



europeana
think culture

Multilinguality from an European(a) perspective

Sjoerd Siebinga

MLIA4DL, Trento, 09-09-209

Overview

- What is MLIA (MultiLingual Information Access)
- What can we learn from the past
- Sanity Check
 - what can we realistically achieve
- Where does MLIA belong in the Design process
- What kind of MLIA system does Europeana need
- Requirements for the White Paper

What does MLIA mean to me?

- Transcending linguistic and cultural boundaries in knowledge acquisition.

Which MLIA components can we identify

- Interface translation
- Browsing with controlled vocabularies
- Searching on a monolingual baseline
- Query translation and expansion
- Result translation

Where have we come from?

- Good:
 - The EU commission made multilinguality a spearhead of their IT frameworks.
 - Each project accepted needed to do something with MLIA
- Bad: (We all seem to be flogging the same dead horse)
 - No overall coordination from the EU commission to encourage or enforce cross-project and cross-funding-round collaboration
 - No organizational body to take the result from the projects and move it to the next level.
 - Everyone is exploring the same subjects and abandoning them when follow-up funding is denied.



I think we are here now because of this aim.

Some where more successful then others => Micheal with their collection descriptions translations claiming to have a fully multilingual solution

Where are we now?

- MLIA software is pretty mature and delivers usable results
- MLIA resources are incomplete and often very expensive
 - This is – in my opinion at least – the major oversight in the vision of the EU commission.
 - Just fund one project with a reasonable budget (+20 mln euro) that does two things:
 - 1) define the standards to be used for all the language resources.
 - 2) create the language resources for all official EU languages.
 - Setup a small European agency that maintains this data and makes it freely available until the world ends.



With all the money in the world there is not a software company that can give us pan-europeana coverage.

This is logical if you think about the business sense it makes to invest heavily in small european languages like slovenian

What are the key stumbling blocks to move ahead

- Lack of translation and NLP resources for all European Union languages
- Lack of understanding by stakeholders of the commitment required to be successful
 - Providers
 - Business
 - Developers
 - Users



In the next slides I will go into the details

language tags in each content field, attributes specifying which controlled vocabularies they use, making the controlled vocabularies freely available, documenting their cataloguing rules

long term commitment, MLIA will stifle rapid agile innovation

MLIA needs to be part of the core design as an add-on it will simply not work

Maybe the user interaction should be more central to the design

What do we need from the Providers?

- Language attributes for each content field
- Attributes specifying which controlled vocabularies are used for each controlled field
- Making the controlled vocabularies freely available (in SKOS?)
- Documenting and sharing their rules for creating metadata

Of course using a standard vocabulary

This is especially important when you keep in mind that the interoperability that Europeana nows has is mainly syntactic

Business needs to understand:

- What the real cost are of any software component
 - 20% R&D and build; 80% maintenance, e.g. update, extend, etc
- That updating, improving and extending ML resources needs permanent skilled staffing and budget
- That implementing a MLIA solution is a product life-time commitment
- That MLIA prerequisites are also the basis of functionality they really want
 - For example: named entity extraction, co-referencing, vocabulary alignments, geo-temporal extraction and browsing, etc.

I can't stress this subject enough. Often even permanent development staff is treated as a necessary evil.

The operational cost will never go away and will not substantially go down

Developers need to understand:

- That MLIA is not an add-on, it needs to be part of the core-design (Martin Braschler).
- There are no **standards**, because everyone implementing a standard deviates from the standard.
- The world is schema-less: lets embrace it
 - The system needs to be designed with scalability and extendibility in mind.
 - Instead of specifying everything lets work from the premise that we ignore what we don't need.



This is the reason why intergatiing oai-pmh is still so much work.

Instead of specifying everything lets work from the premise that we ignore what we don't need

Convention over configuration. using overhype word agile

- Open, Social, Linked => Interfacing the user
 - What does Google do: massive query expansion under the hood.
 - What does Apple do: focus on the right level of abstraction and user interaction
- User interaction design should be more important
 - Why do the iPhone and iPod have such an enormous market share
 - Because they revolutionarized the way users interact with the system

Apple stuff we need to focus more attention on the interface we want to have with our users

You can say a lot of good things about google, but not that they are the most gifted interface/ user interaction designers

50/50 percent distribution between engineers and user interaction designers

Making it easy for the user to do the right thing

Why do developers choose certain libraries over others. Because they offer a better level of abstraction to interact with the underlying functionality

Europeana Requirements

- The MLIA system must be able to be maintained and extended by non-developers
 - Therefore, the Europeana MLIA component must provide tooling for the complete workflow
- The system must be clusterable and replicatable.
- Convention over Configuration design: for example, being able to add languages without recompiling the code.
- It must run on a Linux system, preferably on the Java VM.
- Documentation, technical hand-over, needs to be part of the roadmap.



No hard coding MLIA support

Requirements for the White Paper (1/2)

- Describe realistic requirements and assumptions to work under
 - We must be really honest about what we need to acquire success and long-term sustainability
- Roadmap for commitment from the business to MLIA
 - Scoping expectations and explaining the pervasive integration of MLIA in the Europeana Operational Service

If we take the above into account. What do we need. This white paper must be our sanity check, maybe even our service contract

Requirements for the White Paper (2/2)

- Building Europeana as an Information hub
 - The KOS needs to be part of the core system in order to be sustainable and reusable for future extensions
- Finally, the Europeana MLIA workflow needs to provide tools to interact with the complete workflow, i.e. from preparing dictionaries, dataset preparation to end-user interaction.
 - In other words, how are we going to support the 80% maintenance.

If we take the above into account. What do we need. This white paper must be our sanity check, maybe even our service contract

Thank you very much for you attention

- Any questions?