|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | |  |  |  |  | | |
| |  |  |  | | --- | --- | --- | | |  |  | | --- | --- | | [Picture Placeholder: Pletschacher, Stefan](https://sp.uibk.ac.at/sites/eu-news/Metadata/_layouts/userdisp.aspx?ID=68) | https://sp.uibk.ac.at/_layouts/images/blank.gif | | | [Pletschacher, Stefan](https://sp.uibk.ac.at/sites/eu-news/Metadata/_layouts/userdisp.aspx?ID=68)https://sp.uibk.ac.at/_layouts/images/blank.gifNo presence information | | **ALTO**  ​When implementing ALTO support in our evaluation tools we encountered a number of issues/ambiguities of the Schema and its documentation which we would like to share with you (in addition to what was already reported/proposed to the ALTO board):  ·         Hyphenation at the end of a text line (element HYP)  o   No documentation for attribute ‘CONTENT’ (is it a string, a Unicode entry hex or dec, ASCII code?; type is ‘anySimpleType’)  o   No HEIGHT attribute (there is HPOS, VPOS and WIDTH but no height; the hyphenation is however a physical object on the page)  ·         Usage of coordinates for Circle and Ellipse  o   Circle has attributes HPOS, VPOS and RADIUS. It is not clear if HPOS and VPOS represent the centre of the circle or to top left corner of the bounding box  o   Ellipse has attributes HPOS, VPOS, HLENGTH, VLENGTH. It is not clear if HPOS and VPOS represent the centre or the top left corner of the bounding box. Also it is not clear if HLENGHT and VLENGHT are meant to be width and height or horizontal/vertical radius.  ·         Documentation of block types  o   There are two block elements with a generic TYPE attribute: illustration and composed  o   The documentation for both mentions a few examples (photo, map, drawing, chart and table, advertisement) but there does not seem to be a maintained list.  o   Mapping to other formats (like PAGE) is impossible unless a controlled vocabulary is used.    During testing we also made a few observations with regard to the interpretation of the ALTO "standard" in existing material:  ·         Wrong document dimensions and region coordinates in most ALTO attachments in the IMPACT Repository  o   Reason 1: Deskewing by FineReader Engine (coordinates relative to deskewed image, which is not available)  o   Reason 2: Resized image (coordinates relative to image of different size compared to image in dataset; region coordinates off by a scaling factor) [update] Further investigations showed that this is due to the use of different units, see also discussion below [/update]  o    Examples: 00458540, 00480616  ·         Wrongly used GraphicalElement  o   GraphicalElement is according to the schema documentation intended to “separate blocks. Usually a line or rectangle.”  o   Within the ALTO files in IMPACT in many cases GraphicalElement seems to have been wrongly used for pictures etc. (which should be of type Illustration)  o   Examples: 00458404, 00480616  ·         Missing reading order (IDNEXT) - although optional it would be important to have the reading order explicitly stored with all elements since the (implicit) order within XML files is not guaranteed to be preserved by XML parsers. |
| https://sp.uibk.ac.at/_layouts/images/blank.gif |  |
|  | |
| |  |  |  | | --- | --- | --- | | |  |  | | --- | --- | | [Picture Placeholder: Pletschacher, Stefan](https://sp.uibk.ac.at/sites/eu-news/Metadata/_layouts/userdisp.aspx?ID=68) | https://sp.uibk.ac.at/_layouts/images/blank.gif | | | [Pletschacher, Stefan](https://sp.uibk.ac.at/sites/eu-news/Metadata/_layouts/userdisp.aspx?ID=68)https://sp.uibk.ac.at/_layouts/images/blank.gifNo presence information | | When using mm10 as MeasurementUnit (which is the default) the maximum resolution is limited to 254dpi due to Integer as data type for certain attributes (e.g. HPOS, VPOS).  Besides, mm10 as a unit descriptor is extremely misleading (not to say wrong in the mathematical sense) and should be changed to something like mm10e-1. |
| https://sp.uibk.ac.at/_layouts/images/blank.gif |  |
| |  |  |  |  | | --- | --- | --- | --- | | Posted: 11/29/2012 11:41 PM | [View Properties](https://sp.uibk.ac.at/sites/eu-news/Metadata/_layouts/listform.aspx?PageType=4&ListId=%7BF19479D3-D770-4F43-A916-68416DB51BD0%7D&ID=3) | https://sp.uibk.ac.at/_layouts/images/blank.gif | [[Reply](https://sp.uibk.ac.at/sites/eu-news/Metadata/_layouts/listform.aspx?PageType=8&ListId=%7BF19479D3-D770-4F43-A916-68416DB51BD0%7D&RootFolder=%2Fsites%2Feu%2Dnews%2FMetadata%2FLists%2FTeam%20Discussion%2FALTO&ContentTypeId=0x0107&DiscussionParentID=3) **Reply**](https://sp.uibk.ac.at/sites/eu-news/Metadata/_layouts/listform.aspx?PageType=8&ListId=%7BF19479D3-D770-4F43-A916-68416DB51BD0%7D&RootFolder=%2Fsites%2Feu%2Dnews%2FMetadata%2FLists%2FTeam%20Discussion%2FALTO&ContentTypeId=0x0107&DiscussionParentID=3) | | |
| |  |  |  | | --- | --- | --- | | |  |  | | --- | --- | | [Picture Placeholder: Bauer, Joachim](https://sp.uibk.ac.at/sites/eu-news/Metadata/_layouts/userdisp.aspx?ID=26) | https://sp.uibk.ac.at/_layouts/images/blank.gif | | | [Bauer, Joachim](https://sp.uibk.ac.at/sites/eu-news/Metadata/_layouts/userdisp.aspx?ID=26)https://sp.uibk.ac.at/_layouts/images/blank.gifNo presence information | | Hi Stefan,  here my already outstanding reply on your points outlined.  1.1) I agree that the type should be defined for the HYP as "xsd:string" as well as the others. (added this to proposal schema file)  1.2) The HYP element has no HEIGHT, while similar to the SP (space) it is supposed that only one character is in there.  If for consistency to the String node the HEIGHT should be added here, what's then about "SP" element?  2.1) Regarding the circle: from mathematical point only the center I would think the center makes sence. From design of the ALTO the upper left corner makes it consistent. We have not generated any circles yet, due to we had no need to think about this point and to precise further more since the META/e project. I would propose to define this to center point based on mathematic definition of a circle.  2.2) For the Ellipse I would expect the upper left corner be described by the point HPOS/VPOS and the complete width and height of the rectangle enclosing the ellipse.    3. Regarding the BlockTypes I am not sure if you mixed up "ComposedBlock" with "GraphicalElement", while Illustration and "GraphicalElement" is quite similar. A "ComposedBlock" I understand as a unit, which contains further subunits and can be even described further more. So that is the single one which has a nesting  - even in theory endless possible. So I have a table and in there cells and in there one with a illustration and one with a graphic. So the key point to this is the  <xsd:group ref="BlockGroup"/> inside the "ComposedBlockType".  The differentiation between Illustration and GraficalElement I think is more difficult, but is the same like defining a markup in regards of "What is an article". Depending on the layout of newspapers you cannot define this precised to all cases. Mainly I would extent that the Illustration has any content / meaning (foto, comic, ...), a GraficalElement is just for design / layout purposes.  Perhaps in better wording by native english speakers this can be added. I am limited there with my technical english!  ;-)  I propose to have a call about this next week together to negotiate details and write this directly into the schema proposal. I already started and prepared the file baed on ALTO 2.0 .  regards, jo |
| https://sp.uibk.ac.at/_layouts/images/blank.gif |  |
| |  |  |  |  | | --- | --- | --- | --- | | Posted: 11/29/2012 11:46 PM | [View Properties](https://sp.uibk.ac.at/sites/eu-news/Metadata/_layouts/listform.aspx?PageType=4&ListId=%7BF19479D3-D770-4F43-A916-68416DB51BD0%7D&ID=4) | https://sp.uibk.ac.at/_layouts/images/blank.gif | [[Reply](https://sp.uibk.ac.at/sites/eu-news/Metadata/_layouts/listform.aspx?PageType=8&ListId=%7BF19479D3-D770-4F43-A916-68416DB51BD0%7D&RootFolder=%2Fsites%2Feu%2Dnews%2FMetadata%2FLists%2FTeam%20Discussion%2FALTO&ContentTypeId=0x0107&DiscussionParentID=4) **Reply**](https://sp.uibk.ac.at/sites/eu-news/Metadata/_layouts/listform.aspx?PageType=8&ListId=%7BF19479D3-D770-4F43-A916-68416DB51BD0%7D&RootFolder=%2Fsites%2Feu%2Dnews%2FMetadata%2FLists%2FTeam%20Discussion%2FALTO&ContentTypeId=0x0107&DiscussionParentID=4) | | |
| |  |  |  | | --- | --- | --- | | |  |  | | --- | --- | | [Picture Placeholder: Bauer, Joachim](https://sp.uibk.ac.at/sites/eu-news/Metadata/_layouts/userdisp.aspx?ID=26) | https://sp.uibk.ac.at/_layouts/images/blank.gif | | | [Bauer, Joachim](https://sp.uibk.ac.at/sites/eu-news/Metadata/_layouts/userdisp.aspx?ID=26)https://sp.uibk.ac.at/_layouts/images/blank.gifNo presence information | | Regarding the units onyl short following point: Yes, the meaning of the unit need to be described (like 10th of millimeters).  Regarding the values I do not really think it is well done, but the type definitions varies! So for blocks you onyl can define in integer as you outlined, but for Strings, ... - the smaller units you have float! I do not really like something like this because it makes it more complicated to "get" / understand the format, but at least from purpose and functionality it would work out. A block you cannot messure nor analyze more precised then you already calculated as minimum.  So perhaps all values nevertheless could be defined as float, but we should think once more about negative consequences, e.g. sizing of files, correctness, ...  Regards and good night, jo |