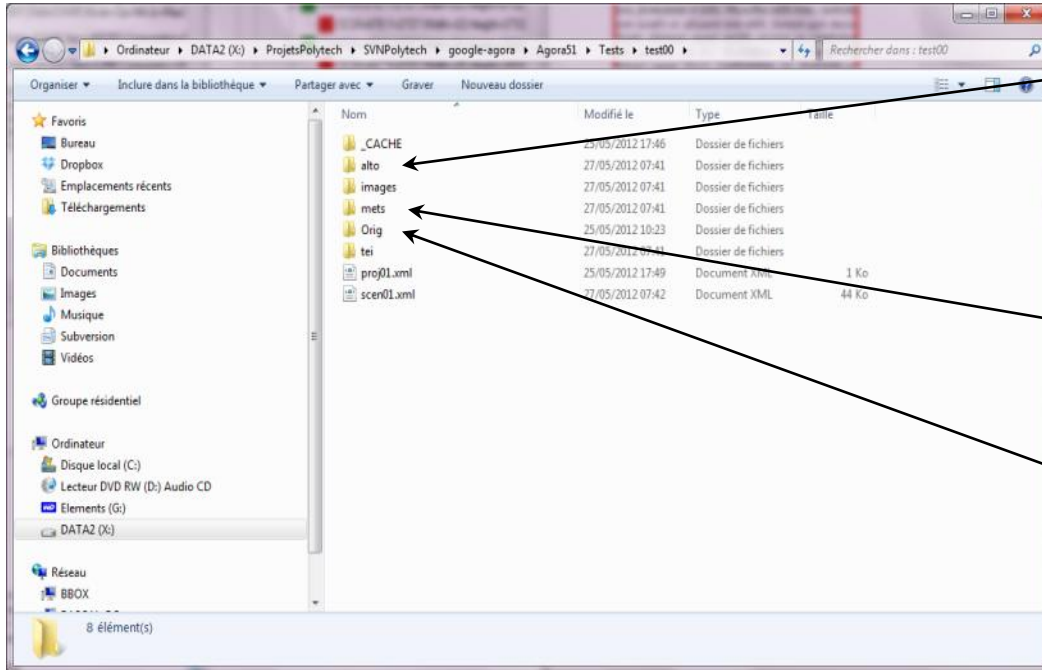


Document

This scenario exports its results in ALTO format. In the project directory, directories were as follow:



Alto directory contains:

- ALTO generated files
- thumbnails of each EOC (png format) present in each ALTO file

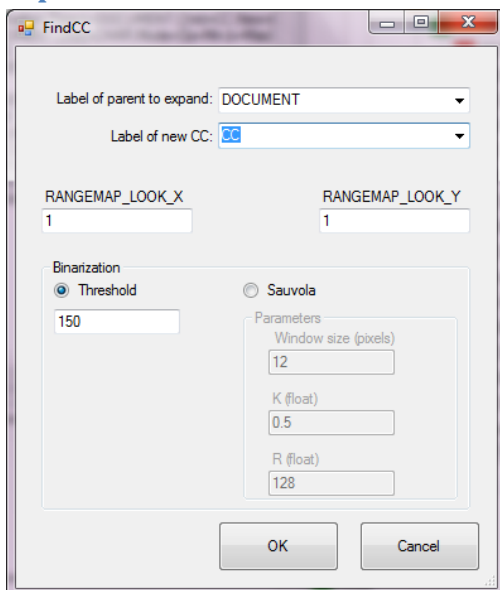
METS files. Their structure is the same as the ALTO files.

Original images (directory's name is stored in project file)

Details of the scenario

By double-clicking on a scenario step, you can visualize its parameters. This double-click is only active for “complex” steps !

Expand an EOC : FindCC



This operator will expand any parent EOC by adding childs EOC to it. These childs EOC will be connected component identified in the parent's XY bounding box. 2 binarization pre-treatment algorithms are possible. RANGEMAP_XXX parameters will be useful for complex layouts with many blocks in the document; in this tutorial, default values are used.

Insert new EOCs: classify by features

Classify By...

Label of parent: DOCUMENT

Label of new EOC: NOISE

Label of child: CC

Replace child

Features of child:

	a0	b0
EUCLIDIAN_X	0	15
EUCLIDIAN_Y	0	15
RANGEMAP_X	0	0
RANGEMAP_Y	0	0
CHILDS_NUMBER	0	0

Coefficients:

0	Ch0
1	...
1	...
0	+Infini
0	+Infini
0	+Infini

Comparators:

0	Ch0
...	...
...	...
...	...
...	...
...	...

Reset

OK

Cancel

Add

Del

This means: a new EOC “NOISE” will be inserted between EOC “DOCUMENT” and one of its child EOC “CC” if rules are satisfied.

For this operator, rules are:

- Size of the child EOC is considered (i.e. EOC can be everywhere)
- Coefficients will be applied to each feature before comparison with comparators. Here, the “0” means simply that this corresponding feature doesn’t care.
- between 0 and 15 pixels in X, and between 0 and 15 pixels in Y

Specifications of “between” are done by the following comparators:

FeatureMatrixComparator

EUCLIDIAN_X

Graph's EOC Feature

Pattern's or Graph's EOC Feature

Reset

OK

Cancel

FeatureMatrixComparator

EUCLIDIAN_Y

Graph's EOC Feature

Pattern's or Graph's EOC Feature

Reset

OK

Cancel

Here, “a0” and “b0” are the features of the pattern child “CC”, “a1” and “b1” are the features of the actual child “CC”. For specifying “between”, we specify “a1 > a0” and “b1 < b0” thanks to the showed settings.