

The presentation, which will be given to you in the next minutes, has the title "Registry of the Fine Arts Collections", but the scope is broader than to show you only another resource for browsing the art web. Because such data sources are not only about presentation and attractiveness. All that data must be created first of all, and the accurate curatorial work can not be replaced by some harvesting strategy.

We have accustomed to the fact, that the same things in the real world (and even in museum collections) can have different names. This is common for geographic places, characters, tangible things and many other entities.

Moreover, every name can have different written forms reflecting the order of its subsequent parts and can also look totally dissimilar when using various scripts common for many languages mankind speaks. But these common facts are sometimes very complicating the data retrieval in nowadays world.

For example, lets try to search the greatest on-line source of the European cultural content for some artifacts connected anyway with Vienna. Vienna was the capital of Austria-Hungarian monarchy till the end of first World War and many of our grandparents were born into this multinational state. German language was the official one and widely spoken, thus we should try first with this.

We are able to query more than sixty thousand results using

the original German name of Vienna. It's pretty logical and we can suppose that many of them will be also relevant.

Bécs - Hungarian name for Vienna gives much less results despite of the fact that this language was the second most widely used, especially in the eastern part of Austria-Hungarian Empire. This can simply reflect the inferior contribution of Hungary to the Europeana's data pool, but shows clearly the weakness of disconnected name variants.

For the interest of organizers - Romanian name. This returns very similar number of results as in the previous case. Please pay attention to the representativeness of objects found. Such gauntlets really don't look smart enough to be integral part of classical Viennese fashion.

Trying to find anything about Jewish Vienna using Yiddish is completely pointless in Europeana despite of her integration with Judaica Europeana project and flourishing culture of more than quarter million of Jewish inhabitants reported from the pre-war Vienna.

The librarians have already found the answer. They call it authority files and use them for co-referencing many various names with their single chosen (authoritative) form.

Our project to introduce the good of authority files to the museum people had been already presented 3 years later in Athens.

Now only in short. National Authority Files in National Library had been recognized as insufficient for the needs of museum curators and archivists too.

Hence we aimed at interfacing original librarians' authorities and museum needs. During the work on this project, Association of Galleries turned to us with the request for help. And now, definitely, we get to the Registry of Fine Arts Collections.

This facility existed from the beginning of eighties, of course in the form of paper cards, without any notion about modern technologies. In the beginning of new millennium, galleries started to miss its benefits because the Central Registry had been set aside in-between. Three years ago, Association of Galleries decided to recreate the registry as newly based,...

...fully computerised system which could be able to not only help with organisation of exhibitions (on the side of curators), but also provide the abundant cultural content to the public.

After the initial cleanup of individual datasets in the local collection databases of 26 galleries, all the datasets had been merged and deduplicated. At the end we have more than 80 thousand works of art caught up in the database and some six and half thousand artists.

One third of all artist names were duplicates and it took a lot of time and effort to eliminate them. For example here: the pre-war entertainer, actor and painter Emil Arthur Pitterman-Longen is known under at least two pseudonyms, his long name practically eggs on misspellings and even his birthdate is not very clear.

But to be sure, we have authority files in hand. And now even in the enhanced, museum-friendly form, thanks to the previously mentioned project.

This is very simple search interface to the museum authority files. Such an "Czech museum Google".

We can search for terms in different connotations.

For example the name of famous Czech writer and journalist Karel Čapek.

And here is the output. In the pull-down menus on the top you can choose its format - this is detailed view in the form acceptable for museum people. There are different views for librarians, but they look horribly, at least for me. Just after the preferred form of the name you can see other forms and pseudonyms. They can be used for searching too and would lead you to the same record! Small symbols of magnifying glasses indicate the authority linked fields. Clicking to them will bring you to the appropriate record of the authority referred. Well, lets browse to the Čapeks brother Josef.

His authority record looks very similarly and the downward reference to Karel Čapek is in evidence. Many magnifying glasses point to geographical, personal, corporate or subject authorities.

This authority record is provided with comprehensive curriculum vitae describing such details as the artistic development and personal life of Josef Čapek. At the end, links to some external sources lead directly to the pages dedicated to Josef Čapek there.

In wikipedia,...

...there is significantly shorter info in English than in Czech of course. But we can find the interesting fact, that Josef Čapek invented the word "robot" at least.

Authorities of The National Library of Czech Republic. They are the original authorities taken as a base for enriching them in museum authority files.

You can see that only few fields essential for identification and librarians' work are here.

In the joint catalogue of National Library...

...we can find almost six hundred different books written BY or written ABOUT Josef Čapek.

AbART database is another source...

...which provides comprehensive information especially about exhibitions and exhibition catalogues arranged with works of particular artist.

And definitely the Registry of Fine Art Collections is referenced from here...

...to show all of 167 Čapek's fine art works.

Here in detailed view, with authoritative form of his name replacing the original (and misspelled though) name entered by a curator.

But how do the names get into such authority database? Who and in which manner connects the authoritative forms of names with the unpreferred ones? Here you can see screenshots of the original client for database update. Nothing what could be called "user-friendly" and full of librarians' MARC tags.

The editor in use recently doesn't only look better and up-to-date, but provides the full set of authority fields which can be edited via web browser directly during the work on the Registry.

For technically savvy: the museum authority record is not stored in simple MARC format, but the MARC XML with a lot of CIDOC CRM structuring had been implemented with cooperation with our technological partner.

The museum authority system is ready for the working use by now. There are two slightly different approaches to it. Authorities can be used during the curatorial description of objects, while processing them in collections. This is very slow, but systematic and accurate. The other approach is much popular nowadays - at least from the point of view of oversized EU-funded projects. Authorities can be simply exploited for postprocessing of huge, but poorly standardized datasets. Combination with relevant heuristic analysis and operator supervision can help to solve even such problems as Finnish gauntlets in search results for "Vienna".

A bit of applied museology and futurology at the conclusion. The way how to transmit the information from museums to cybernauts was pretty straightforward in the past. Web content had been simply recycled from collection documentation.

Now we are trying to use authority files at documentation process and web content is no longer simplistic documentation remake.

Web 2.0 brings the phenomena of re-generating web content and we must to enhance our content creation abilities with massive use of authority files. Failing to it would implicate relative lowering of A-M-L share in the information world.

The idea of Web 3.0 explicitly builds on the semantic bonds in web environment. And the existence of shared authorities is fundamental for this downtrend.

If the cloud wins and museums fail, their role will be unenviable, with minimal status and irrelevant influence on the community.

But if we succeed in adapting ourselves in the changing world, the web democracy will work for us and our mission.

At least, we should add that small magnifying glasses to all authority link-able fields in the meanwhile. The web 2.0 and web 3.0 will struggle with us then, not against.

END