Challenges and potential solutions to implement temporal aspects in INSPIRE specifications

Presentation to: INSPIRE Conference 2015, Lisbon

Author: Dominique Laurent

Date: 26th May 2015

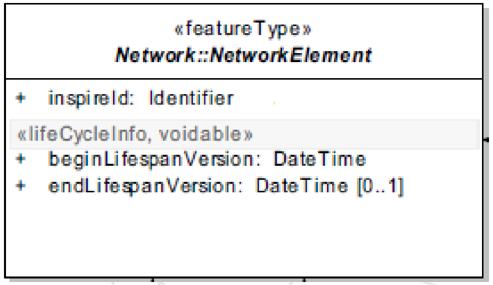
ISN.15.064

the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

ELF (European Location Framework) project

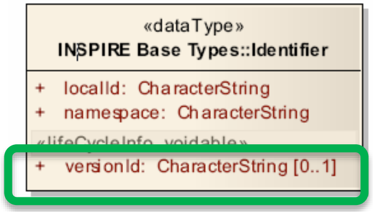
- Main objectives
  - Implement INSPIRE
    - Transform data (themes **AD, AU, BU, CP, GN, HY, TN**, EL**, LC, ...)** and set up download services
    - Offer operational platform for data and services
  - Exploit this platform by applications
    - Internal products (BaseMap, GeoLocator, ELF Cadastre, ..)
    - Applications (internal : insurance, real estate .. + external application developers)
- Main partners: NMCAs + technology providers
- From 2013 to 2016
- 50 % funded by European Commission

Theory: INSPIRE mechanism for incremental updates


«featureType»
Network::NetworkElement
+ inspireId: Identifier
«lifeCycleInfo, voidable»
+ beginLifespanVersion: DateTime
+ endLifespanVersion: DateTime [0..1]

- ★ Most of INSPIRE data models include:
  - ★ Inspire identifier
    - – unique
    - – **persistent**
  - ★ Temporal attributes related to the data base life-cycle
    - – beginLifespanVersion
    - – endLifespanVersion

} Versioning of objects


«dataType»
INSPIRE Base Types::Identifier
+ localId: CharacterString
+ namespace: CharacterString
«lifeCycleInfo, voidable»
+ versionId: CharacterString [0..1]

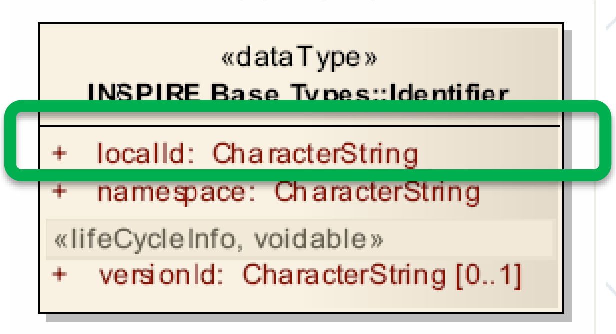
- ★ Users can get incremental updates (e.g. evolutions between t_1 and t_2) just by querying on temporal attributes
  - ★ beginLifespanVersion after t_1 and before t_2 ⇒ to get new (created) objects or new versions of modified objects
  - ★ endLifespanVersion after t_1 and before t_2 ⇒ to get old (deleted) objects or old versions of modified objects

the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

Practice: many issues

- Many data producers deliver only valid data (by regular releases) but do not give access to historical data
  - => Users can’t get information about old /deleted objets

- Not always persistent identifiers
  - Persistent identifiers missing in source data
  - Persistent identifiers lost during the transformation process
    - Split features
    - Merge features



«dataType»
INSPIRE Base Types::Identifier

+ localId: CharacterString
+ namespace: CharacterString

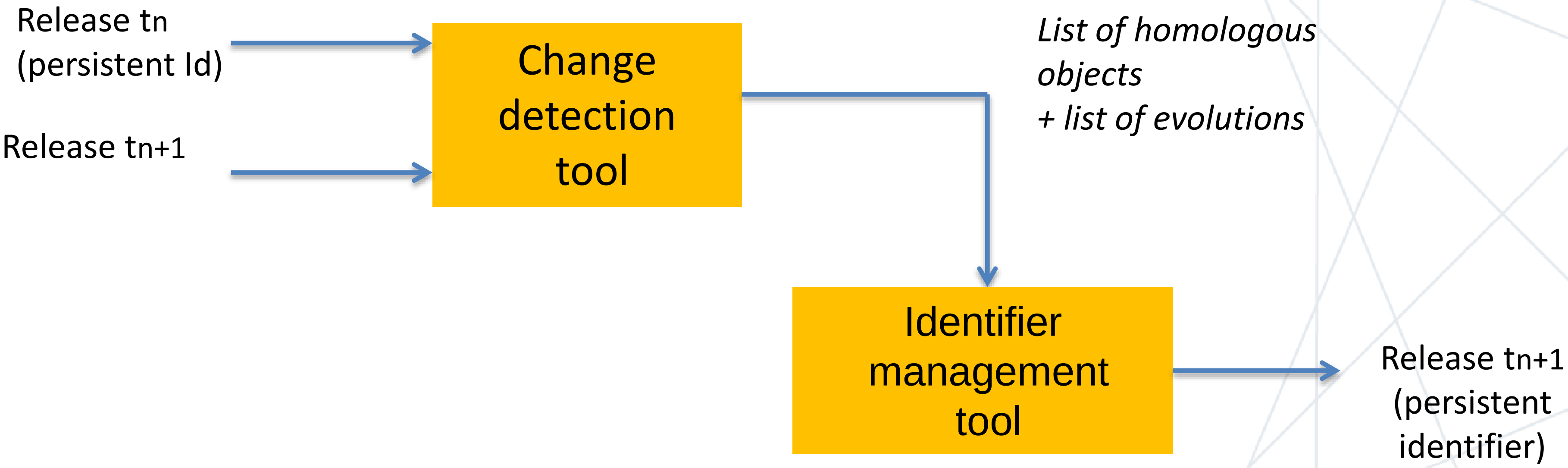
«lifeCycleInfo, voidable»
+ versionId: CharacterString [0..1]

- Temporal attributes
  - May be missing in source data
  - May be no longer reliable due to transformation process
    - More data in source than in INSPIRE
      - » Overdetections
    - Main source data + ancillary data to fill INSPIRE (e.g. by joining tables)
      - » underdetections

the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

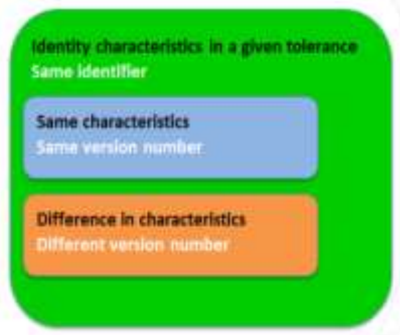
The ELF proposal: the change detection tool

- Principle: ensure **persistent** identifiers and provide temporal attributes, by comparing objects coming from 2 releases at t_n and t_{n+1}



Persistent identifier rules

- ★ The inspire identifier is about the database feature not about the real-world entity
- ★ General issue:
  - ★ When is an object considered as modified (same identifier, new version)?
  - ★ When it is considered as a new object (new identifier)?
- ★ An object is defined by a set of properties (geometry, semantic, ...)
  - ★ Significant change in main properties => new object => new identifier
    - – **Which are the main properties?**
    - – How big the change should be?
  - ★ Any other change => new version



Identity characteristics in a given tolerance
Same identifier

Same characteristics
Same version number

Difference in characteristics
Different version number

Persistent identifier rules: state-of-play

- INSPIRE context
  - No common rules
    - Life-cycle rules up to each data provider
    - Some good practice examples in the Methodology (guidelines for harmonisation)
- ESDIN (ELF predecessor) project:
  - Proposal for some common life-cycle rules for ”simple” objects
- Current IGN change detection tool
  - Rules for ”simple” objects
- => in all existing rules, **geometry is considered as main property, as identity property**

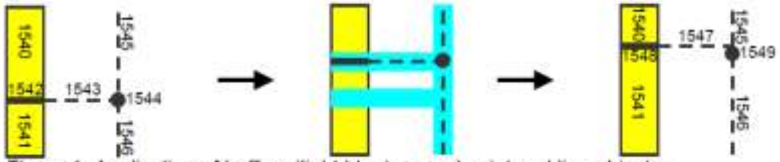
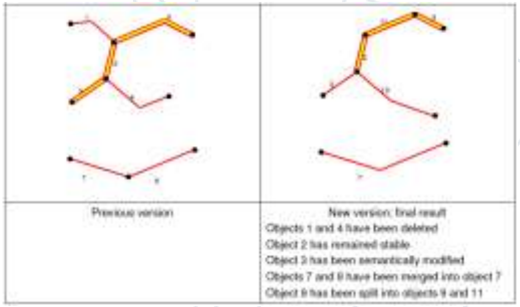
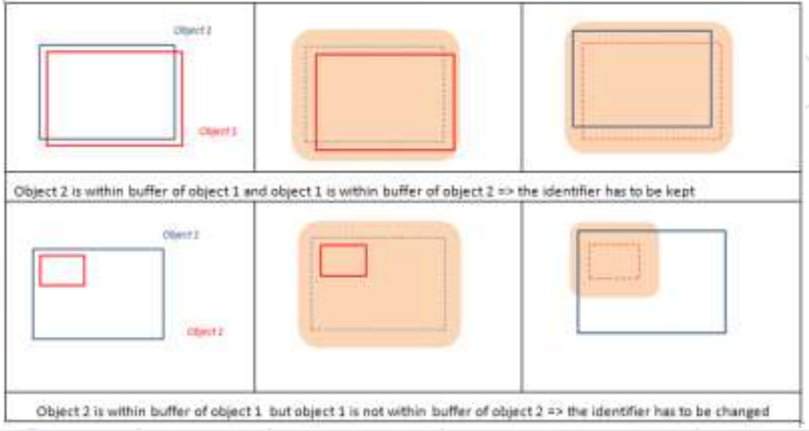


Figure 1: Application of buffers (light blue) around point and line objects.



Previous version

New version: final result
Objects 1 and 4 have been deleted
Object 2 has remained stable
Object 3 has been semantically modified
Objects 7 and 8 have been merged into object 7
Object 9 has been split into objects 9 and 11



Object 2 is within buffer of object 1 and object 1 is within buffer of object 2 => the identifier has to be kept

Object 2 is within buffer of object 1 but object 1 is not within buffer of object 2 => the identifier has to be changed

Persistent identifier rules: new challenge

- Existing rules are based on simple objects, i.e. objects with single direct geometry

- But INSPIRE data models include more complex patterns
  - Objects with generic geometry (GM_Object, GM_Primitive)
  - Objects with multiple representations (AD, BU, CP, ...)
  - Objects with indirect geometry
    - Aggregate objects (set of other objects) : Road, watercourseLinksequence, ...
    - TN properties attached to geometry by linear referencing
  - Objects with no geometry at all : AD components

- => **need for new rules for persistent identifiers**

the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

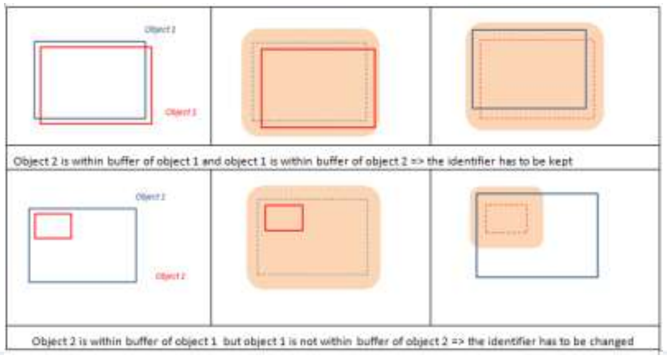
Persistent identifier: ELF rules

- General rule:

One of the identity characteristics of an object is its **classification:** in general, this classification is given by the **feature type**

NOTE : In some specifics cases, the feature type may be completed by a classification attribute (e.g. generic feature type such as NamedPlace, GovernmentalService).

- Basic rule: **geometry is an identity characteristic for spatial objects**
  - simple objects are homologous if they are in one another buffer



Object 2 is within buffer of object 1 and object 1 is within buffer of object 2 => the identifier has to be kept

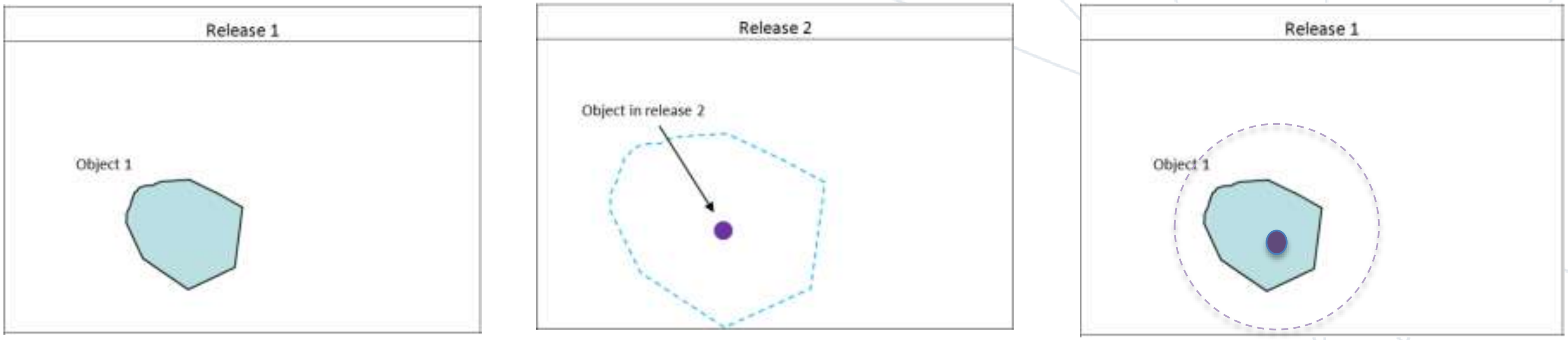
Object 2 is within buffer of object 1 but object 1 is not within buffer of object 2 => the identifier has to be changed

Buffer size is depending on data accuracy

the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

Persistent identifier: ELF rules

Generic geometry



A small lake was represented as a surface in release 1 and then as a point in release 2

★ Proposed rule : **geometry type is an identity characteristic**

- Even with staying in tolerance threshold, object 2 is considered as new object
- Change detection tool will compare:
  - points with points
  - lines with lines
  - surfaces with surfaces

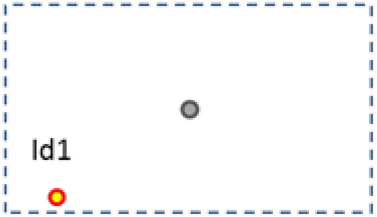
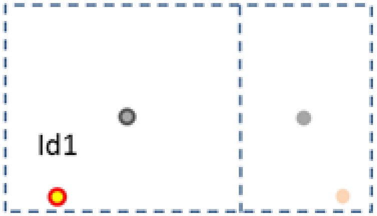
the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

Persistent identifier: ELF rules

Multiple representation

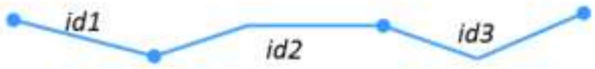
★ General rule:

The identity characteristic is **the reference geometry**

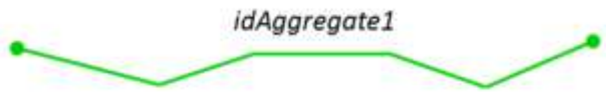
Release t1		The address has 2 geometric representations : ○ Entrance (default) ● Parcel (In this illustration, the "parcel" representation of address is the parcel's centroid).
Release t2		The parcel has been split: - Its reference geometry (entrance) remains the same => the identifier is kept. - its centroid ("parcel" representation of the address) has moved significantly, more than the tolerance => life-cycle information is updated (new version ...)

Persistent identifier: ELF rules

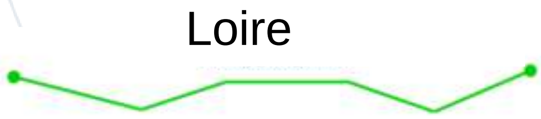
Indirect geometry: aggregate objects



Scenario 1: **set** of objects



Scenario 2:object **with derived geometry**



Scenario 1: object with **thematic identifier or Geographical name(s)**

★ Proposed rules :

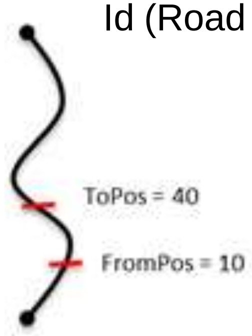
- **thematic identifier or geographical name(s) is an identity characteristic**
- Coarse geometry checking (in case thematic id or name not unique)
  - by ”wide” bounding box

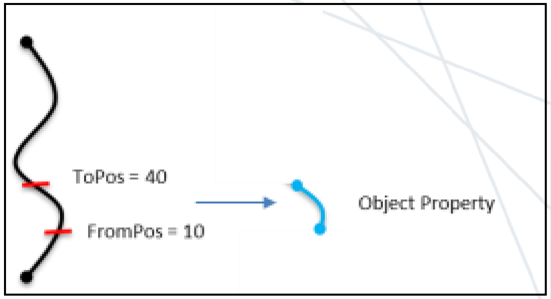
the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

Persistent identifier: ELF rules

Indirect geometry: TN properties

Id (Road element) =12





Scenario 1: **relative location**

Scenario 2:object **with derived (absolute) geometry**

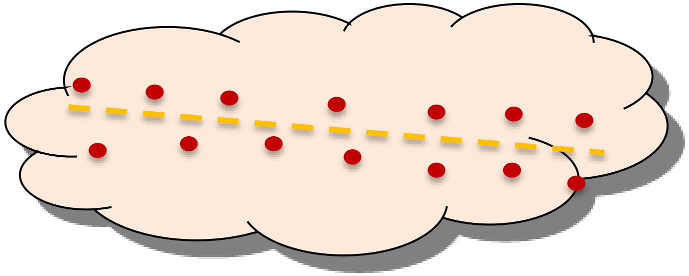
★ Proposed rule :

– scenario 2 : **absolute geometry is an identity characteristic**

the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

Persistent identifier: ELF rules

No geometry: AD components



In INSPIRE, ThoroughfareName has no geometry at all

But this geometry may be approximated by the cloud of AD points

★ Proposed rule for aggregate objects) :

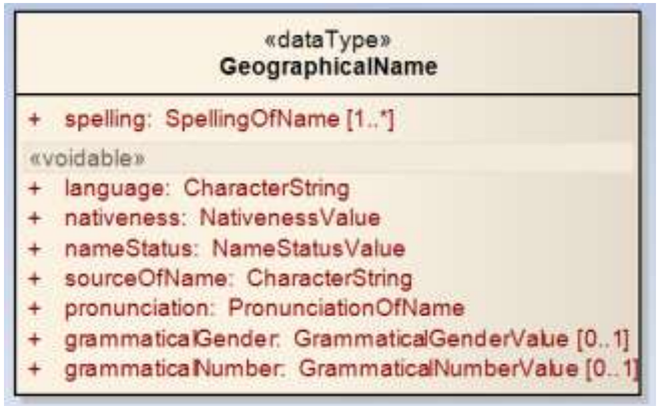
- **thematic identifier or geographical name(s) is an identity characteristic**
- Coarse geometry checking (in case thematic id or name not unique)

the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

Persistent identifier: ELF rules

Geographical names

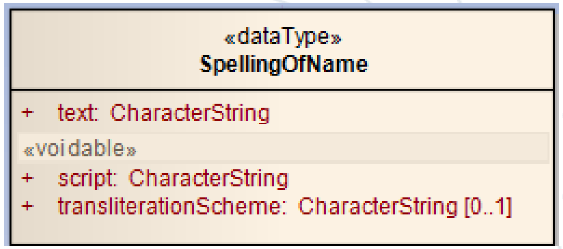
★ What does it mean : same geographical name(s)?



«dataType»
GeographicalName

+ spelling: SpellingOfName [1..*]

«voidable»
+ language: CharacterString
+ nativeness: NativenessValue
+ nameStatus: NameStatusValue
+ sourceOfName: CharacterString
+ pronunciation: PronunciationOfName
+ grammaticalGender: GrammaticalGenderValue [0..1]
+ grammaticalNumber: GrammaticalNumberValue [0..1]



«dataType»
SpellingOfName

+ text: CharacterString

«voidable»
+ script: CharacterString
+ transliterationScheme: CharacterString [0..1]

- A feature may have one or several name(s) with attached metadata
  - » Selection rule for **reference** name(s) up to NMCA (e.g. status = official)
  - » **Key metadata** : language + nativeness
- A name may have several spellings
  - » Reference spelling : original one (without transliteration)
- Small change in spelling is acceptable

the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

Conclusions

- ★ INSPIRE mechanism (persistent id + temporal attributes) remains ideal solution ; but it's achievable only if
  - ★ Data producers manage temporal aspects in source data
  - ★ Data producers give access to whole data set including historical data
  - ★ Source model is close to INSPIRE one

=> ... So it is not achievable by all data producers on short term

- ★ The change detection tool developed by ELF project may provide short-term solution (but not ideal one)

the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

Conclusions

★ Life-cycle rules:

  ★ Are currently up to each data provider

    – Existing life-cycle rules are mainly for simple features with direct single geometry

    – INSPIRE schemas include more complex modelling concepts

    – ELF has proposed rules for these specific cases

  ★ Potential interest to propose more harmonised rules in INSPIRE ?

    – ELF rules might be used as starting point in MIG to develop guidelines about temporal aspects

the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)