C6 - Service Validation Report

Draft I1.00

GSE Land Information Services

ESRIN/Contract No. 19407/05/I-LG

Stage 2 of the Earthwatch GMES Service Element Scaling Up Consolidated GMES Services

Book Captain:	Stefan Kleeschulte (UAB)
Contributing authors	George Büttner (FÖMI)
	Oscar Gomez (UAB)
	Michael Rumberg (TÜV)

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY 3

2. SCOPE 4

3. RELATED DOCUMENTS ..... ERROR! BOOKMARK NOT DEFINED.

3.1 INPUT **ERROR! BOOKMARK NOT DEFINED.**

3.2 OUTPUT ..... **ERROR! BOOKMARK NOT DEFINED.**

4. QA/QC PROCESS 5

5. VALIDATION RESULTS 7

5.1 M2.1 SAAR-MOSEL (INFOTERRA) 7

5.2 M2.1 WESER (INFOTERRA) 8

5.3 M2.6 WESER (GEOVILLE) 10

5.4 M2.6 SAIL (GIM) 12

5.5 M2.4 SAIL (GIM) 15

5.6 M2.4 WESER (GIM) 17

5.7 M1.1 MADRID (INDRA) 18

5.8 M1.1 LILLE (EUROSENSE) 20

5.9 M1.1 BARI (PLANETEK) 22

5.10 M1.1 MUNICH (HG-GDS) 24

5.11 M1.1 WEIMAR / JENA (HG-GDS) 26

6. VALIDATION ISSUES AND CONCLUSIONS 28

6.1 PROCESS RELATED COMMENTS 28

6.2 PRODUCT RELATED COMMENTS 28

1. EXECUTIVE SUMMARY

This document describes the results of the quality control process during phase 1 of the GSE Land project.

The QA/QC process of by the QA team has revealed some practical problems in the pre-operational execution of the production and validation process. The QA-team calls the process “pre-operational” as it believes that many of the issues addressed below can be solved in a later stage of the project, when the production and validation requirements are better defined (finalised) and more clear to all involved parties. Main findings are:

- Difficulties in availability and access to high-resolution reference data (e.g. orthophotos) for the product validation.
- Product accuracies are dependent on the validation method.
- Product mapping guidelines need to be improved for some products to overcome ambiguities in object definition. This ambiguity can result in a different interpretation of the object during production and validation.
- The QA methodology needs to be better documented.

Some of the issues originate from the fact that the QA method was developed in parallel to the production of the first data sets.

The status of the validation and the findings for each the individual products and production sites is summarised in chapter 5.

2. SCOPE

The objective of the validation report is to summarise the findings and the results of the independent quality assurance and validation carried out by the QA technical team. The QA technical team is part of the European Topic Centre on Terrestrial Environment (ETC/TE).

The QA team has performed an independent assessment of the product quality according to the QA guidelines agreed in the GSE Land project. This QA approach is ideally based on the following components:

- A qualitative verification at this beginning of the project (after finalisation of 30% of the mapping area) to detect main issues that might influence the quality of the final product.
- A validation with quantitative feedback on product quality. The validation has been carried out with 2 different methods. An new one developed by GSE Land (cluster approach) and a “old” one (point sampling) to guarantee comparability of the accuracy values with original specifications, based on a point validation approach.
- Re-interpretation of land cover objects, respectively sampling points without knowledge of the original interpretation results from the service provider.
- Re-interpretation based on high-resolution reference data (e.g. aerial photographs, Very High Resolution Satellite Imagery (1m)), preferably from the same year as the satellite image used by the service provider.
- Availability of scanned topographic maps (geocoded), preferably recently updated.
- Reference data available in useable GIS format, ready for integration into validation software (ArcView).

The result of the validation provides information on the ***overall accuracy*** of the product. Additional accuracy parameters are calculated for service provider information only. These should serve to improve the production process. Their communication to the end user is to the discretion of each service provider.

3. QA/QC PROCESS

The QA/QC process had been split into two different steps. A verification with qualitative feedback to the service providers and a validation with a quantitative evaluation of the production results.

The verification was done for the following sites. Main comments are included.

- M2.6 Weser – 0.25 ha MMU for an exhaustive mapping of all artificial features is challenging (CLC class 1)
- M2.6 SAIL – Rejection of product by verification due to insufficient quality.
- M2.1 Saar-Mosel – good verification results, no major comments
- M2.4 Weser - Rejection of product by verification
- M1.1 Madrid, Weimar – difficulties for the verification of land use / functional objects.

The validation approach is based on a selection of 2.5x2.5 km validation plots (“clusters”) for regional products (urban products 1.25x1.25 km). Within these clusters the QA technical team re-delineates and re-interprets the objects. This re-interpretation should be done on high-resolution image data (aerial photographs, VHRSI) in combination with other support material (e.g. topo maps). The re-interpretation is performed by the QA team without knowledge of the results created by the service provider, i.e. in a fully blind and independent way. Certainly the blind approach has some subjectivity, but this should be reduced to a minimum by detailed mapping and interpretation guidelines. The major advantage of this approach is the validation of the geometry and the semantic content of an object, which cannot be achieved by a point sampling approach.

The development of the validation approach had to consider some important boundary conditions that have influenced its final design:

- Development of the validation approach in parallel with the production of first products.
- Design of an approach that minimises the need for high-resolution reference data (e.g. orthophotos, IKONOS).
- Selection of sample plots (clusters) before the results of the classification are known (i.e. the actual distribution of classes).
- Cost-efficient, independent approach.
- European-wide applicability.

The validation starts with 1% of the total production area. If the product specifications are met within this area, the validation is finished. If the product fails the validation by a small difference, the sample size is increased by 1% (up to a maximum of 5%). The idea of this approach is that only a small number of samples is needed to verify if a product is really good (or bad). Only in case of being on the edge, the sample size needs to be enlarged to confirm the tendency.

Upon request by the service providers and supported by user community the validation of the mapping products was done in two different ways. In addition to the method described above – which represents the new approach for product validation proposed by the QA team of the GSE Land project – a more traditional point sampling approach is implemented.

This decision is based on the fact that the original accuracy target values were based on such a point sampling, not on the stricter cluster approach. Both values will be communicated in phase 1 to familiarise service providers as well as clients to the more strict values.

4. VALIDATION RESULTS

4.1 M2.1 SAAR-MOSEL (INFOTERRA)

Location / phase	Saar-Mosel 1st delivery
Product/ nomenclature	M2.1
Service Provider	ITD-Delphi
Area (km^2)	28.288
Processing status (percent of total)	100
Start date of validation	13.06.2006.
End date of validation (1st iteration)	16.06.2006.
Photointerpreters	B Kosztra, M Bíró
Technical preparations	G Maucha, R Pataki
Validation software	InterView GSE 1.0

4.1.1 Validation

Input data

Orthophotos / VHRSI	Limited availability of orthophotos.
Original satellite images	Available.
Topographic maps	The topographic maps sent by ITD could only be opened in B&W.
Other reference material	None.

4.1.2 Validation process

Number of validation clusters	45 (2.5 x 2.5 km)
Number of sampling points	
Additional iterations	None.
Number of additional samples	None.

4.1.3 Comments / remarks

The cluster validation is based on a blind re-interpretation (without knowledge of the service provider interpretation) of the land cover objects within the cluster area (2.5x2.5 km). The approach is sensitive to thematic and geometric differences between both interpretations.

For the overall accuracy, the service provider and the reference interpretations are overlaid and the differences (in area) are quantified.

The topographic maps could only be opened in black & white format, even though the data sent by ITD was in colour. The problem is being investigated.

Almost no aerial photographs (i.e. independent data) were used for the validation. No orthophotos were available for the French part of the implementation site. For Rheinland-Pfalz images were not available in time. Only for a small part of the Saarland orthophotos were provided. ITD is still trying to overcome this bottleneck through continuous discussions with the user organisations.

The satellite images were provided in a mosaic. The validation team would prefer original scenes as the overlap between scenes (2 dates) can be used to better discriminate certain land cover classes.

4.1.4 Results / findings

Comments:

- The product was validated on the original satellite data. A re-interpretation on the original data cannot compare the mapping results to a high resolution data source, thus it mainly gives an explanation how much information was extracted from the images, but it does not say how well it compares with “reality” (the reality in the aerial photograph).
- Lack of independent high resolution images (orthophotos) for most of the production area.
- The validation results were reviewed by S. Meirich and by G. Büttner, the team leader of the technical team. This review showed that in a first validation round there had been a systematic confusion of a number of classes. This resulted in a first, very low accuracy value. The second interpretation of the reference data by the team leader identified those confusions and corrected them. The technical team members were briefed about the results of this review and the confusion should be avoided in future.
- About 10% of the disagreement (error) between validation and production is attributed to a confusion of arable land and pasture.

4.1.5 Overall accuracy

- Target accuracy: 80% +/- 3%
- Product accuracy: 79.6% - target achieved.

4.2 M2.1 WESER (INFOTERRA)

Location / phase	Weser, 2nd delivery
Product/ nomenclature	M2.1
Service Provider	ITD-Delphi
Area (km^2)	45.979

Processing status (percent of total)	100
Start date of validation	11 April 2006
End date of validation (1st iteration)	2 June 2006
Photointerpreters	B Kosztra, M Bíró
Technical preparations	G Maucha, R Pataki
Validation software	InterView GSE 1.0

4.2.1 Validation

Input data

Orthophotos / VHRSI	Available. Acquisition 2001 and 2003.
Original satellite images	Available. Year: 2005
Topographic maps	Available.
Other reference material	None.

4.2.2 Validation process

Number of validation clusters	29 + 56
Number of sampling points	
Additional iterations	None.
Number of additional samples	None.

4.2.3 Comments / remarks

The product was validated two times. First after 30% of the production area was finished (29 clusters) and after the completion of 100% (additional 56 clusters). The results of both validations were combined to one overall accuracy value.

The cluster validation is based on a blind re-interpretation (without knowledge of the service provider interpretation) of the land cover objects within the cluster area (2.5x2.5 km). The approach is sensitive to thematic and geometric differences between both interpretations.

For the overall accuracy, the service provider and the reference interpretations are overlaid and the differences (in area) are quantified.

The aerial images were taken 2 to 4 year before the satellite image on which the land cover map was produced. The interpretation by the technical team was cross-checked with the satellite data to avoid disagreements because of this time difference.

The spatial distribution of the validation plots (clusters) was not optimal, as it was adapted to those sites where aerial photographs existed - mainly urban areas which are in general a bit more complex in their structural composition.

4.2.4 Results / findings

Comments:

- Orthophotos were available for 28 of the 56 clusters of the second validation round. For the clusters without orthoimages, the original satellite data was re-interpreted.
- For two clusters no satellite data were available.
- The confusion of arable land and pasture contributes to about 5% of the error matrix.

4.2.5 Overall accuracy

- Target accuracy: 80%
- Product accuracy
  - Cluster approach: 81.1% - target achieved.
  - Buffered cluster approach 10m: 87%

  The buffer is used to eliminate small differences in object delineations within the given geometric precision of the product. But it also washes out a number of small polygons within the buffered area.

The results of the validation illustrate very well the influence of the validation method on the final accuracy value.

4.3 M2.6 WESER (GEOVILLE)

Location / phase	Weser, first delivery
Product/ nomenclature	M2.6
Service Provider	GeoVille
Area (km^2)	10.032
Processing status (percent of total)	30%
Start date of validation	28 April 2006
End date of validation (1^{st} iteration)	5 May 2006
Photointerpreters	B Kosztra, M Bíró
Technical preparations	G Maucha, R Pataki
Validation software	InterView GSE 1.0

4.3.1 Validation

Input data

Orthophotos / VHRSI	Orthophotos available for first iteration

	only. Date unknown.
Original satellite images	Some geometric distortion (25-30m) between satellite images 53_242 and 54_242 (cluster 44, 46)
Topographic maps	Available, except for cluster 44 (respectively 11 sampling points)
Other reference material	None.

4.3.2 Validation process

Number of validation clusters	2 x 24 (2 iterations)
Number of sampling points	3530
Additional iterations	
Number of additional samples	

4.3.3 Comments / remarks

The product was evaluated with different methods:

- Cluster based area sampling

  For the cluster based area sampling 2% of the total area were validated. For 16 out of 48 clusters aerial photographs were available.

  The cluster approach is based on a blind re-interpretation (without knowledge of the service provider interpretation) of the land cover objects within the cluster area (2.5x2.5 km). The cluster approach is sensitive to thematic and geometric differences between both interpretations.

  For the overall accuracy, both interpretations are overlaid and the differences (in area) are quantified.

- Random point sampling

  Orthophotos were only available for 115 sample points. 11 sample points were not covered by topographic maps.

  The point sampling is based on a distribution of validation points over the complete production area with a minimum number of points per class. The sampling approach is designed to assess the thematic accuracy of the product. The validation points are placed towards the centre of the objects (polygons) to avoid border problems.

In case of non-availability of the aerial photographs (and topo maps) the validation was performed on the original satellite data.

4.3.4 Results / findings

Comments:

- The exhaustive mapping of all artificial area with 0.25ha is a challenge (going to the limit of what is possible) with the satellite data used.
- Some confusion exists if an object can be classified by its land use respectively its land cover. This is a nomenclature issue, not necessarily a weakness of the production.

4.3.5 Overall accuracy

- Target accuracy: 95% +/- 3%
- Product accuracy:
  - Cluster approach: 88.8%
  - Buffered[1] cluster approach 10m: 91.8%
  - Stratified random point sampling: 95.7% - target achieved.

The results of the validation illustrate very well the influence of the validation method on the final accuracy value. An accuracy of close to 89% with the cluster approach was considered by the QA team as very good and should be considered equal to the 95% from the point sampling.

4.4 M2.6 SAIL (GIM)

4.4.1 Product submitted for validation

Location / phase	Sail
Product/ nomenclature	M2.6
Service Provider	GIM
Area (km^2)	19120
Processing status (percent of total)	100% of the area has been produced. Due to difficulties in the availability of reference data in the NL, the validation area only covers about 85% of the work area.

[1] The buffer was used to eliminate small differences of object delineations within the given geometric precision of the product. But it also washes out a number of small polygons within the buffered area.

Start date of validation	10th Aug 2006
End date of validation (1st iteration)	29th Aug 2006
Photointerpreters	B Kosztra, G Büttner
Technical preparations	G Maucha, R Pataki
Controlled by	G Büttner, B Kosztra
Validation software	InterView GSE 1.0

4.4.2 Validation

Input data

Orthophotos / VHRSI	Available. Very good quality. NL data have projection problem. Acquisition date not provided.
Original satellite images	Available.
Topographic maps	Available. NL data have projection problem Projection of BE data has been corrected by FÖMI
Other reference material	None

4.4.3 Validation process

Number of validation clusters	25 clusters validated (UK 12, BE 5, FR 8). 1 FR cluster could not be validated, because of large geometrical shift All 5 NL clusters could not be validated because of projection problem with reference data
Number of sampling points	None
Additional iterations	None.
Number of additional samples	None.

4.4.4 Comments / remarks

- The cluster validation is based on a blind re-interpretation (without knowledge of the service provider interpretation) of the land cover objects within the cluster area (2.5x2.5 km). The approach is sensitive to thematic and geometric differences between both interpretations.

- For the overall accuracy, the service provider and the reference interpretations are overlaid and the differences (in area) are quantified.
- Because of problems with projection (topomaps and orthophotos) in NL data, the Netherland part of the SAIL region could not be validated.

  The origin of the projection problem (shift) is unclear, as the reference data appear correct when the service provider checked them. It seems to be a software problem as GIM used ArcGIS to prepare the data, while FÖMI uses ArcView 3.2 to display the data.

  Due to the problem in NL, the planned 1% minimum area for validation could not be achieved. For alternative validation sites reference data are not available.
- In the UK part there is a significant geometric shift (appr. 50 m) between some of the satellite images and the orthophotos while there is a good fit between the orthophoto and the topographic map. The origin of this shift should be checked by the service provider to avoid impacts on the final mapping product.
- The Belgian topographic maps were reprojected at FÖMI to eliminate a 1300m shift.
- The French clusters are usually covered by more than one orthophotos. These orthos are not in good registration with each other, and sometimes several 10 m deviations can be seen, which makes validation interpretation difficult. For an overlapping area, the orthophoto showing better registration with the topographic map has been considered. Cluster 23 has not been interpreted, because the corresponding orthos showed too high misregistration (20 – 50 m).

4.4.5 Results / findings

Comments:

- Due to insufficient high-resolution reference material, only 85% of the validation clusters could be checked. Since no additional reference material is available the cluster size cannot be extended.
- Satellite images in the UK part have 26 m pixel size and no SWIR band. This is practically unsuitable to deal with the 0.25 ha MMU.
- The excellent quality hi-res data have shown that lots of objects (e.g. farms, forests) exist in the area having size close to the MMU. Looking at the coarser resolution satellite images many of these objects are impossible to delineate. It is suggested to compare the object size, delineated by the SP and that of the validator.
- Analysing the contingency matrix computed from the SP's classification and the validation, the largest disagreements (errors) are as follows:
  - 200-110: urban fabric classified as agriculture – 2.4% (of total area).
  - 110-120: industry / commercial classified urban fabric – 1.8 %.
  - 110-200: agriculture classified as urban fabric – 1.4%
- The resolution of the satellite data is on the edge for the detection of small objects (close to the MMU) – especially in small farmsteads rural areas.
- Probably due to the production without topographic maps, class 120 (industrial and commercial areas) has a rather low producer accuracy (30%) – confusion with urban fabric (110) and agriculture (200).

In view of the importance of this class for the further downstream production (sealing levels), it is recommended to review this class before starting the downstream service.

4.4.6 Overall accuracy

- Target accuracy: 95% +/- 3%
- Product accuracy: 85% (this is in the same range as the M2.6 product for the Weser region – 89%). – Target achieved, nonetheless it is recommended to review in particular class 120 which had a low producer accuracy (30%).

4.5 M2.4 SAIL (GIM)

Location / phase	Sail
Product/ nomenclature	M2.4
Service Provider	GIM
Area (km^2)	19120
Processing status (percent of total)	100% of the area has been produced. Due to difficulties in the availability of reference data, the validation area only covers about 50% of the planned area.
Start date of validation	16th Aug 2006
End date of validation (1st iteration)	21st Aug 2006
Photointerpreters	G Büttner
Technical preparations	G Maucha, R Pataki
Controlled by	B Kosztra
Validation software	InterChange GSE 1.0

4.5.1 Validation

Input data

Orthophotos / VHRSI	Available (except FR). Very good quality
Original satellite images	Available.
Topographic maps	Available for UK and BE (BE was with wrong projection) Not available for NL and FR
Other reference material	None

4.5.2 Validation process

Number of validation clusters	17 clusters validated (UK 12, BE 5) 14 clusters not validated (FR 9, NL 5), because of missing reference data.
Number of sampling points	
Additional iterations	None.
Number of additional samples	None.

4.5.3 Comments / remarks

The validation of changes is a complex task. The validation does not only need to validate changes mapped by the service provider, but should also find out if all existing changes are mapped. It was decided to validate the changes for the same clusters that were used to assess the quality of the base product (M2.6). This was done to avoid the need to request additional high-resolution reference data.

The validation of changes (increments or decrements) of polygon areas and / or codes in the M2.6 base map for 2005 was based on comparison of the image data taken in 2000 and 2005.

Because of missing topomaps (NL and FR), and missing orthophotos (FR), only 55 % of clusters could be validated.

Due to the large pixel size (15-25 m) of 2000 and 2005 images as well, it is hardly possible to map small changes (a few ha and below). Very small number of changes has been identified

For UK: 2000: 14 m (SPOT), 2005: 26 m (SPOT)

For BE: 2000: 25 m (SPOT), 2005: 10 m (Ikonos)

The quality of the change mapping is impaired by the rather coarse resolution of the targeted minimum mapping unit (0.25 ha or 2x2 pixels at 25m). The different pixel sizes for the two dates add an additional difficulty to the mapping exercise.

4.5.4 Results / findings

Comments:

- Very difficult access to reference data.
- Pixel size is not optimal for mapping changes with very high details.
- Very small amount of changes have been found during validation.
- The frequent lack of SWIR band made validation (but also photointerpretation) more difficult.

Overall accuracy

- Target accuracy: 95% +/- 3%
- Product accuracy: not yet specified.

4.6 M2.4 WESER (GIM)

Location / phase	Weser
Product/ nomenclature	M2.4
Service Provider	GIM
Area (km^2)	33439
Processing status (percent of total)	50% (although 100% have been processed by GIM, only the area corresponding to the previously validated M2.6 area was checked)
Start date of validation	7th Aug 2006
End date of validation (1st iteration)	14th Aug 2006
Photointerpreters	G Büttner
Technical preparations	G Maucha, R Pataki
Controlled by	B Kosztra
Validation software	InterChange GSE 1.0

4.6.1 Validation

Input data

Orthophotos / VHRSI	For half of the clusters good quality colour orthophotos are available
Original satellite images	2005: 10 satellite images 2000: 11 of the available 19 satellite images used – the images do not cover the entire area (see map).
Topographic maps	Good quality colour topomap mosaics
Other reference material	None

4.6.2 Validation process

Number of validation clusters	25 clusters. Only 12 clusters have orthophotos.
Number of sampling points	None

Additional iterations	None.
Number of additional samples	None.

4.6.3 Comments / remarks

The validation of changes is a complex task. The validation does not only need to validate changes mapped by the service provider, but should also find out if all existing changes are mapped. It was decided to validate the changes for the same clusters that were used to assess the quality of the base product (M2.6). This was done to avoid the need to request additional high-resolution reference data.

The validation of changes (increments or decrements) of polygon areas and / or codes in the M2.6 base map for 2005 was based on comparison of the image data taken in 2000 and 2005.

The distribution of clusters is uneven, because the existing orthophotos have been used as much as possible.

4.6.4 Results / findings

Comments:

- Satellite images for 2000 have lower resolution (20 m) and no SWIR band, therefore consistent delineation of very small changes (less than 1 ha) was hardly possible.
- A very small number of changes was identified during the validation.

4.6.5 Overall accuracy

- Target accuracy: 95% +/- 3%
- Product accuracy: not yet specified.

4.7 M1.1 MADRID (INDRA)

Location / phase	Madrid 2nd delivery (100%)
Product/ nomenclature	M1.1
Service Provider	Indra
Area (km^2)	1059
Processing status (percent of total)	100
Start date of validation	17 May 2006
End date of validation (3rd iteration)	28 Aug 2006
Photointerpreters	B Kosztra
Technical preparations	G Maucha, R Pataki

Validation software	InterView GSE 1.0

4.7.1 Validation

Input data

Orthophotos / VHRSI	Colour orthophotos available, reference year 2004
Original satellite images	Reference year August 2005
Topographic maps	Available.
Other reference material	None.

4.7.2 Validation process

Number of validation clusters	19 (1.25x1.25 km)
Number of sampling points	874
Additional iterations	None.
Number of additional samples	None.

4.7.3 Comments / remarks

The validation was based on a blind re-interpretation (including delineation) of the land cover objects within the cluster sites (cluster sampling).

For the cluster sampling a second accuracy value was derived by buffering the object delineations. This buffer eliminates small differences in object delineation in the SP and reference interpretation. For the urban atlas product a buffer of 7.5 m (to both sides of the line) was agreed.

A maximum of 900 sampling points was checked in the point validation. The number of points mentioned above lists those points for which a clear interpretation was possible.

Points falling on or close to land cover boundary have been coded by 999 and were excluded from the above mentioned number of “valid codes”.

4.7.4 Results / findings

Comments:

- The product was evaluated three times (without topographic maps, with B&W orthophotos, with colour orthophotos).
- The seven clusters validated earlier by using b&w orthophotos have been checked again and corrected if needed.
- The colour orthophotos proved to be much more useful than the b&w ones used in the previous version of the validation, not because of the colour information, but also the better resolution and timeliness.

- Some of artificial/urban areas is still interpreted as 999 (unidentifiable area), but this number is much less than in the previous versions. 999 code is used frequently on suburban (agricultural) areas when the use of the area was not clear (210, 230 or 320).
- Newly constructed areas are difficult to classify (e.g. residential or public).
- Significant part of agricultural land is hard to identify because of the lack of suitable topographic maps and the missing (not available) SWIR band of satellite image.

Analysing the contingency matrix computed from the SP's classification and the validation, the largest disagreements (errors) are as follows:

- 131-210: arable land classified as mine: 4.6%
- 240-320: natural grassland / shrubs classified as mixed agriculture: 4.3 %
- 310-320: natural grassland / shrubs classified as forests: 3.2 %
- 230-320: natural grassland / shrub classified as pastures: 2.3%
- The producer accuracy of the urban fabric (111 & 112) is below 65%.

This means that most of the disagreements stem from the classification of agricultural & seminatural lands and not from built-up.

4.7.5 Overall accuracy

- Target accuracy: 90% +/- 3%
- Product accuracy: 68.2% (cluster sampling)

  72.7% (cluster sampling with 7.5m buffer)

  66.6% (point sampling)

4.8 M1.1 LILLE (EUROSENSE)

Location / phase	Lille
Product/ nomenclature	M1.1
Service Provider	Eurosense
Area (km^2)	554
Processing status (percent of total)	15%
Start date of validation	12. June 2006
End date of validation (1st iteration)	18. July 2006
Photointerpreters	B Kosztra, M Bíró
Technical preparations	G Maucha, R Pataki
Validation software	InterView GSE 1.0

4.8.1 Validation

Input data

Orthophotos / VHRSI	Available. Good quality, but date unknown
Original satellite images	Available. Date: 2005. Image quality appears not optimal for the usage.
Topographic maps	Available.
Other reference material	None.

4.8.2 Validation process

Number of validation clusters	4 (1.25 x 1.25 km)
Number of sampling points	871
Additional iterations	One. With ortho-photos.
Number of additional samples	None.

4.8.3 Comments / remarks

The cluster areas (randomly distributed) are mainly distributed in non-urban parts (>80% of the validation are was on agricultural lands) of the production area and thus avoid those structural more complex areas. The clusters were validated twice, as for the first validation no ortho-photos had been available (corrupted transfer media).

Satellite image has limited quality: no SWIR (short wave infrared) band for vegetation discrimination, low contrast and insufficient sharpness (considering the 0.25 ha MMU).

Distinction between 210 (arable land) and 230 (pastures) was possible based on ortho-photos only, not in the original satellite images. However, the date (year) of the ortho-photo was unknown.

A maximum of 900 sampling points was checked in the point validation. The number of points mentioned above lists those points for which a clear interpretation was possible.

Points falling on or close to land cover boundary have been coded by 999 and were excluded from the above mentioned number of “valid codes”.

4.8.4 Results / findings

Analysing the contingency matrix computed from the SP’s classification and the validation, the largest disagreements (errors) are as follows:

– 210-230: pastures classified as arable land – 23.4% (of total area)

– 210-112: discontinous urban fabric classified as arable land – 2.9 %

– 112-121: industry / commercial classified as discontinuous urban fabric – 1.2%

This means that distinction between arable land and pastures proved to be the weakest point. Mistakes in urban classes contribute to significantly less extent to the inaccuracy of the product.

4.8.5 Overall accuracy

- Target accuracy: 90% +/- 3%
- Product accuracy: 66.4% (cluster sampling)
  72.6% (point sampling)

4.9 M1.1 BARI (PLANETEK)

Location / phase	Bari
Product/ nomenclature	M1.1
Service Provider	Planetek
Area (km^2)	1400
Processing status (percent of total)	About 50 %
Start date of validation	11th Aug 2006
End date of validation (1st iteration)	23rd Aug 2006
Photointerpreters	B Kosztra
Technical preparations	G Maucha, R Pataki
Controlled by	G Büttner
Validation software	InterView GSE 1.0

4.9.1 Validation

Input data

Orthophotos / VHRSI	Available. B&w, taken 7 years before the satellite image
Original satellite images	Available, except SWIR band
Topographic maps	Available, 1:25.000 (colour) 1:50.000 (b&w)
Other reference material	None

4.9.2 Validation process

Number of validation clusters	5 + 1; the 5 randomly selected clusters did not include built-up areas, therefore an

	additional cluster was selected
Number of sampling points	
Additional iterations	None.
Number of additional samples	None.

4.9.3 Comments / remarks

- An extreme amount of efforts was needed to collect the ancillary material, as topographic maps and orthophotos were initially not available for validation. The finally provided reference material is rather old which limits its use.
    - Topographic maps were delivered as vector maps. In a first delivery even without coordinates.
    - Orthophotos are available on a central server, but only for viewing. The data cannot be downloaded or otherwise be integrated into the validation software.
    - The provision of orthoimages to project partners outside of Italy seems to be a general problem. The export of such data is not allowed under the current legislation.
    - The satellite images were initially delivered without coordinates (only pixel coordinates).
- The 7 years difference between the acquisition date of the satellite image and that of the orthophoto made validation difficult.
- Orthophotos were provided in different projection than the topo map or the satellite image. Registration was produced by FÖMI.
- The satellite image lacks a SWIR band, has a low contrast and insufficient sharpness (considering the .25 ha MMU). Consequently several polygons have been coded by 999 (unknown, impossible to interpret).
- The actual number of land cover classes in the validation set is restricted to 8 out of possible 23 classes.
- The urban classes cover less than 20% of the validation area.

4.9.4 Results / findings

Comments:

The major main sources of disagreement between the product and the validation are:

- 230-320: natural grassland / shrub classified as pastures: 11.3%
- 210-320: natural grassland / shrubs classified arable land: 4.0 %
- 230-210: arable land classified as pastures: 2.0%
- The producer accuracy for the classes 112 (discontinuous urban fabric) and 122 (transport network) is quite low (< 30%).

4.9.5 Overall accuracy

- Target accuracy: 90% +/- 3%

- Product accuracy: 69.1% (cluster methodology)
  71.1% (point sampling)

4.10 M1.1 MUNICH (HG-GDS)

Location / phase	Munich
Product/ nomenclature	M1.1
Service Provider	Geo Data Solutions GmbH
Area (km^2)	1195
Processing status (percent of total)	73%
Start date of validation	3rd Aug 2006
End date of validation (1st iteration)	30th Aug 2006
Photointerpreters	B Kosztra
Technical preparations	G Maucha, R Pataki
Controlled by	G Büttner
Validation software	InterView GSE 1.0

4.10.1 Validation

Input data

Orthophotos / VHRSI	62 good quality colour orthophoto, the sampling represents the central part (appr.35%) of the total area
Original satellite images	1 satellite image, covers only 90% of the total area. SPOT-5, Date: 16.10.2002.
Topographic maps	1 large low-quality B&W topomap mosaic, covers only 42% of the total area.
Other reference material	None

4.10.2 Validation process

Number of validation clusters	13 clusters.
Number of sampling points	659 (northern part not yet mapped)

Additional iterations	None.
Number of additional samples	None.

4.10.3 Comments / remarks

- Cluster size is smaller than usual, because of the size of the ortho-photos. Represented area is 1% of total. Location is defined according to the availability of ortho-photos and topomap – only central part is represented.
- A maximum of 900 sampling points was checked in the point validation. The number of points mentioned above lists those points for which a clear interpretation was possible.
- Points falling on or close to land cover boundary have been coded by 999 and were excluded from the above mentioned number of “valid codes”.
- Better spatial coverage of the point sampling than with cluster sampling. 13 orthophotos are included in cluster sampling, and all the 62 available orthos are covered with point sampling (also outside of topo map).

4.10.4 Results / findings

Comments:

Largest disagreements between the Service Provider’s and the validation interpretation are (shown by both validation methods):

- Discontinous urban fabric classified as continuous urban fabric
- Pastures are classified as arable land

Overall accuracy

- Target accuracy: 90% +/- 3%
- Product accuracy: 60.0% (cluster approach)
  63.9% (point sampling)

4.11 M1.1 WEIMAR / JENA (HG-GDS)

Location / phase	Weimar-Jena
Product/ nomenclature	M1.1
Service Provider	Geo Data Solutions GmbH
Area (km^2)	590 km2
Processing status (percent of total)	80%– missing satellite data
Start date of validation	1st Aug 2006

End date of validation (1st iteration)	23rd Aug 2006
Photointerpreters	B Kosztra
Technical preparations	G Maucha, R Pataki
Controlled by	G Büttner
Validation software	InterView GSE 1.0

4.11.1 Validation

Input data

Orthophotos / VHRSI	160 colour orthophoto available in 3 blocks (Apolda/Jena/Weimar). Photos were taken in Sept/Oct 1997.
Original satellite images	1 satellite image, covers only 80% of the total area. SPOT-5, Date: 20.06.2003.
Topographic maps	1:25.000 scale b&w topomaps exist for the entire area
Other reference material	None

4.11.2 Validation process

Number of validation clusters	4 (one of the clusters has been shifted from uninterpreted area)
Number of sampling points	868
Additional iterations	None.
Number of additional samples	None.

4.11.3 Comments / remarks

- Limited image quality: Probable no SWIR band, consequently there are not enough colour differences. Pixel size is OK (2.5 m) but image is not sharp enough (considering the 0.25 ha MMU).
- Urban classes covered about half of the validation area.
- A maximum of 900 sampling points was checked in the point validation. The number of points mentioned above lists those points for which a clear interpretation was possible.

- Points falling on or close to land cover boundary have been coded by 999 and were excluded from the above mentioned number of “valid codes”.

4.11.4 Results / findings

Comments:

- 112-142: sport and recreation classified as built-up – 13.8%
- 111-112: discontinuous urban fabric classified as continuous urban fabric – 10.4 %
- 141-121: industry & commercial classified as urban green area – 1.8%
- 240-142: sport and recreation classified as mixed agriculture – 1.7%
- The producer accuracy of the urban classes lies below 60%.

4.11.5 Overall accuracy

- Target accuracy: 90% +/- 3%
- Product accuracy: 54.8% (cluster sampling)
  71.2% (point sampling)

5. VALIDATION ISSUES AND CONCLUSIONS

5.1 PROCESS RELATED COMMENTS

- Upon request by the service providers and supported by the end users, GSE Land products were validated using two different approaches – the newly developed cluster approach and the previously used point sampling.

  Original target values were based on point sampling, therefore it was decided to communicate both values.

- There is a lack of independent validation data. The availability of reference data (e.g. maps, high resolution image data) is below expectations and needs for the correct execution of an independent product validation.

  Validation on original data sources is possible, but does not fulfil the criteria of independent validation.

- The reference data is partly not in a useable format for validation. This means the data that was delivered cannot be integrated into the validation software, e.g. the data is part of an application (city atlas) and cannot be extracted, topographic maps are not delivered as raster files, quality of the scan is low, data are delivered without coordinates.

- The communication of the cluster location to the service providers only after the finalisation of the mapping process delays the validation process in those cases where the service providers still have to go to their users and obtain the high-resolution image data.

- The selection of cluster locations should consider the actual distribution of land cover classes in the mapping area.

- Corine Land Cover level 1 information should be used to used to control the selection of validation clusters to avoid that clusters fall into the sea or that important land cover classes are not included in the sample (e.g. no urban areas in the validation of M1.1 products).

5.2 PRODUCT RELATED COMMENTS

- The mapping of “functional areas” in product M1.1 is very difficult to validate as their concept is based on land use criteria, which are not possible to validate without sufficient reference material. Even level 3 of the nomenclature is in several cases not possible to achieve in the validation process.

  In addition, within these functional areas some objects are not mapped even though they are larger than the minimum mapping unit (especially green areas).

- The artificial non-agricultural green areas (class 140), also called “green urban areas” caused a particular problem for the urban atlas. The M1.1 product differentiates public and private green areas. While public green areas (publicly accessible and, in most cases also, owned) are separated, private green areas are considered to be part of the urban block (also if they are bigger than the MMU). This also applies to green spaces between apartment blocks, as these green areas do not allow public access.

  The mapping guidelines will be review to review this ambiguity.

- The land use versus the land cover aspect of some classes is an issue. Trees in an “urban context” should be classified as green urban areas (e.g. park), otherwise as forest. This class

is rather well defined for larger urban areas and trees within the urban structure, i.e. park. But trees on the edge of urban areas might be subject to different interpretations.

Similar “problems” exist with pastures and semi-natural grasslands. Both can be used for grazing, but the intensity of use is different as well as the degree of human influence (e.g. fertilization). In central Europe most grasslands in the neighbourhood of settlements should be considered to be under human influence, i.e. pastures.

Also trees on wetlands can cause interpretation problems as they can be classified as wetland or forest.

- The homogeneity and comparability between the Urban Atlas products for the individual sites needs to be improved.
- The mapping quality of non-urban areas of the M1.1 product (Urban Atlas) is not sufficient. Up to 25% of the error contribution stem from these surfaces.

  As a consequence of this fact, the M1.1 specifications were reviewed and the mapping criteria for non-urban classes have been adapted (decrease of minimum mapping unit in non-urban areas to 1 ha, overall accuracy thresholds are still be to discussed).
- The M1.1 urban atlas will provide products with different specifications to the local and the European clients. Only the European product will be validated by the QA team, the local product only by the local user.

  There is a risk that the local product will continue to have the same quality related problems as the products described in this document, as only the symptoms of the problem are addressed in the mitigation, not the cause.
- The mapping of class 130 (mine, dump and construction sites) was difficult for most service providers and has a low producer accuracy in several products.
- The resolution of the input satellite data is often “on the edge” with respect to the minimum mapping unit.