City-Zen Open liked data for more Visible Historical and Cultural Heritage

Prof. Basma Makhlouf Shabou
University of Applied Sciences of Western Switzerland
HES-SO GENEVA

Prof. Maria Sokhn
University of Applied Sciences of Western Switzerland
HES-SO Valais Wallis

Archival preservation and assessment: Relevant Scientific & professional Initiatives
ALA-ICA 2017 – Mexico City
Outline

- Context
- Motivation
- Challenges
- City-Zen system
- Conclusion & Future work
Context

- Lead
  - University of Applied Sciences (HES-SO Valais - Geneva)
- Partners
  - HES-SO IS-NET
  - Hasler Foundation
  - State archives of Valais Switzerland
  - Digital Switzerland
Motivation

30 years ago...

How neighborhood looked ..?

One single application
Motivation

- Information systems
  - Mostly rely on up-to-date content on attractive places
  - Additionally use multimedia digital archives
    - Historical data: photos, films, documents

- Interesting to explore evolution of cultural interests over time
Challenges (1/3) - Multimedia retrieval

- Support different queries

Simple location queries

Temporal queries

Spatial queries

Multimedia similarity queries

Radar mode
Challenges (2/3) - Entity recognition and Linking

- Different sources and formats
  - Schema and data integration
- Entity recognition
- Linking with external data sources: e.g. DBpedia
Challenges (3/3) – Crowdsourcing & Quality control

- Enable provision of user content data
- **Motivate** participation
  - Incentives
    - *intrinsic* or *extrinsic* motivations
- Control the *quality* of crowdsourced data
City-Zen high level system design

Provide integrated interactive access to heterogeneous contents

Datasets → Data integrator → Multimedia retrieval engine → Content enricher and curator → Entity recognition and linking → Frontend

Frontend

Data visualization
Dataset

- Uses metadata
  - 2 datasets (*DigitalValais* and *MediathequeValais*)
  - Three languages (French, German and Italian)
  - Title and description
Dataset - Model
Dataset - Model
Dataset - Linking

CulturalInterest :locatedIn City

prefix dbpedia: <http://dbpedia.org/ontology/>
select distinct ?place
where {
?place rdf:type dbpedia:Place
}
Dataset – triple store design
Dataset – Quality control

Information Quality

General Information Qualities applied on sources

Trustworthiness

Reliability

Authentication

Cognitive Accessibility

Technical Accessibility

Juridical Accessibility

Institutional Context

Socio-cultural Context

Exploitability

Representativeness

Uniqueness

Information Richness

Specific information Qualities applied on contents

11.2017
## Information Quality Levels

<table>
<thead>
<tr>
<th>Score</th>
<th>**</th>
<th>**</th>
<th>***</th>
<th>****</th>
<th>*****</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### General Information Qualities applied on sources

#### Trustworthiness

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Provided by 1 non-institutional information source</th>
<th>Provided by 1 institutional information source</th>
<th>Provided by 1 or more institutional information sources</th>
<th>Provided by 2 institutional and citizen's testimony information sources</th>
<th>Provided by institution and professional testimony sources with exhaustive description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authenticity</td>
<td>Provided by source non-identified</td>
<td>Provided by Citizen's authentic testimony</td>
<td>Provided by Authentic professional testimony</td>
<td>Provided by multi-source information</td>
<td>Provided by authentic multi-source information</td>
</tr>
</tbody>
</table>

#### Exploitability

<table>
<thead>
<tr>
<th>Cognitive Accessibility</th>
<th>French</th>
<th>French/ German/Italian</th>
<th>French/ German/Italian/ English/Spanish</th>
<th>Multi-language (7 languages)</th>
<th>Multi-language (more than languages)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical Accessibility</td>
<td>Easy to understand for the French tourist or citizen</td>
<td>Easy to understand for the French/ German/ Italian tourist or citizen</td>
<td>Easy to understand for the tourist or citizen lambda</td>
<td>All tourists and citizens can understand easily</td>
<td>All tourists can understand easily</td>
</tr>
<tr>
<td>Juridical Accessibility</td>
<td>Information is accessible but non-consultable physically</td>
<td>Information is accessible but non-consultable physically</td>
<td>Information is accessible but non-consultable physically</td>
<td>Information is accessible but non-consultable physically</td>
<td>Information is accessible but non-consultable physically</td>
</tr>
</tbody>
</table>

#### Representativeness

<table>
<thead>
<tr>
<th>Institutional Context</th>
<th>Recommended by Office of Tourism but information not up to date (of last year)</th>
<th>Recommended by Office of Tourism but information not up to date (of last year)</th>
<th>Recommended by Office of Tourism and information is up to date</th>
<th>Recommended by Office of Tourism information is up to date</th>
<th>Recommended by Office of Tourism information is up to date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socio-cultural Context</td>
<td>Information is not scarce, in a context summary (abstract)</td>
<td>Information is not scarce, in a context summary (abstract)</td>
<td>Information is scarce by the exhaustiveness of the context</td>
<td>Information is scarce by the exhaustiveness of the context</td>
<td>Information is scarce by the exhaustiveness of the context</td>
</tr>
</tbody>
</table>

### Specific information Qualities applied on contents

#### Score

<table>
<thead>
<tr>
<th>**</th>
<th>**</th>
<th>***</th>
<th>****</th>
<th>*****</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Uniqueness

- Density of information, lacks of precision
- Abundance of format, media, information typology, subject, period / context
- Density, lacks of precision
- Abundance of format, media, information typology, subject, period / context
- Selection of information, lacks of precision
- A selection of format, media, information typology, subject, period / context
- Exclusive Information, precise and indexed
- Scarcity of format, media information typology subject, period / context

#### Information Richness

- The Information will be searched by personal words, there are no feedback available, the information will be limited audio, with an Impersonal source (ex. Tourism Office)
- The information will be searched by a list of keywords, the feedback will be slow but available, the information will be limited in Visual, with an Impersonal source (ex. Tourism Office)
- The information will be searched by a list of Keywords, the Feedback will be slow, the information will be diffused in Visual or in Audio, with an Impersonal source (ex. Citizen’s testimony and Tourism Office)
- The information will be searched by a Numerical Index, the Feedback will be Fast, the Information will be diffused in Visual and Audio, with a Personal and an Impersonal source (ex. Citizen’s testimony and Tourism Office)
Dataset – Quality control

Example: information reliability metrics
Define the time interval

1953 2017

EXPLORE CULTURAL INTERESTS
1965

Assemblée générale

1979

Un moment de récréation dans une vie très active : la brisolée annuelle !

Section féminines
Conclusion

Datasets

Data integrator

Multimedia retrieval engine

Entity recognition and linking

Content enricher and curator

Integration of all 3 parts

Frontend

Evaluation on the full integration
Future developments...
(1) Crowdsourcing

- Issues with the existing data
  - Part of the data is *not complete*
    - *Missing* information: GPS, subject, location
  - *Conflicting* data from different sources
    - e.g., two different dates for the same event

- Initiative for collecting new digital data from individuals
  - Many individuals have data about the region they live
(2) Project extension

- Enable a more performing quality assessment focusing on the metadata use
- Enabling a multilingual functionality: over the structured and non structured extracted data
- **Historical and Cultural Heritage exploration based on geolocalisation** (Radar mode service)
- Relevance assessment of City-Zen when applied in pilot cities
QUESTIONS?

Prof. Basma Makhlouf Shabou
basmamakhlouf-shabou@hesge.ch

Prof. Maria Sokhn
maria.sokhn@hevs.ch

Thank you!