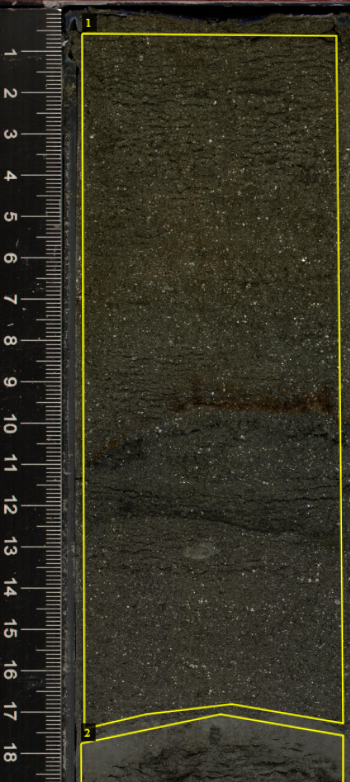
**Description of data annotation**

This description includes marking tags indicating the relevant information for further processing. Semi-structured data type is used as output, which allows changes to the structure and content at run time, forming a dynamic data structure integrated with the analyzed data source. Semi-structured data can be in .json, .xml or .csv file format.



**6,5 cm**

" 362-u1480e-2h-2-a\_shlf7853651\_20160813145856.jpg6748938": {

"filename": " 362-u1480e-2h-2-a\_shlf7853651\_20160813145856.jpg",

"size": 6748938,

"regions": [

{

"shape\_attributes": {

"name": "polygon",

"all\_points\_x": [

400,

1630,

1660,

1120,

710,

410,

400

],

"all\_points\_y": [

1590,

1600,

4930,

4840,

4890,

4960,

1590

]

},

"region\_attributes": {

"Lithology": "Medium-sand\_sandstone",

"Depth": "0.5\_17.0"

}

}

],

"file\_attributes": {}

}

Figure 1 Example of .json file format content of the main tags used in this study. The "name" tag defines the shape of the region of interest. The tags “all\_points\_x” and “all\_points\_y” are X and Y pixel coordinates of the analyzed image forming the polygon with 7 coordinates. The tag "region\_attributes" includes the attributes defined for "Lithology" and "Depth". Attribute "Depth" is measured in cm, in the format: start\_and separated by ‘\_’. The example image belongs to site U1480, Hole E, Core 2h, Section 2. (DUTTA and ZISSERMAN, 2019).

The json output file type is widely used in markup structures that require quick, secure and easy handling (Fig. 1). It is a format widely used on the internet and compatible with unstructured databases. (HAMOUDA and ZAINOL, 2019; STREKALOVA and BOUAKKAZ, 2017; MARTINEZ-MOSQUERA, NAVARRETE and LUJAN-MORA, 2020). The file name is defined as a primary initial tag designated as an access key. The tags "filename", "size", "regions" and "file\_attributes" define the primary structure with information from the analyzed file (image, video or audio). The subtag “shape\_attributes” stores information on areas of interest with pixel X and Y coordinates, in addition to specific user-defined attributes, which in this case are “lithology” and “depth” attributes.

References

HAMOUDA, S., ZAINOL, Z., 2019. Semi-Structured Data Model for Big Data (SS-DMBD). 8th International Conference on Data Science, Technology and Applications (DATA 2019). DOI: 10.5220/0007957603480356.

STREKALOVA, Y.A., BOUAKKAZ, M., 2017. Semi-structured Data. In: Schintler L., McNeely C. (eds) Encyclopedia of Big Data. Springer, Cham. https://doi.org/10.1007/978-3-319-32001-4\_183-1.

MARTINEZ-MOSQUERA, D., NAVARRETE, R., LUJAN-MORA, S., 2020. Modeling and Management Big Data in Databases—A Systematic Literature Review. Sustainability, 12, 634. DOI:10.3390/su12020634.