
LEOPOLDINA ONLINE – IN SITE DIGITAL CONTENT AGGREGATING SERVICE

Piotr Malak; piotr.malak@uwr.edu.pl; (University of Wrocław, Faculty of Letters, Main Library)

Increasing amount of digitised content is a very promising fact. It promises a wide and easy access to digitised pictures of valuable content and artefacts. But reality shows us it may be contrary – we encounter problems searching digitised content. Metadata aggregating services solve the problem. Splendid example of such is Europeana¹ providing universal access to European digital cultural heritage resources. Locally we may know digital libraries metadata aggregators, such as Polish FBC (Federation of Digital Libraries)². Very functional they have limitation, though. They provide access to limited, in terms of formats, sources. Organisations aiming towards digital transformation face a new challenge – combining variety of available digital content formats. An attempt to address problem of access to different digital resources in a Leopoldina online³ platform. Developed by University of Wrocław (UWr)⁴, in cooperation with PSNC⁵ – Polish national meta data aggregator for Europeana, it aims to aggregate and deliver digital resources of various UWr units. In current paper we present goals of the project, short description of different digital resources available, metadata handling and common, universal search engine design.

<https://doi.org/10.52036/1335793X.2021.1-2.34-37>

Digitisation is a not underestimatable initiative of preserving and delivering cultural heritage, output of scientific research, memories and many, many other remarkable as well as trivial content. One may assume that an early phase of individual, focused, single goal oriented digitisation is already over. Except of traditional enormous (in terms of volume) web content (websites etc.) we can observe increasing amount of digitised content. In order to secure fluent and convenient access to those resources a big picture approach is required. An example of which is given by Europeana – a pan-European metadata aggregator. Other initiatives of that kind are digital libraries aggregators. Despite of their maturity and proficiency in delivering digital content there are limitations of their functionality. However extended, metadata aggregators are still limited to process certain resources – mostly digitised books and documents.

Leopoldina online, a project of University of Wrocław, with PSNC as subcontractor, attempts to handle digital representations of any objects, despite of digital files format, form or size, and despite of different, dedicated standards of metadata. The project aims to digitize and popularise online valuable scientific resources of the University of Wrocław and launch digital services that allow increasing the intensity and efficiency of exploitation of the scientific

achievements of the University of Wrocław. A global action of digitising, substantive development and online access to ca 23 thousands of digital objects was set as project deliverables.

In parallel to project preparation, an global survey was conducted among University units. The survey was the first ever attempt of analysing the volume, types and partially state of various Cultural Heritage and scientific resources stored in different units and places. Results gave us an overall insight into index, administrative and physical locations of resources. The survey was also used to communicate Leopoldina online aim to wide public of University staff, and to collect statements of will to contribute to project. Finally 16 following units (in alphabetical order) applied with their resources:

1. Archives of the University of Wrocław,
2. Botanical Garden,
3. Cartography History Workshop,
4. Department of Climatology and Atmospheric Protection,
5. Department of Geoinformatics and Cartography,
6. Department of Physical Geography,
7. Geological Museum,
8. Institute of Archaeology,
9. Institute of Art History,
10. Institute of Geography and Regional Development,

11. Institute of Geography and Regional Development,
12. Library of the Historical Institute,
13. Mineralogical Museum,
14. Museum of the University of Wrocław,
15. Natural History Museum,
16. Wrocław University Library.

In short, the first task of the project was to adapt or establish room with professional equipment, appropriate to types and kinds of reported resources (like flat or 3D scanners, micro scanning heads or digital cameras). This task enables launching wide digitisation program of different resources of UWr, including nature exhibits, fossils, stones, minerals, etc. resulting in 66 thousands of digitised object available in a new created Digital Museum of UWr.

Another task was preparing integration layer for different presentation systems used for digital resources of UWr. Those are following systems (in alphabetical order):

1. AtoM (Digital Archive),
2. dLibra (Digital Library and Repository),
3. dMuseum (Digital Museum),
4. Huesca (Scientific Information System),
5. Midas Browser (hierarchical dictionaries (thesauri) creation system and CH objects registration system).

As an aggregating service Leopoldina online actively, almost in real-time mode, gathers new metadata sets from the source digital services, presents them to users in an uniform, basic form, and redirects to original objects in source systems. The source systems are responsible for delivering the full metadata descriptions and for appropriate presentation of most of aggregated digital objects.

A short descriptions of above mentioned digital services (systems) and their role is given in the next section.

INPUT RESOURCES PRESENTATION AND DATA AGGREGATION

University of Wrocław, recognised as a research one, is aiming towards data-driven research university. This direction is a result of innovation, research and development policy of the university authorities.

One of the ways to reach this goal is providing online access to wide scope and broad volume of publication and research data of the scientific output of University activities. In order to deliver scientific production to wide public UWr developed policy of encouraging its staff to publish in digital form as big part of their work as possible. Common platform for providing digital, scientific, content is UWr repository, available at: <https://repozytorium.uni.wroc.pl/>.

Currently the main source of digitised objects are resources of university libraries, Main and units. It has been established as an initiative of Wrocław University Library in order to enable preservation and popularisation of reach library resources recognized as the cultural heritage. Except of digitised books and manuscripts it presents also cartographical resources, music materials, drawings, etc. Digital Library contains over 110 thousands of digital objects.

University of Wrocław is also in a comfortable situation of academic institution possessing very rich resources of valuable, quite often unique, scientific and cultural heritage. Chosen examples of collections are available at Digital Museum of University of Wrocław portal: <https://www.muzeumcyfrowe.pl/dlibra?language=en/>. It serves as exhibition point for digital resources of, among others, Museum of University of Wrocław, Geological Museum, Mineralogical Museum, Museum of Natural History, Institute of Archaeology, Department of Physical Geography or Botanical Garden. Scans and pictures of different types and kinds of resources are available through Digital Museum service. All together digital museological collections consist of 66 thousands of digital objects.

Another source of digitised objects is Archive of University of Wrocław. Digitised resources of the Archive are available under https://atom.bu.uni.wroc.pl/index.php/?sf_culture=en. As a serving point for digital archives AtoM system (Access to Memory: <https://wiki.ica-atom.org/>) has been selected and implemented. University Archive belongs to network of national Polish Archives, thus provides objects description according to national archiving rules and, as for use of AtoM – according to ICA metadata standards.

UWr cultural heritage digital data are stored and published by MIDAS-browser service, according to MIDAS-Heritage standards (<https://www.dcc>).

ac.uk/). The service is available under: <https://dk.bu.uni.wroc.pl/> URL, but unfortunately only in Polish.

Another, quite rich in content, source of digital data is Scientific Information System (SIS) – HUESCA, which registers current scientific publications of UWr staff. System acts only as data storing service for analytical purposes, thus there is no user-friendly interface for presenting the data.

All different data storage and/or presentation systems communicate with the envisaged aggregation service using common communication protocols. The dictionary data, as well as update requests use HTML GET protocol and XML and JSON files in order to send up-to-date data. Metadata of digitised objects are transmitted using OAI-PMH protocol (oai_dc and oai_qdc) with respect to METS and LIDO standards. For digital objects created as representations of nature exhibits as well as fossils and minerals we applied and DarwinCore metadata description standard.

DATA PRESENTATION

Different kinds and type of analog resources as well as different forms and formats of their digital representations require different end system with appropriate user interfaces. One of main goals of Leopoldina online project was to deliver an general service for presenting to wide public those different resources. In order to fulfil this aim a proper, universal presentation layer was required. However, the presentation layer was build on the data gathered by data acquisition layer.

First proposal of such layer was based on the best practices of ontology and thesauri creating. As a linking part CIDOC ontology was proposed. This attempt, meeting most requirements of semantic web, was, however, not user oriented. In theory appropriate approach, leading to semantic retrieval of information was quickly criticized by test users. One of the main problems was labelling layer, not corresponding to natural language, but to ontological description of ideas. The main objection of test users was common labelling of authors, contributors, publishers and other CH object-related persons under one label: name, which was obviously misinforming.

Using information architecture principles as well as UX approach we introduced user-focused labelling and data presentation layer. As the main source

of data for Leopoldina are publications (from SIS, Repository and Digital Library) the bibliographic description pattern was used as the leading one. Title and author(s) are given without any labelling. For other related persons their role is provided, eg. trans[lator], print[er], etc. Other key meta information provided in unified digital object presentation within Leopoldina online portal are: keywords, description, identifier, rights (access and usage), language of the original object (if applicable) and object type. The rest of metadata object description is available in the source services, with respect to standards used in particular systems.

Leopoldina online offers also universal, common indexing and searching services and facilities. This offers users possibility of corporate retrieval of digitised content by the following searching criteria:

- objects
- persons people
- institutions
- a separate subset of proper names
- the place where the artifact was found
- events
- collections
- group publications
- property rights
- advanced queries that combine keywords from different description fields
- free search, when the user has no special needs and is interesting

CONCLUSIONS

Aggregating metadata services, as Leopoldina online, appears to be reasonable answer for needs of delivering to users different types of digital content. On the information retrieval and access level they map different standards of descriptive metadata, and do index different metadata fields. On presentation level they deliver an universal layer for data publishing to the users. Developing of such aggregators require joining, synergically, UX research, information architecture analysis and design as well as clear and universal rules for descriptive metadata providing. Experiences with Leopoldina online development and implementation shows it is reasonable to set full spectrum data presentation to the source digital services (like digital library, repository, museum or archive) while aggregating vital, from retrieval and access perspective, metadata in one aggregating system. Such a system should

provide universal presentation template, keeping in mind decrease of information volume and value, in comparison to original source digital services. Further phase, having place currently at UWr, is tuning the system according to users experiences and expectations. We are redesigning indexing layer now, providing new mapping of different metadata fields in accordance to analysed users journeys.

ACKNOWLEDGEMENT

Leopoldina online was implemented as part of the Digital Poland Operational Program for 2014 – 2020, Priority Axis 2 “E-administration and open government”, Actions No. 2.3 “Digital availability and usefulness of public sector information”, Sub-measure 2.3.1 “Digital access to sector information public from administrative sources and science

resources”(type II of the project: digital access to science resources).

Implementation period: 1.11.2017 r. – 31. 10. 2020 r.
Project value: 14 040 2Project co-financing from the EU: 11 882 300, 67 PLN

References

1. Europeana: <https://www.europeana.eu>
2. Federacja Bibliotek Cyfrowych: <https://fbc.pionier.net.pl/?lang=en>
3. <https://leopoldina.pl/>
4. Uniwersytet Wrocławski: <https://uni.wroc.pl/en/>
5. Poznańskie Centrum Superkomputerowo-Sieciowe (PSNC): <https://www.psnk.pl/>

BUILDING TAXONOMIES ON THE BASE OF VOCABULARY IN REAL USE IN THE SYSTEM, LEOPOLDINA.PL USE CASE

Agnieszka Łętowska; agnieszkaletowska7@gmail.com; (Freelance programmer and linguist, collaborator of Leopoldina online project)

Taxonomies are attempts of ordering world in semantic field. They are also a useful tool of knowledge organization, facilitating domain information retrieval. Taxonomy or thesauri creation requires extensive and thorough analytical process, engaging information and knowledge from variety of sources. Typically it is a long-term activity. Such traditional approach failed for knowledge organization gathered on Leopoldina.pl platform. The aim of the service is to present heterogenic and differentiated digital resources of University of Wrocław (UWr). Due to lack of success of creation an universal taxonomy in given period of time, an bottom-up approach was proposed. The entire range of vocabulary indexed in Leopoldina.pl platform was divided into discipline categories which were used as the base for taxonomies and thesauri creation. In current paper we describe the bottom-up process of taxonomy creation. Programming tools (in Python) used for domain dictionaries creation are presented. We provided also evaluation of external organized knowledge sources (like Wikipedia, GBIF – the Global Biodiversity Information Facility and other domain-specific thematic portals) for automatic knowledge handling and taxonomies creation.

<https://doi.org/10.52036/1335793X.2021.1-2.37-41>

INTRODUCTION

Taxonomy is a hierarchical way of describing the world. For centuries, the term taxonomy was used only to describe the systematics of living organisms.

Nowadays, this term is also used to classify knowledge from various fields in Knowledge Organization Systems (KOS). In information websites, taxonomy is the division of website content into general