Mapping the Darwin Core with a national data exchange standard: feedbacks and issues

Julie Chataigner - julie.chataigner@mnhn.fr
Project manager « Standardization of biodiversity data » SINF

For the « SINF Standardization of biodiversity data » Working Group

October 2014
Summary

1. Context

2. Mapping issues

3. Conclusion
Context

The Information System Nature and Landscapes on biodiversity (SINP)
• Led by the French Ministry in charge of ecology

Goals:
• Dissemination of Public Data
• Reuse of these data to compute indicators for reporting on the status of environment: i.e.:
  – to compute indicators for the European reporting,
  – to enrich the Natural Heritage National Inventory (INPN), ...
• **MNHN** is in charge of developing a data exchange standard about bio and geo diversity in France

• **A working group** is composed by the main biodiversity stakeholders in France
Elements of the standard:

- Focus on the observation (occurrence of taxon): “A taxa is observed at a specific place and time by given observers”
Context

- Elements of the standard:
  - Focus on occurrence of taxon
  - Identified by a persistent identifier: URL + UUID
  - Modular, extendable by:
    - Thematic extensions (defined by the Working Group)
    - Additional attributes (defined by the users)

The SINP Standard

Core

IdPersistent

AddAttributes

Extensions
Context

- **Elements of the standard:**
  - Focus on occurrence of taxon
  - Identified by a persistent identifier
  - Modular
  - Documented by a data dictionary
  - Format: GML schema

```xml
<xs:schema elementFormDefault="qualified"
  targetNamespace="http://www.naturefrance.fr/sinp/" version="0.8.2">
  <xs:import namespace="http://www.opengis.net/gml/3.2"
    schemaLocation="http://schemas.opengis.net/gml/3.2.1/gml.xsd"/>
  <xs:element name="AttributAdditionel" type="sinp:AttributAdditionelType">
    <xs:complexType name="AttributAdditionelType">
      <xs:sequence>
        <xs:element name="valeurAttribut" type="xs:string"/>
        <xs:element name="uniteAttribut" type="xs:string"/>
        <xs:element name="thematiqueAttribut" type="xs:string"/>
        <xs:element name="nomAttribut" type="xs:string"/>
        <xs:element name="definitionAttribut" type="xs:string"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

V1 December 2013

V beta April 2014
• The SINP standard is congruent with:

  – The European Inspire Community:
    • Mapping with Observation & Measurement guidelines (work in progress)

  – The GBIF Community:
    • Mapping with Darwin Core
## Context

Differences between DWC and SINP:

<table>
<thead>
<tr>
<th></th>
<th>DWC</th>
<th>SINP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definition</strong></td>
<td>Less strict</td>
<td>Strict with specific and accurate rules</td>
</tr>
<tr>
<td><strong>Mandatory terms</strong></td>
<td>5</td>
<td>13 at least (it depends on conditions)</td>
</tr>
<tr>
<td><strong>Controlled vocabulary / checklist</strong></td>
<td>Some terms – recommandation</td>
<td>All terms - mandatory</td>
</tr>
<tr>
<td><strong>Cardinality</strong></td>
<td>0..1 – simple format</td>
<td>Multiple - hierarchical</td>
</tr>
</tbody>
</table>

These differences impact the mappings:
- DWC → SINP
- SINP → DWC
From SINP to DWC
issues and choices
Exchange the minimum of information

• Fill the mandatory terms required by DWC
  – institutionCode, collectionCode, catalogNumber, scientificName, basisOfRecord
  – Use the recommended controlled vocabulary of DWC

  Has been done for 7 years

• Fill the terms that can alter the understanding of the information
  – OccurrenceStatus: indicate if the taxa is observed (present) or not (absent) during the event (observation)
Issue 1. Optimize the work of data managers

• Find the right balance between standardization efforts and dissemination of reusable information

• In the line of thought that unstructured data is worst than no data
  – Difficulty to reuse
  – Volumetry of data ++++ versus added value +

➔ The DWC terms based on concatenated information are excluded from the mapping
Ex : dynamicProperties, higherGeography...
Issue 2. Map a hierarchical standard with a simple flat standard

• The habitat term example
• Habitat

A category or description of the habitat in which the Event occurred.

SINP

<table>
<thead>
<tr>
<th>Occurrence is in * Habitat</th>
</tr>
</thead>
<tbody>
<tr>
<td>idHabitat</td>
</tr>
<tr>
<td>refHabitat</td>
</tr>
</tbody>
</table>

DWC

| Occurrence habitat        |

The mapping is not done.

Loops + links with registries

Concatenated data
However if the query is simple : the mapping is done
Example : individualCount

<table>
<thead>
<tr>
<th>SINP</th>
<th>DWC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occurrence</td>
<td>Occurrence</td>
</tr>
<tr>
<td>objectCounted</td>
<td>individualCount</td>
</tr>
<tr>
<td>: codeList count</td>
<td></td>
</tr>
</tbody>
</table>

Simple query

\[
\text{objetCounted} = \text{« individual »}
\]

The mapping is done
Issue 3. Make the SINP data visible on the GBIF website

- kingdom
- phylum
- class
- order
- family
- genus

National CheckList **TAXREF** is published on GBIF
ID: 0e61f8fe-7d25-4f81-ada7-d970bb2c6d6
Issue 3. Make the SINP data visible on the GBIF website

Filter of basisOfRecord – GBIF website

- Specimen
- Fossil
- Living Specimen
- Observation
- Human Observation
- Machine Observation
- Material sample
- Literature Occurrence
- Unknown evidence

controlled vocabulary of basisOfRecord

→ How to select data where basisOfRecord = Occurrence?
Issue 3. Make the SINP data visible on the GBIF website

Occurrence must have «decimalLatitude» and «decimalLongitude» terms (i.e. x,y)

SINP: Location:
1. at the spatial object (point, line, polygon)
2. by reference to a spatial object like municipality, ... (identifier)

But not X, Y terms

We create and share the centroid

The length of the Bounding box

The spatialObject

decimalLatitude and decimalLongitude

coordinateUncertaintyMeters

footprintWKT
Issue 4. Promote actors involved in the data production

We met difficulties to exchange both the actors of the primary data production and the actors of the standardized data production
From DWC to SINP
DWC → SINP

• Focus on mandatory terms of SINP standard
  – date, taxon name, localisation, organisms of observers and traceability of the verbatim data
    = 23 dwc terms

→ Only bring back terms that totally match the SINP terms

Exemple : ReferenceBiblio and associatedReferences are close but not the same

Describe the occurrence

Can describe the occurrence or the taxon or any document linked in a way or another to the occurrence
To conclude

• Document (in french) is available:
  http://www.naturefrance.fr/actions/groupe-de-travail-standardisation-des-donnees-biodiversite-du-sinp

• Update cycle:
  – Sinp: once a year (January)
  – Dwc: ?

The mapping will be updated once a year after the update of the SINP
Thank you for your attention

http://inpn.mnhn.fr/accueil/donnees-referentiels

http://www.naturefrance.fr/actions/groupe-de-travail-standardisation-des-donnees-biodiversite-du-sinp