

L'IDENTIFICAZIONE DEGLI OGGETTI GEOGRAFICI IN UN SISTEMA INFORMATIVO

Manuela Corongiu (*), Gabriele Garnero (**), Mauro Rossi (*)

(*) Ente LaMMA - CNR Ibimet - Regione Toscana, via Madonna del Piano 10, 50019 Sesto Fiorentino (FI), tel.+39055 4483041, fax +39 055 444083, E-mail corongiu@lamma.rete.toscana.it, mr.sit@tiscali.it

(**) Università degli Studi di Torino - DEIAFA – V. Leonardo da Vinci, 44 – 10095 Grugliasco (TO) Tel +39 011 670.5518, Fax +39 011 670.5516 – E-mail gabriele.garnero@unito.it

L'informazione geografica sta uscendo dal "ghetto" dei GIS per diventare una delle dimensioni, quella spaziale, di un sistema informativo multidimensionale e multimediale. In questo contesto è importante poter disporre di un identificatore degli oggetti indipendente dal dominio applicativo, di adeguata robustezza applicativa e di persistenza nel tempo. In questa memoria vengono esaminate le proprietà strutturali dell'UUID DCE e la sua piena rispondenza ai requisiti di unicità universale, come proprietà intrinseca, oltre che di persistenza nel tempo. Vengono inoltre proposte delle modalità gestionali per risolvere il problema del lookup service. Tale identificatore è stato sperimentato come foreign key applicandolo ai dati relativi alla sezione campione di Gattinara (094130) prodotti nel contesto della produzione prototipale per l'aggiornamento della CTR alla scala 1:10.000 della Regione Piemonte, realizzata applicando le specifiche IntesaGIS.

The geographic information is going out from the "ghetto" of the GIS to become one dimension, the spatial one, of a multidimensional and multimedia informative system. In this framework, it is paramount to have an object identifier which is independent from the applied dominion. Furthermore it must have an adequate applicative robustness and a time durable persistence. In this document the UUID DCE structural property, its full correspondence with the universal uniqueness requirements (as intrinsic property) and the time durable persistence are examined. Some management practices to solve the problem of the lookup service are also proposed. Such identifier was tested as foreign key applying it to the data referred to the Gattinara sample section (094130). These data were produced during the prototypal production for the "Regione Piemonte" CTR updating (scale 1:10.000); which was realized applying the IntesaGIS's specifications.

Il contesto informativo

Per una corretta posizione del problema occorre tenere conto della tumultuosa evoluzione dell'Information Technology, dell'estendersi del processo di automazione e della sua integrazione nel vasto contesto della comunicazione.

L'Informazione Geografica, con l'avvento del GPS e delle reti wireless, sta uscendo dal ghetto riservato a pochi esperti per diventare oggetto di largo consumo: ne consegue che gli oggetti attualmente gestiti dalla Cartografia Numerica o dai GIS entreranno a far parte di un più vasto Sistema Informativo dove la gestione della dimensione spaziale sarà una delle funzionalità previste, come la gestione della dimensione temporale o di quella multimediale.

In un tale contesto risulta evidente il ruolo giocato dall'identificazione degli oggetti e dall'esigenza sempre più evidente di una sua robustezza applicativa e persistenza nel tempo. Un'idea di tale processo la si può dedurre per similitudine dal ruolo giocato dal Codice Fiscale italiano, che, progettato ai fini fiscali, si è rapidamente imposto come valido strumento per identificare in modo univoco ogni cittadino, estendendosi *naturalmente* a tutti i campi che coinvolgono l'identificazione di una persona, dalle varie "anagrafiche", sino alle transazioni bancarie, ai contratti ed in genere a tutti i documenti amministrativi. Nello specifico geografico un'analogia immediata con

l'identificazione di una persona è quella relativa ad un edificio ed al ruolo che questo gioca nei molteplici flussi informativi e gestionali.

Le caratteristiche di un OID, Object Identifier

Gli attuali limiti dei software GIS costringono ad escludere dal discorso gli *internal identifier*, utilizzati dal sistema per la gestione degli oggetti, e di restringerlo agli *external identifier*, da utilizzare in relazioni tra dati di sistemi diversi, in un ruolo molto simile a quello delle *foreign-key* dei sistemi relazionali. L'OID non è da considerarsi neppure a sostituzione dei vari identificatori applicativi, tematici o gestionali, quali ad esempio quelli delle particelle catastali o quelli della rete stradale, di cui deve essere garantita la sopravvivenza, prima di tutto come doveroso rispetto del linguaggio dei loro gestori e dei loro utenti.

L'unicità universale

Lo scambio dei dati in una rete sempre più globale impone che l'OID risulti unico non solo in uno specifico contesto applicativo, ma risulti tale in senso assoluto, cioè sia un UUID, Universally Unique Identifier. Si ritiene fondamentale a questo proposito distinguere quando l'unicità è proprietà intrinseca all'identificatore stesso o è garantita da un Ente gestore, in quanto godono di caratteristiche molto differenti tra loro.

L'unicità universale come proprietà intrinseca. L'UUID

A questo tipo appartengono gli identificatori la cui unicità a livello universale sia garantita solamente dal rispetto di alcune regole di generazione. Appartiene a questo caso l'UUID che fa riferimento allo standard ISO/CIE 11578:1996, assunto dall'ANS.1 come OID nel documento X667 (ITU-T X667), e che è oggetto di questo documento. Per evitare confusioni si indicherà con UUID solo tale famiglia di identificatori.

Unicità garantita da un Ente gestore ed universalità basta sul Namespace

Questo tipo di identificatore, proposto anche per l'informazione geografica, si basa su una codifica, spesso una numerazione progressiva, la cui unicità sia garantita da un Ente gestore che ne cura l'attribuzione. La sua unicità a livello universale viene poi garantita associandolo ad un Namespace, tipicamente un indirizzo della rete web. Questa modalità, indicata anche nelle prescrizioni dell'OGC, si basa sull'unicità di una URN che rende universale qualsiasi codifica che sia unica nel suo contesto. Un tale tipo di identificazione è fortemente inadeguato allo sviluppo prefigurato nell'introduzione, in quanto risente dei seguenti limiti:

- l'unicità dipende dall'esistenza di un'Autorità di registrazione e dall'efficienza del processo di registrazione stesso;
- la sua sopravvivenza dipende da quella dell'Ente. La dipendenza da un'Autorità di registrazione può risultare obsoleta, qualora la gestione dell'oggetto in questione passi ad un'amministrazione differente;
- le procedure di registrazione possono risultare inutilmente gravose in un contesto informativo distribuito che vede l'interazione di numerosi *data administrator* e può essere la causa di indesiderati blocchi alla prosecuzione di un lavoro;
- la sua universalità e la sua sopravvivenza dipende dalla sopravvivenza del Namespace;
- questa modalità per sua natura limita fortemente la codifica al contesto applicativo.

La persistenza di un OID

Si indica con persistenza la caratteristica di robustezza di un OID ai fini della sua durata nel tempo, indipendentemente dall'evoluzione della tecnologia che ne elabora l'oggetto identificato e dall'ente che ne cura la gestione amministrativa. Questa proprietà dipende da diverse caratteristiche di cui deve godere l'OID:

- non riuso di un OID. Una corretta gestione di un OID richiede che questo non venga più utilizzato qualora l'oggetto identificato venga modificato o cancellato;
- assenza di una logica integrata nella struttura dell'OID ed in particolare l'indipendenza da contesti tematici o amministrativi che ne limitino il dominio applicativo o la sua validità nel tempo;
- indipendenza strutturale da un Ente gestore che ne condizionerebbe la validità nel tempo, ad esempio la sua unicità, facendola dipendere dalla sua esistenza;
- numerosità adeguata alla quantità degli oggetti da codificare.

La persistenza e la robustezza della struttura dati

La persistenza di un identificatore è strettamente connessa ad una adeguata analoga qualità della struttura dati atta a garantire la persistenza degli oggetti identificati (Corongiu et al., 2007). Questo proprietà fondamentale è ignorata dal normale approccio, derivato dalla Cartografia Numerica, che identifica l'oggetto geografico con la sua componente geometrica. Si pensi ad esempio ad un edificio, la cui struttura dati sia costituita dalla sola componente spaziale, il suo contorno di distacco dal suolo. Una modifica di quest'ultima porta ad una cancellazione della geometria precedente e, per le regole di non riuso dell'identificatore, anche dell'edificio, con la conseguente perdita di tutte le relazioni esterne a questo correlate.

La verifica della correttezza di un OID e delle sua esatta assegnazione

L'uso di un OID richiederebbe la possibilità di verificarne:

- la correttezza della sua digitazione. Questo normalmente si basa sulla verifica mediante un *check-sum*. L'UUID non ha valori di *check-sum*
- la corretta attribuzione dell'OID ad una istanza. Questo si può ottenere solo con identificatori la cui composizione sia strettamente connessa al dominio applicativo, quale ad esempio il Codice Fiscale, contraddicendo la caratteristica di indipendenza da logiche applicative.

L'UUID

Come detto precedentemente, con questo termine ci si fa riferisce all'UUID definito nello standard ISO/CIE 11578:1996 e che è stato assunto dall'ANSI come OID nel documento X667 (ITU-T X667) cui si rimanda per una sua esposizione di dettaglio. La codifica dell'UUID è basata su una granularità spazio-temporale e tipologica. A questa si aggiunge un ulteriore insieme di valori destinato a garantirne l'unicità nel caso di una collisione temporale in fase di generazione.

La componente spaziale, MAC address o nodo

La suddivisione spaziale, denominata *nodo*, consiste nel valore dell'indirizzo MAC (ISO/IEC 8802-3) della scheda di rete presente nel computer utilizzato per la sua generazione. L'uso del MAC address, un identificatore cablato nell'hardware di ogni scheda di rete e la cui unicità è gestita dall'IEEE, è la chiave di volta su cui si basa l'unicità universale della codifica dell'UUID.

La componente temporale, *time stamp* o datario

Questo dato, denominato *time* (tempo) o *time stamp* (datario) rappresenta il valore del tempo corrispondente all'istante di generazione dell'UUID espresso in termini di intervalli di 100 nanosecondi trascorsi a partire dalla mezzanotte del 15 ottobre 1582, la data della riforma gregoriana del calendario. In un giorno possono essere definiti, con un solo computer (più precisamente una sola scheda di rete) 864 miliardi UUID distinti, per ogni tipo di UUID. Questo dato assicura un'adeguata, anzi ridondante, numerosità per la codifica.

La tipologia degli UUID, *Version* e *Variant*

Sono previsti diversi tipi di UUID, articolati per *Versione* e in *Variante*. Per una descrizione di dettaglio si rimanda al documento X667 (ITU-T X667) dell'ASN.1. In questo documento si fa riferimento alla versione DCE time-based ed alla sua variante DCE.

Il campo per garantire l'unicità: la sequenza di clock , o *clock sequence*

La codifica dell'UUID è stata progettata anche per far fronte ad una codifica in tempo reale come richiesto ad esempio da transazioni connesse alla comunicazione tra computer. In questo contesto, che definiamo *dinamico*, devono essere previste delle collisioni temporali tra due o più richieste di generazione dell'UUID, con conseguente assegnazione dello stesso valore per il tempo. Per ovviare a questo inconveniente è stato previsto un campo denominato *sequenza di clock*, da utilizzare per diversificare UUID generati nello stesso istante.

La composizione e la rappresentazione dell'UUID

I campi utilizzati per la codifica dell'UUID precedentemente descritti sono opportunamente concatenati in una stringa di 16 ottetti (byte) che può esser considerata anche come un numero intero senza segno di 128 bit. Per il dettaglio della sua composizione, cioè come i campi che lo compongono vengono formattati nei 16 byte, e per la sua rappresentazione, si rimanda sempre al documento X667 (ITU-T X667) dell'ASN.1.

Per la sua rappresentazione sono previste diverse modalità, ad esempio:

- esadecimale: **33809f10-7efa-11da-b9ea-0040f41cca8e**
- esadecimale senza separatori: **33809f107efa11dab9ea0040f41cca8**
- intero decimale senza segno : **68458467992998175305782945840285665934**
- come URN **urn:uuid:33809f10-7efa-11da-b9ea-0040f41cca8e**
urn:uuid:68458467992998175305782945840285665934

L'indirizzo MAC ed il requisito d'opacità

L'uso dell'indirizzo MAC in certi contesti può violare la privacy od essere utilizzato per l'attacco da parte di virus (es. virus Melissa) (Corongiu et al. 2007). Il suo uso previsto da questo documento si configura in un contesto esente da tali problemi.

L'UUID e l'identificazione degli Oggetti geografici in un Sistema informativo

Asincronicità della codifica

La maggior parte dei dati geografici non richiede un'assegnazione dinamica dell'identificatore, in sincronia con il loro utilizzo in rete. Si pensi ad esempio a tutti quei sistemi di identificazione con criteri di codifica che prevedono una preventiva validazione da parte di un'autorità di registrazione. In altri termini, l'identificazione di tali oggetti può essere generata in tempi e anche su sistemi diversi da quelli della loro fruizione in rete. Questa proprietà di asincronicità è significativa per:

- ottimizzare le procedure di generazione dell'OID. Inoltre rende non necessario l'uso del campo di sequenza del clock, che potrebbe essere utilizzato per riportarvi un valore di check-sum, anche se questo richiederebbe la definizione formale di un'ulteriore versione dell'UUID;
- dissociare l'indirizzo MAC utilizzato per la sua generazione da quello del sistema che lo rende disponibile in rete, risultando in tal modo indipendente dal requisito di opacità.

L'adozione dell'UUID nella versione e nella variante DCE

Da quanto sviluppato in questo documento emerge la proposta di adottare, come *external identifier* per l'identificazione degli oggetti nei Sistemi Informativi Geografici, l'UUID nella versione e variante DCE, basato sull'indirizzo MAC. Questa scelta garantisce:

- l'unicità universale a livello strutturale, cioè indipendente da certificazioni esterne;

- l'indipendenza applicativa ed una numerosità ridondante rispetto a quella degli oggetti da identificare, godendo quindi di robuste prerogative per la persistenza dell'OID;
- l'adozione e la condivisione in molteplici domini indipendentemente dai criteri tematici o tassonomici che li definiscono;
- semplicità ed autonomia operativa per la sua generazione,

L'indirizzo MAC deve essere legato all'indirizzo fisico e non multicast.

Per gli oggetti per i quali occorre invece un'assegnazione dell'identificatore con modalità sincrona al suo utilizzo, quali ad esempio i dati climatici o di monitoraggio, e quindi con possibile richiesta di un requisito di opacità, ci si può avvalere delle altre versioni dell'UUID, quali il GUID o quelli generati con modalità pseudo-casuale. Questi casi non rientrano in questa nota di approfondimento.

La strutturazione dei domini spazio temporali per il *lookup-service*

Il diffondersi della disponibilità nella rete web di dati, ha evidenziato la necessità di collegare un dato al server che lo fornisce. Il servizio di tale funzionalità è denominato *lookup-service*.

L'indipendenza del sistema utilizzato dalla generazione dell'UUID da quello del server di rete preposto alla distribuzione permette di pensare una gestione degli indirizzi MAC utilizzati per la codifica degli oggetti in modo da associarli opportunamente ai domini dei dati codificati al fine di gestirne il *lookup-service*.

Infatti si può pensare ad una URN in rete dove a sottoinsiemi di indirizzi MAC, utilizzati per la codifica degli UUID degli oggetti, corrisponda l'indirizzo del nodo di rete del server dove gli oggetti sono reperibili.

L'asincronicità temporale permette di progettare procedure di generazione degli UUID tali da generare anche tutti gli 864 miliardi UUID di un giorno. Grazie a questa generazione tabellare è possibile pensare di avvalersi della granularità temporale per ripartire opportunamente i domini degli oggetti gestiti da un server, senza intaccare le caratteristiche dell'UUID.

La generazione tabellare può essere utilizzata non solo per non perdere l'enorme quantità di identificatori generabili in un giorno, ma di recuperare anche quelle corrispondenti ai secoli passati. Inoltre può essere estesa, garantendo la correttezza delle operazioni, anche per utilizzare gli indirizzi MAC di schede non più funzionanti od obsolete.

La ripartizione spazio-temporale può esser quindi utilizzata per suddividere i domini degli oggetti codificati per una efficace gestione sia del *lookup-service* sia per associare loro alcuni dati di metainformazione, compresa quella relativa ai proprietari degli stessi o dei loro *data administrator*.

Si nota, per inciso, come l'associazione di un Namespace all'UUID risulti del tutto non necessaria ai fini di associare ad un dato la metainformazione sul suo fornitore, essendo questa una delle tante metainformazioni che possono essere associate ad un oggetto. Un tale approccio risulta inutilmente ingombrante e vincola l'identificazione ad un contenuto applicativo che può mutare violando pesantemente il requisito dell'assenza di logica incorporata nella generazione dell'identificatore.

La registrazione di un UUID

Per taluni oggetti può essere richiesta una registrazione presso una qualche Autorità che ne garantisca alcune proprietà sostanziali o formali, quali ad esempio la certificazione del dato ai fini legali. Questa attività, distinta da quella non necessaria per garantirne l'unicità, può essere sempre compiuta e può essere associata ad un singolo UUID od avvalersi della ripartizione spazio-temporale suaccennata.

La sperimentazione

L'adozione degli UUID è stata sperimentata nel contesto della ricerca COFIN2004 "*Strumenti GIS in ambiente WEB per l'analisi e il governo del territorio*", nell'ambito della quale è stata effettuata una prima analisi applicativa sull'uso dell'UUID come External Object Identifier.

Oggetto della sperimentazione sono stati i file di fornitura in formato Shape, relativi alla sezione campione di Gattinara (094130) prodotti nel contesto della produzione prototipale per l'aggiornamento della CTR alla scala 1:10.000 della Regione Piemonte, realizzata applicando le specifiche *IntesaGIS*.

La sperimentazione ha visto la ristrutturazione degli Shape-file con l'assegnazione contemporanea degli UUID e quindi delle relazioni necessarie a materializzare i collegamenti tra le varie componenti degli oggetti e basate sulla condivisione degli UUID, assumendo il ruolo di *foreign key*. Gli UUID sono stati assegnati alle varie istanze di oggetti, come attributo alfanumerico di nome UUID; tale struttura dati è pronta per un loro inserimento nel GML di cui è in fase di messa a punto la definizione dello schema.

La generazione degli UUID è stata fatta con programmi sviluppati in linguaggio C a partire da quanto proposto nei documenti di riferimento (P.Leach et al. 2005). Per la loro generazione è stata sperimentata anche la modalità tabellare.

Bibliografia essenziale

Di seguito si riportano alcune citazioni bibliografiche ritenute essenziali, rimandando a quanto riportato nel documento (Corongiu et al., 2006) per un quadro complessivo.

M.Corongiu, G.Garnero, M.Rossi (2006) "L'UUID - Universal Unique Identifier come external Object Identifier dei dati del DBT" - Progetto COFIN2004 "Strumenti GIS in ambiente WEB per l'analisi e il governo del territorio.", in corso di pubblicazione (<http://www.freegis-italia.org>)

M.Corongiu, G.Garnero, M.Rossi (2007) "L'implementazione dell'UUID in un dataset IntesaGIS like: Comunicazione per INSPIRE" (<http://www.freegis-italia.org>)

(DCE C706) The Open Group, CAE Specification, DCE 1.1:Remote Procedure Call, Document number C706

ITU-T X667 - ITU-T Recommendation X.667 (2004) | ISO/IEC 9834-8:2005, Information technology Open Systems Interconnection Procedures for the operation of OSI Registration Authorities: Generation and registration of Universally Unique Identifiers (UUIDs) and their use as Object Identifier components – (<http://www.itu.int/ITU-T/studygroups/com17/oid/X.667-E.pdf>)

P.Leach, M.Mealling, R. (RFC4122 05) draft-mealling-uuid-urn-05.txt - Network Working Group Internet-Draft 2005 – (<http://www.ietf.org/internet-drafts/draft-mealling-uuid-urn-05.txt>)