

## D4.6 Validation methodology CANTATA



Project number:	ITEA05010
Document version no.:	v2.0
Status:	Final
Edited by:	Roel Truyen, Philips Healthcare

**ITEA ROADMAP DOMAINS:**

Major: Services & Software creation  
Minor: Cyber Enterprise

**ITEA ROADMAP TECHNOLOGY CATEGORIES:**

Major: Content  
Minor: Data and content management

**TABLE OF CONTENTS**

**1 Introduction .....2**

1.1 General scheme .....2

1.2 Description of components .....2

1.2.1 Test Dataset .....2

1.2.2 Ground Truth Annotation and Description .....3

1.2.3 Content Analysis and Results Description .....3

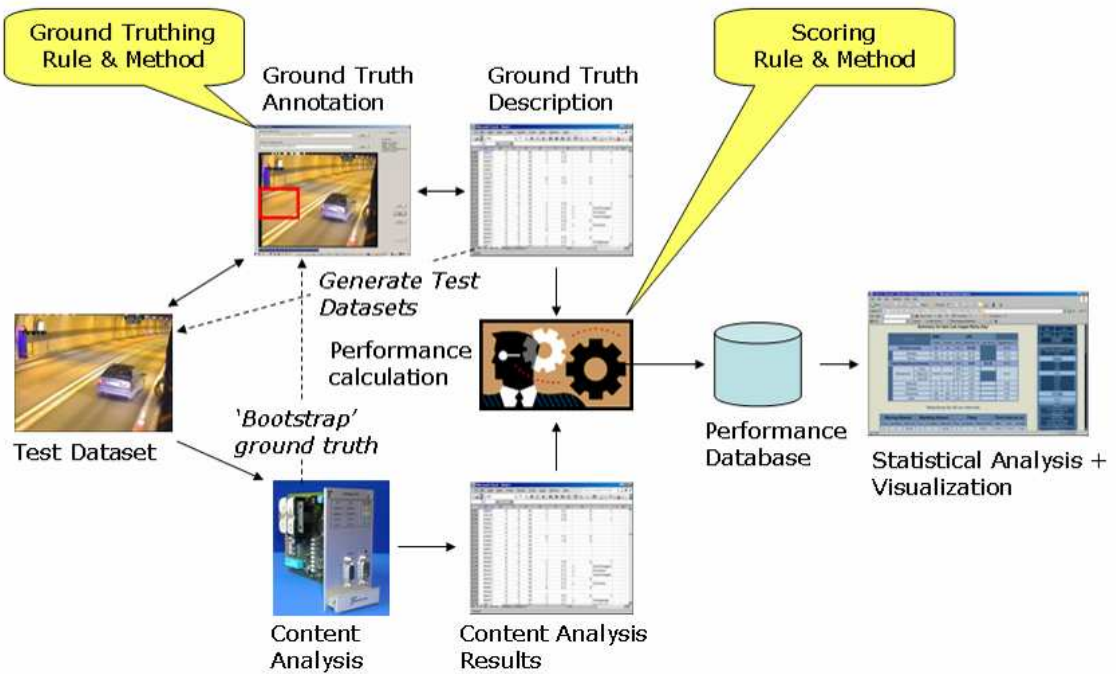
1.2.4 Performance calculation - metric .....3

1.2.5 Statistical Analysis & Visualization .....3

1.3 Description of validation methodology .....4

**1 Introduction**

**1.1 General scheme**



**Figure 1 - General scheme for MCA validation**

**1.2 Description of components**

In Figure 1 the main components of a validation methodology are depicted for offline validation (not involving a user). Following components should be clearly defined in the validation methodology in order to obtain meaningful validation results.

**1.2.1 Test Dataset**

A test dataset is a set of digital content (images, movies, sounds) with which testing is conducted to determine the performance of a MCA device. It must be different from the set of images used to

train/develop or validate the algorithm. The test database represents the target events and the target conditions for which the device is intended. The test database usually includes a relevant spectrum of content characteristics and acquisition hardware, but may include difficult cases as well.

There are various methodologies to obtain an appropriate database for device testing:

- Collection during field testing, where data is collected and interpreted in real use
- Enrichment for rare events (“needle in a haystack” problems)
- Enrichment for a stress test
- Generate test datasets automatically based on a ground truth description  
e.g. by using the Object Video Virtual Video Tool

### **1.2.2 Ground Truth Annotation and Description**

Depending on what should be evaluated, the ground truth should be defined and annotated. There is a large choice in ground truth rules and methods, including differences in:

- source of ground truth, e.g. using a different source of information or an expert (panel)
- unit of analysis and ground truth, which can be an entire image sequence or a small subset of a frame
- format of ground truth, ranging from a simple yes/no answer (per unit of analysis) to full spatiotemporal coordinates

Ground truth is annotated using an annotation tool that generates output in an agreed format. In cases where it is too cumbersome to manually annotate the ground truth this can be ‘bootstrapped’. In this annotation paradigm the VCA algorithm is used to generate a first draft of the ground truth which can then in a second phase be manually corrected with the annotation tool. Therefore, three different methods for the annotation are proposed (manual, semi-automatic or completely automatic annotation).

### **1.2.3 Content Analysis and Results Description**

The device under test is the content analysis algorithm that is applied to the test dataset and writes its results in an agreed format suitable for comparison with the ground truth.

### **1.2.4 Performance calculation - metric**

Scoring rules and methods should be defined how to compare the annotated ground truth with the results of the content analysis algorithm. This performance rule is often called the metric, and should be well described in the validation methodology.

### **1.2.5 Statistical Analysis & Visualization**

The content analysis algorithm will be applied to a set of content that is representative of the typical intended use of the algorithm – possibly involving a representative set of users/readers as well. For every content item (image, sound, movie) and for every user a (set of) performance measure(s) will be available. The validation methodology should describe what statistical analysis methods will be used to analyze this data and what hypothesis rules apply if these are tested. Also the visualization of the performance results is an important aspect that should be specified in the methodology.

### **1.3 Description of validation methodology**

We should first determine what a content analysis algorithm should do and what role it plays in a real application. This was described in D4.2 under Scenario's and Use Cases. From this we can write down the requirements for validation of the different scenario's with specification of the datasets needed to perform such a validation, the definition of the ground truth and the annotations that need to be done. Once we know these we can describe the different validation components as mentioned in 1.2 for each of the scenario's.