e-infrastructures austria · Deliverable

Metadata and Accessibility

e-infrastructures austria · Deliverable

Metadata and Accessibility

Work-Package-Cluster:	**Cluster I: Metadaten** **Cluster L: Projektübergreifende Fragestellungen (aus nicht-technischer & technischer Sicht)**
Leitung des Clusters:	Susanne Blumesberger (Cluster I) Andreas Jeitler (Cluster L)
Datum:	15.09.2016
Version	1.0
AutorInnen:	Andreas Jeitler Universitätsbibliothek Klagenfurt andreas.jeitler@aau.at Susanne Blumesberger Universitätsbibliothek Wien susanne.blumesberger@univie.ac.at

Kurzbeschreibung (Deutsch):	Metadaten und Barrierefreiheit
Description (English):	Metadata Accessibility / Accessibility Metadata

Schlagwörter (Deutsch):	Barrierefreiheit, Metadaten
Keywords (English):	Accessibility, Metadata

Table of Contents

1. Abstract

Digital media like e-books, electronic magazines or other kinds of electronic documents build the foundation to allow people with certain disabilities to take a self-determined part within the scientific community. For example, access to information for blind people has never been as easy as today. But not all documents may a priori be used by all users, since they may miss some accessibility features. Several attempts have been made to assure that electronic documents will be accessible by as many users as possible (web content accessibility guidelines (WCAG) [1], Section 508 [2]). Beside these considerations, the most accessible document would be of no use, if you will not find it within a repository. To ensure that also people with disabilities will find the documents they search we plan to investigate the accessibility of metadata, after all data that describes other data, and their capability to describe the accessibility of objects they describe.

2. Introduction

e-Infrastructures Austria [3] is a three-year partnership project, started in January 2014 and sponsored by the Austrian Federal Ministry of Science, Research and Economics (BMWFW) [4]. The project's objective is the coordinated establishment and development of repository infrastructures for digital resources in research and science throughout Austria. To achieve the different goals the project's workload was divided into work package clusters.

The project partners understand the importance of accessibility for as many users as possible and therefore installed a work package cluster on cross-project issues that discusses, besides other issues, the matters of accessibility and works on proposals and guidelines [5]. In cooperation with the "Metadata" cluster [6] (see also [7]) this workgroup investigates the ways how metadata could be defined to be as accessible as possible. Second basic interests are ways to describe the accessibility of the objects described by specific sets of metadata (accessibility metadata).

In this paper we would like to describe our intention to investigate how to apply basic ideas of given international accessibility guidelines on the concept of metadata and evaluate sets of accessibility metadata to ensure people with certain disabilities may determine if a given document would be usable and accessible by them. To answer these questions we would like to define some basic terms first.

3. Metadata

Metadata have been used in libraries for more than a century, for digital objects new metadata types have been adopted. Metadata fulfil some tasks like identifying items uniquely worldwide, describing the context of the objects, supporting retrieval, recording access permissions and rights information, recording technical parameters and so on.

So we can distinguish in

- descriptive *metadata*, the descriptions of what a digital object it is, when and by whom it was made, but also its name and where to find it,
- *structural metadata* tell about logical and structural relationships,
- *technical metadata* give information about file type, which hardware and software was used when creating the object, and which platform or software will be needed to use it. Technical metadata should also include things such as image size (images) or frame rate (video),
- *administrative metadata* include rights management and documented preservation activities. Any change or modification to the object should be included in administrative metadata. (see also [8]),
- *accessibility metadata* are a new kind of metadata that describe an object's ability or disability to be accessible for users with specific needs. Accessibility metadata lie somewhere between descriptive and technical metadata.

Metadata are an essential part in repository design, since they describe and structure the content stored within the repository.

Some theses about metadata:

1. Metadata help to find data, they support the search for the objects
2. Metadata give access to data, they regulate also the application to the objects
3. Metadata structure the data
4. Metadata give further information about the data
5. Metadata are valuable separately, they are sometimes also independent data
6. Metadata need standards, which allows to share them
7. Metadata help archiving and preservation of resources, because they give us the information about long term preservation
8. Metadata have to be accessible for all, also for disabled persons or persons with another first language
9. Metadata are data. They can be used as data from one person, and for another person they ae describing other data. It depends on the perspective. Per example a video can be used as describing metadata for a diagram or it can be seen as research data.

4. Metadata Accessibility

Within this paper we would like to use the term accessibility in regard to a person's ability or disability to use, perceive or find information due to a specific impairment.

But why would information be inaccessible to people with certain impairments? To answer this question let's first take a look on a basic model of communication introduced by Shannon and Weaver in 1949 [9]. The basic idea of communication would be the process of sending messages or transferring information from a sender to a receiver via a channel or media. A human being as receiver perceives information via his or her five senses.

If one sense is limited or even disabled at all (for example if you are blind and therefore not able to access visual information) other senses should compensate the loss. In case of blindness tactile and audio senses have proven to compensate the visual sense quite well. If a person lost his or her hearing abilities, the visual or tactile sense may serve as a compensator.

But even if a person is able to perceive information he or she may possible not be able to understand the message in a cognitive manner (people with learning disabilities) or the message gets lost or is altered on its way from sense to the brain (dyslexia etc.). And even in case all senses work quite fine and the message would reach the destination, it may be possible that a person would not be able to interact with a computer or other device, because he or she has issues with motoric control and therefore cannot use the device.

A person's impairment(s) and the resulting disabilities in regard to information-consumption are quite individual and hard to compare.

The simple model shown only reflects a one-way information transmission for ease of understanding. Within today's information society people with disabilities themselves are functioning as data sources. Therefore the authoring tools to create information must meet the same accessibility requirements [10].

In case of web accessibility we may rely on international guidelines [1] which promise help in achieving the goal of accessible content. Even PDF got an ISO standard to meet accessibility requirements called PDF/UA [11]. Luckily WCAG 2.0 was designed in a modular manner as general guidelines which may be applied to specific technologies. Implementations exist for HTML, CSS and even Text etc. There is no implementation for metadata yet, but we will try to work on that.

4.1 Four principles of accessibility defined by the WCAG 2.0

The Web Content Accessibility Guidelines depend on four principles. Content must be:

- **Perceivable:** Information and user interface components must be presentable to users in ways they can perceive.
- **Operable**: User interface components and navigation must be operable.
- **Understandable**: Information and the operation of user interface must be understandable.
- **Robust**: Content must be robust enough that it can be interpreted reliably by a wide variety of user agents, including assistive technologies.

We tried to investigate how to meet accessibility requirements specific to metadata. Concepts like easy to read need special attention, since most metadata are text. Easy to read is an accessibility feature many people may benefit from. For many deaf persons spoken languages like English or German are foreign languages, since they primary communicate in sign language. The best way to fit the needs of this specific group of users would be sign language metadata. But since not so many people know sign language, providing such metadata would be quite difficult. Even if sign language is some person's native language, this does not mean he or she would not understand written language – maybe just not that well. What we may provide is a simpler version of the given written language. Easy language will also serve the needs for people with learning difficulties or even people with other mother tongues who would not understand a foreign language quite well.

We are aware of the fact, that talking about easy language is a challenge within the scientific community. Most scientists depend on their specific terms or subject-specific language, which may be quite hard to understand for some people but are seen as a necessity. A glossary that describes individual terms or similar concepts could help in this matter. In this case we would talk about data describing metadata.

In most cases not the metadata on their own function as accessibility hazards, but the way they are represented within the user interface. Let's take a look on some examples.

4.2 Example 1: Tag Cloud

The concept of a tag cloud may be used to represent how often articles with certain topics or tags are viewed on a website or blog. Individual tags are seen as single clouds or balloons. Each time they are viewed, they ascend within the tag cloud. Imagine you look down on the cloud from above. Objects that are located higher would appear bigger. Since we are dealing with textual objects, "bigger" means "bigger font-size". In the given example "Foto des Monats" ("image ot the month") would be the most viewed topic because it's biggest in the cloud. Another way to show a tag's position in the cloud is by using colour. Objects related more to the ground would look paler due to the longer distance. So the darker an object, the closer it is to the viewer hence we look down from above such objects fly higher.

Tag clouds, as shown in this example, rely on visual characteristics (font size and colour) to represent their structure. Two characteristics a blind person may not use quite well. Neither font-size nor colour plays a big role in screen reader usage. In most cases a screen reader[1] would represent a tag cloud as an unrelated list of tags or topics without any relation.

One way to make tag clouds usable by blind users is to order the cloud by number of views for a tag. A blind user would know, what comes first in the list was viewed most often, and so on. We could even append the number of views like "Foto des Monats: 599 views" to offer more Information.

As we see, the metadata "number of views" by itself is just a number. The number itself is no accessibility problem, but the way it is used in the user interface can be.

4.3 Example 2: DBIS colour flags

For a given library DBIS offers a list of licensed databases that may be accessed by users of the institution [12]. Some databases may offer free access, while others may be used

---

[1] A screen reader is a piece of software that reads text represented by a computer system. In contrast to a computer screen, where structure content and design are represented side by side to be recognised at once, screen readers "read" the content sequentially. A blind user will not know a page's entire content and structure till he or she reaches the end of the page.

only within the IP range of a library or restrict access even more (you have to enter a password). DBIS not only shows a list of the available databases, but also their access model. Access models are shown as colour flags. In case of Klagenfurt university library the colours were defined as follows: *Green* means free access, *yellow* means campus-wide access (IP range) and *blue* means campus-wide access with additional username and password. Would DBIS only use these colours as single way to differentiate between the values this would be a violation of success criteria 1.4.1 of WCAG 2.0. To avoid this behaviour DBIS colour flags additionally show textual information in form of specific characters (in this case F, C and U). Despite the internal implementation of the access level (may be a number or character) the metadata's actual degree of accessibility depends on its representation in the user interface.

Die Datenbank ist ...	
U	für die Universität Klagenfurt lizenziert und campusweit zugänglich; externer Zugang mit AAU-Benutzernamen und Passwort via Proxy möglich.
C	für die Universität Klagenfurt lizenziert und campusweit zugänglich.
F	frei im Web zugänglich.
Bitte beachten Sie die Nutzungsbedingungen und Copyright-Bestimmungen des Verlages/Herausgebers!	

Gesamtangebot (43 Treffer)	Zugang	
ACM Digital Library	Campus+	U
Audio Engineering Society Paper Search	frei	F
Bookboon.com / Studium	frei	F
Cooperative Patent Classification	frei	F
datasheets360	frei	F
DEPATISnet	frei	F

4.4 Example 3: traffic light

Another way to represent – at least three – states with colour coded information could be a traffic light. Within a traffic light you get red, orange and green states, which could lead to misinterpretation, if a user suffers from red-green colour-blindness. But luckily, we got three distinct lamps positioned on top of one another. Therefore, we may interpret a given state by the position of the light glowing at this time. Convention dictates that red is always the upper light, while green is always the lowermost. The traffic light itself is an image, so it may not be read by a blind person a priori. By using alternative text (like red, orange, green) we may serve this user group as well.

As the three examples demonstrate, metadata should always be regarded in context to their representation within a repository's user interface.

Our high aims within the e-infrastructures Austria project are to allow everyone access to both data and metadata.

To meet these expectations, we will have to consider some thoughts:

1. At first we should consider that metadata are also data. Metadata do not only describe data, but in some cases they give important information about the context.
2. We should think about how we may support data-providers to describe their objects intelligible to all, for example in a generally understandable language (easy to read). This also means to offer fields for multilingual explanations – maybe even sign language, which would require other formats than text, like video.

3. We also need to think about the accessibility of help texts.
4. To achieve the goal of universal accessibility we have to learn and understand the different needs of people with specific impairments, as well as people with migration background.
5. An additional benefit of these efforts is that we all get generally understandable metadata and data, we all can use for our work.

4.5 How to achieve this goal?

In order to determine the accessibility of within a given project, we propose the following modus operandi:

- Pick metadata samples to observe. For example, the set of metadata used in a specific repository.
- Apply WCAG 2.0 on user interface
    - Understand metadata representation in the related user interface (perception)
    - Understand metadata input and search forms (operability)
    - How understandable are naming and values? Are input, search and results supported by help systems for users? (understandable)
- Create a catalogue of tips how to handle specific metadata.
    - Or refine an already existing catalogue on each new project.

To test this approach, we plan to apply a case study on *Phaidra*[13], a repository hosted by the university of Vienna. The results will be published separately in time. Phaidra in its current state needs some enhancements towards accessibility, which makes it a perfect test case. In addition, Phaidra is a candidate to act as a repository for literature optimized for the needs of people with certain disabilities in conjunction with accessibility metadata as discussed in the next section of this paper.

5. Accessibility Metadata

Beside the idea of having metadata that are as accessible for as many users as possible metadata themselves may also describe the degree in which a given object can be used by persons with specific disabilities. Let's take a look at a use case to understand the idea.

Since printed media are generally not quite accessible for blind and visually impaired people universities and libraries had to think about ways to offer these persons an alternative way to perceive their literature. Several Austrian universities therefore offer special digitalisation and processing services to transform printed materials into a format, a specific user may need [14][15]. At the end of this process we gain documents with diverse degrees in quality with regard to the recognized content (OCR). Until now people with certain disabilities had no easy way to determine if a given document provided by these services may fit their specific needs on accessibility, since no standards existed so far.

Schema.org [16] recently included several properties, defined by the *accessibility metadata project* [17], within its *CreativeWork* [18] class to identify the accessibility qualities of a publication. Many of these values were derived from the *IMS Global Access for All (AfA) Information Model Data Element Specification* [19]. Since this metadata can be expressed in XHTML content documents using RDFa [20] or microdata attributes [18][21], they can even be included in EPUB3 documents [22].

Let’s take a look on some examples of accessibility metadata.

- *Accessibility features* are content’s particular enhancements to improve the accessibility for specific users. Examples for accessibility features would be alternative text, audio-description, or braille.
- On the other hand, *accessibility hazards* describe characteristics of resources that are physiologically dangerous to some users (like flashing, which may cause epileptic seizure).
- AccessibilityAPIs define Interfaces between an Operation System and assistive technology like screen readers or braille terminals. A given object or source may implement one or more of these APIs allowing assistive technology to interact with it.
- accessibilityControl: Keyboard / mouse / voice etc.

5.1 Example: Accessibility Metadata for video

```
<body itemprop="video" itemscope="" itemtype="http://schema.org/VideoObject">
   <meta itemprop="accessibilityControl" content="fullKeyboardControl">
   <meta itemprop="accessibilityControl" content="fullMouseControl">
   <meta itemprop="accessibilityFeature" content="captions">

   <h1>Martin Luther King's Speech (video): 'I Have a Dream' </h1>
   <video src="http://www.youtube.com/watch?v=smEqnnklfYs" controls="">
      <track src="captions.webvtt" type="captions"/>
   </video>
</body>
```

The given example shows the use of accessibility metadata to describe video content. The example describes a YouTube video which is *fully controllable by keyboard* (which makes it usable for blind users), *controllable by mouse* and *uses captions* for textual transcriptions (what makes it usable for deaf and hard of hearing persons).

6. What are the next steps?

Within the *uniability* [2] workgroup on *digital literature processing for print disabled persons,* where some of the members of Cluster L are involved, we tend to work toward a joint repository to hold literature prepared for specific needs of people with certain disabilities. Within this workgroup we will continue to evaluate, if the proposed

[2] Uniability is the association for the equality of people with disabilities and chronic conditions at Austrian universities and colleges[23].

accessibility metadata provided by projects like the accessibility metadata project may fit the goals, and if necessary how to apply them to our needs.

Within the planed repository we intend to map a user's individual needs to objects serving those needs. Therefore, we will have to provide:

- A *set of metadata* to map user's needs to object properties: Schema.org / Accessibility Metadata Project / access4all defined such metadata.
- The infrastructure to serve different versions of an object, possibly optimized for specific user groups. Search for a suitable repository is still pending. One of the candidates could be an accessibility enhanced future version of Phaidra.
- Tools (possibly as part of the previously mentioned repository) to allow users to specify their needs and settings.

The joined work on Cluster I and Cluster L of project e-Infrastructures Austria built the foundation of our future efforts in the direction and provided us with new ideas and approaches to reach the goal.

7. References

[1] Word Wide Web Consortium (W3C), 'Web Content Accessibility Guidelines (WCAG) 2.0', 27-Jun-2016. [Online]. Available: http://www.w3.org/TR/WCAG20/. [Accessed: 06-Jan-2009].

[2] *Section 508 of the Rehabilitation Act (29 U.S.C. 794d), as amended by the Workforce Investment Act of 1998*. 1998, pp. 105–220.

[3] e-Infrastructures Austria, 'e-Infrastructures Austria project page', 27-Jun-2016. [Online]. Available: http://www.e-infrastructures.at/en/startpage/. [Accessed: 27-Jun-2016].

[4] 'Bundesministerium für Wissenschaft, Forschung und Wirtschaft (bmwfw)'. [Online]. Available: http://www.bmwfw.gv.at/Seiten/default.aspx. [Accessed: 27-Jun-2016].

[5] Project e-Infrastructures Austria, 'e-Infrastructures Austria : Cluster L - Cross Project Issues (technical and non-technical)'. [Online]. Available: http://e-infrastructures.at/en/the-project/work-package-clusters/cluster-l/. [Accessed: 27-Jun-2016].

[6] Project e-Infrastructures Austria, 'e-Infrastructures Austria : Cluster I – Metadata'. [Online]. Available: http://e-infrastructures.at/en/the-project/work-package-clusters/cluster-i/. [Accessed: 27-Jun-2016].

[7] S. Blumesberger, 'Die Welt der Metadaten im Universum von Repositorien.', *Mitteilungen der VÖB. Axams: Steiger, 68 (2015) 3/4*, pp. 515–528, 2015.

[8] S. l. Henriksen, W. Seuskens, and G. Wijers, 'Digitising Contemporary Art: Guidelines for a Long-term Preservation Strategy for Digital Reproductions and Metadata'. 2012.

[9] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana: University of Illinois Press, 1949.

[10] World Wide Web Consortium (W3C), 'Authoring Tool Accessibility Guidelines (ATAG) 2.0', 24-Sep-2015. [Online]. Available: http://www.w3.org/TR/ATAG20/. [Accessed: 27-Jun-2016].

[11] International Organization for Standardization (ISO), 'ISO 14289-1:2012 -

Document management applications -- Electronic document file format enhancement for accessibility -- Part 1: Use of ISO 32000-1 (PDF/UA-1)', *ISO*, 2014. [Online]. Available: http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?ics1=35&ics2=240&ics3=30&csnumber=54564. [Accessed: 27-Jun-2016].

[12] University of Regensburg, 'Datenbank Informationssystem (DBIS)'. [Online]. Available: http://rzblx10.uni-regensburg.de. [Accessed: 15-Sep-2016].

[13] University of Vienna, 'Phaidra - Permanent Hosting, Archiving and Indexing of Digital Resources and Assets'. [Online]. Available: https://phaidra.univie.ac.at/. [Accessed: 27-Jun-2016].

[14] University of Klagenfurt - Library, 'UB Klagenfurt: VIP Services'. [Online]. Available: http://ub.aau.at/cms/en/library-services/vip-services/. [Accessed: 27-Jun-2016].

[15] Uniability - Consortium for equalisation for people with disabilities and chronic deseases at Austrian universities and colleges, 'Uniability - Contact persons in Austria'. [Online]. Available: http://info.tuwien.ac.at/uniability/english/kontakt.htm.

[16] schema.org, 'Home - schema.org'. [Online]. Available: http://schema.org/. [Accessed: 27-Jun-2016].

[17] Accessibility Metadata Project, 'Accessibility Metadata Project | Making Accessible Content Discoverable: A Benetech Led Proposal for Accessibility Metadata in Schema.org'. [Online]. Available: http://www.a11ymetadata.org/. [Accessed: 27-Jun-2016].

[18] schema.org, 'CreativeWork schema'. [Online]. Available: http://schema.org/CreativeWork. [Accessed: 27-Jun-2016].

[19] IMS Global Learning Consortium, 'IMS Global Access for All (AfA) Information Model Data Element Specification', 2012. [Online]. Available: http://www.imsglobal.org/accessibility/afav3p0pd/AfA3p0_DESinfoModel_v1p0pd.html. [Accessed: 27-Jun-2016].

[20] World Wide Web Consortium (W3C), 'RDFa 1.1 Primer - Third Edition', 2015. [Online]. Available: http://www.w3.org/TR/xhtml-rdfa-primer/. [Accessed: 27-Jun-2016].

[21] World Wide Web Consortium (W3C), 'HTML Microdata', 2013. [Online]. Available: http://www.w3.org/TR/microdata/. [Accessed: 27-Jun-2016].

[22] International Digital Publishing Forum (IDPF), 'Schema.org Accessibility Metadata | Metadata | EPUB 3 Accessibility Guidelines'. [Online]. Available: http://www.idpf.org/accessibility/guidelines/content/meta/schema.org.php. [Accessed: 27-Jun-2016].

[23] Uniability - The association for the equality of people with disabilities and chronic conditions at Austrian universities and colleges., 'UNIABILITY Homepage'. [Online]. Available: http://www.uniability.org. [Accessed: 15-Sep-2016].

e-Infrastructures Austria

Sustainable data storage and the provision of data for use by third parties are central roles of science. e-Infrastructures Austria is a federally funded program fort he coordinated expansion and continued development of data repositories across Austria, and is made possible by a grant from the Austrian Ministry of Science, Research and Commerce (BMWFW). This program enables the safe archival and lasting availability of electronic publications, multimedia objects and other digital data from the research and teaching fields. Concurrently, topics relating to research data management and digital archiving workflows will be addressed.

Cluster A	Monitoring of Document Repositories within the Partner Network *Patrick Danowski (IST Austria)*
Cluster B	Planning and Implementation of a „national Survey" for Research Data *Christian Gumpenberger (University of Vienna)*
Cluster C	Designing a Knowledge Network: Development of a reference structure for the construction of Repositories *Paolo Budroni (University of Vienna)*
Cluster D	Infrastructure *Raman Ganguly (Vienna University Computer Center)*
Cluster E	Legal and Ethical Issues *Seyavash Amini (Counsellor-at-law, University of Vienna)*
Cluster F	Open Access *Andreas Ferus (Academy of Fine Arts Vienna)*
Cluster G	Visual Data Modeling *Martin Gasteiner (University of Vienna)*
Cluster H	Life Cycle Management *Andreas Rauber (Technical University Vienna)*
Cluster I	Metadata *Susanne Blumesberger (University of Vienna)*
Cluster J	Permanent backup of the data *Adelheid Mayer (University of Vienna)*
Cluster K	Data from scientific and artistic-scientific research processes *Bernhard Haslhofer (Austrian Institute of Technology)*
Cluster L	Cross-project issues (technical and non-technical) *Andreas Jeitler (Universität Klagenfurt)*

University of Vienna, Library and Archive Services
1010 Vienna, Universitätsring 1
T: +43-1-4277-15172
office@e-infrastructures.at
www.e-infrastructures.at