coppia tessera-data il diario riporta l’elenco degli spostamenti effettuati, con le informazioni relative a tipo di validazione, corsa, linea, fermata e orario di salita e di discesa (reale nel caso del credito trasporti, stimata nel caso degli abbonamenti), distanza percorsa a bordo e tempo di percorrenza.

Anche su questi dati elaborati ed aggregati è necessario eseguire ulteriori controlli, differenziati a seconda del tipo di validazione effettuata.

##### Abbonamenti

Sulle validazioni di tipo *validityCheck* si effettua un controllo sull’esistenza degli archi fra le fermate.

Esistenza archi

Per valutare alcuni degli indicatori di frequentazione a partire dai dati delle validazioni, è necessario conoscere la distanza percorsa a bordo delle singole corse utilizzate da ciascuna tessera. È quindi necessario controllare l’esistenza nel database Rete di tutti gli archi tra fermate successive, dalla fermata di salita alla fermata di discesa stimata, in modo da poter attribuire un valore alla distanza percorsa.

Per le validazioni che non superano questo controllo, è possibile stimare la lunghezza degli archi mancanti e quindi la distanza percorsa, calcolando la distanza in linea d’aria (conoscendo la latitudine e la longitudine) tra le fermate successive per le quali manca l’arco di collegamento.

Nel caso di archi mancanti viene comunque creato un log.

##### Credito trasporti

Le validazioni effettuate con il credito trasporti richiedono necessariamente un controllo sui progressivi fermata, uno sugli archi e un possibile ulteriore controllo sul tempo di percorrenza.

Progressivi fermata

I *checkIn* e i *checkOut* del credito trasporti sono stati abbinati, come descritto nel capitolo precedente, facendo riferimento agli orari delle validazioni, a parità di corsa. A questo punto è quindi necessario controllare che il progressivo fermata associato alla fermata di salita sia effettivamente precedente rispetto a quello della fermata di discesa. In caso contrario, significherebbe che la tessera ha effettuato il check out ad una fermata precedente a quella del check in. Un tale errore potrebbe essere dovuto ad un codice fermata sbagliato oppure ad un’imprecisione nell’attribuzione del progressivo fermata nel database contenente le corse programmate. In ogni caso, ai fini del calcolo del check-out virtuale, le validazioni che non superano questo controllo vengono eliminate e gli errori che hanno causato l’eliminazione sono inseriti in un log. Per tutte le altre applicazioni, per cui questo errore non risulti critico, ci si limita a produrre un log che elenca gli errori.

Esistenza archi

Per valutare alcuni degli indicatori di frequentazione a partire dai dati delle validazioni, è necessario conoscere la distanza percorsa a bordo delle singole corse utilizzate da ciascuna tessera. È quindi necessario controllare l’esistenza nel database Rete di tutti gli archi tra fermate successive, dalla fermata di salita alla fermata di discesa stimata, in modo da poter attribuire un valore alla distanza percorsa. La correzione prevista, per le validazioni che non superano questo controllo, consiste nella stima della lunghezza degli archi mancanti e quindi della distanza percorsa, ottenuta calcolando la distanza in linea d’aria (conoscendo la latitudine e la longitudine) tra le fermate successive per le quali manca l’arco di collegamento.

Nel caso di archi mancanti viene comunque creato un log.

Tempo di percorrenza

Nel caso di validazioni con credito a scalare non è necessario stimare l’orario di discesa, poiché è fornito dalla validazione di *checkOut*. Potrebbe però essere effettuato un ulteriore controllo sul tempo di percorrenza: si confrontano pertanto gli orari di salita e discesa con quelli rilevati dal servizio esercito. Nel caso si riscontrassero incongruenze oppure tale informazione mancasse nel database delle corse a consuntivo, si confronta il tempo di percorrenza desunto dagli orari di *checkIn* e *checkOut* con quello previsto da programmato, producendo una segnalazione di potenziale errore nel caso si riscontrasse una notevole differenza tra i due (per esempio, superiore ai 15 minuti).

Le incongruenze più gravi vengono inserite in un log.