EUROPEAN COMMISSION

Research Executive Agency

Seventh Framework Programme

Cooperation: Space Call 3

FP7-SPACE-2010-1

Grant Agreement: 262371

Enabling Access to Geological Information in Support of GMES

D3.3A: PanGeo Quality Assurance

Version 1

7th August 2012

Dissemination Level:	Public	
Author:	L Bateson, BGS. M Cuevas, IG. G Glanfield, FNPA	Date: 7/8/2012
Checked by (WP Leader):	Stuart Marsh, BGS	Date: 7/8/2012
Approved by (Coordinator):	Ren Capes, FNPA	Date:
Date of Issue:		

CHANGE RECORD

Version 1.0 of 7th August 2012

Section	Page	Detail of change
All		First release

DEFINITIONS

PanGeo Production Pack – PSI data, other information, templates and manuals sent to the geological survey to allow them to produce the Ground Stability Layer and Geohazard Description.

PSI Processing Report – A report, written by the PSIPs, detailing the PSI processing

Ground Stability Layer – The polygons created by the geological surveys outlining areas of ground instability

Ground Stability Polygon – A polygon, within the Ground Stability Layer, drawn around an area of ground instability

Geohazard Description– A summary report, for each PanGeo town, sections linked to each ground stability polygon, provides the geological interpretation for the motion.

PanGeo Production Manual – The instructions for how the geological surveys should create the Ground Stability Layer and Geohazard Description

PanGeo Portal – The INSPIRE compliant web portal through which PanGeo information will be discoverable and available free of charge.

Ground stability– Stability or instability of the ground and everything that is built on it.

Hazard - Something with the potential to cause harm.

Natural Hazard - A natural hazard is a natural process or phenomenon that may cause loss of life, injury or other impacts, property damage, loss livelihoods and services, social and economic disruption or environmental damage. (Council of The European Union – Commission Staff Working Paper – Risk Assessment and Mapping Guidelines for Disaster Management).

GeoHazard (Geological hazard) - A geological process with the potential to cause harm.

Risk - The likelihood that the harm from a particular hazard will be realised.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Within PanGeo Ground Stability Layers are produced, for 52 towns, by 27 different organisations with access to a variety of input datasets. Due to the variety of input datasets and interpretation involved in the creation of the PanGeo service it is difficult to assign an overall accuracy to the PanGeo service.

We address the overall quality of the PanGeo products by consideration of the input data types and the production process. Products for all PanGeo towns will use the PSI and Urban Atlas data in the production process, the quality of these datasets is high and accuracy figures are provided below. Each town will also have access to varying amounts of 'other deformation datasets' and geological data. Experienced interpreters within the national geological surveys will only include high quality datasets therefore their quality is also high.

The PanGeo product and service has been designed by a group who are highly experienced in the interpretation of geohazards from PSI and geological datasets. The result of this process is a detailed procedure which each Geological Survey will follow to produce standard PanGeo Products. These products are then validated by the design team to ensure their quality.

Input	Accuracy	Reliability	Quality
PSI	>2mm/yr	High	High
Other deformation techniques	Typically > PSI	Typically high	High
Urban Atlas	>=85%	High	High
Geological Data	NA	High	High
PanGeo interpretation	NA	High	High

1 INTRODUCTION

This document is the PanGeo Quality Assurance document which aims to provide reassurance to the user as to the, accuracy and reliability of the PanGeo service and its products. In this document, the term 'quality' is used to denote both accuracy and reliability. Accuracy is defined as 'nearness to the truth', and reliability as 'consistency of being fit for purpose'.

The PanGeo product is constructed from a range of input data that has been interpreted by one of the 27 EU geological surveys. The absolute accuracy of a given Ground Stability Layer for a given town will therefore vary depending on the quality of the input data and the level of interpretation.

When considering the question of quality we must consider the following:

1. The input data
2. The interpretation process

These are described in more detail below.

2 QUALITY OF THE INPUT DATA

There are many possible input data types within PanGeo; they can be grouped into the following:

1. **Base Inputs** (those always used to produce PanGeo products
   a. Persistent Scatterer Interferometry
   b. Urban Atlas data
2. **Variable Inputs** (may or may not be used depending upon availability or relevance)
   a. Geological data
   b. Ancillary data

The following sections describe the quality of each input data type in turn.

2.1 PERSISTENT SCATTERER INTERFEROMETRY

Overall accuracy of PSI average annual velocities; **1.0 – 1.8 mm/yr**

Overall accuracy of PSI time series; **3.6 – 8.7 mm**

NOTE; the above figures are based on the findings of the Terrafirma Validation study and are therefore representative of the data processed for Alkmaar and Amsterdam, they are not definitive accuracies for all PSI processing's but do give an idea on the performance of the technique.

PanGeo is built on the achievements of the ESA GMES Service Element (GSE) project Terrafirma. In fact, 27 of the 52 PanGeo towns have been PSI-processed in Terrafirma. The Terrafirma Validation Project (TFVP) was a major PSI validation exercise run as an integral part of Terrafirma Stage 2. During this project, the PSI Providers (PSIPs) involved in PanGeo, i.e. TRE, Gamma Remote Sensing, Altamira Information and Fugro NPA, were "certified" by undertaking a scientifically rigorous validation exercise.

The TFVP consisted of two main parts: a "process analysis" involving the cross-comparison of the output of the different PSI processing chains, including the PSIPs processed slant-range outputs and the analysis of their intermediate results; and a "product validation", in which the geocoded PSI products were validated against ground truth, providing quantifiable tolerances of accuracy.

Two test sites in the Netherlands, Alkmaar and Amsterdam, were selected by consensus amongst all project partners including the PSIPs. Alkmaar is a rural area of forest, dunes, beaches and villages, affected by spatially correlated deformation due to gas extraction. The sparse levelling data available were complemented by modelling of the expected subsidence bowl. This site was studied using ERS (1992-2000), and Envisat (2003–2007). Amsterdam is a typical urban area which includes autonomous and spatially uncorrelated ground motion over the 9.5km long N-S metro line route, which was under construction. Amsterdam was studied using Envisat (2003 - 2007) and surveying, levelling, inclinometer and extensometer data were available since 2001.

The nature of the deformation and the main targets of the PSI analysis were communicated to the PSIPs before the standard PSI processing, corresponding to standard Terrafirma H-1 products, was applied. They were invited to tune their processing, and especially the thresholds on Persistent Scatterer (PS) selection, to obtain a good balance between PS coverage and measurement quality.

Prior to the comparison of the PSI datasets, the origin of the PS time series was set to the first date of each time series, the datasets were all referenced to the same point and trends related to residual orbital errors were removed. In the next stage, differences between the PSIP outputs due to different implementations of the PSI technique were first analysed in radar geometry. DLR used results from its own independent, scientific PSI processing chain as a reference to assess the PSIP results. This procedure was an internal comparison of the different PSI processing chains, recognising that each is based on somewhat different approaches. The next step was the direct inter-comparison of the PSIPs slant-range outputs for deformation velocities, deformation time series, topographic corrections and PS density. **The estimated standard deviations for each PSIP are 0.40 – 0.53 mm/yr for velocities, 1.1 – 4.0 mm for time series and 2.14 – 4.71 m for geocoding.** There is a large difference in the number of PS and the coherence values delivered by each PSIP, which are used as quality indicators within Terrafirma. This rich set of global statistics can be used to provide error bars about the quality of the ground motion estimates derived by PSI, which is key information for the Terrafirma, and therefore PanGeo, end users.

In the final stage, the PSI results of each PSIP were compared to the available ground truth at each site. For Alkmaar, direct velocity validation against levelling shows RMS error of 1.0 – 1.5 mm/yr for ERS, and 1.3 – 1.8 mm/yr for Envisat. Direct time series validation shows RMS error of 6.2 – 8.7 mm for ERS, and 3.6 – 4.8 mm for Envisat. Finally, the comparison of PSI and levelling for the estimated subsidence bowl shows good results, even though the subsidence signal is weak. For Amsterdam, there is a general underestimation of deformation velocity by PSI compared to ground truth. The absolute standard deviation of the double difference in velocity ranges from 1.0 to 1.2 mm/yr. The average RMS errors of single deformation measurements in the time series range from 4.2 to 5.5 mm. The slightly better results achieved at Amsterdam compared to Alkmaar probably relate to their urban and rural settings.

2.2 OTHER DEFORMATION DATA

Data from geodetic surveys such as GPS and levelling data are considered, in general, standard products with a reliability higher than PSI. However, the quality of these data strongly depends on the methods or techniques employed and the equipment used, among other factors. For information on the accuracy and precision of each dataset, refer to the accompanying report or metadata.

2.3 URBAN ATLAS

Overall Thematic Accuracy >=80%

For the purpose of accuracy the Urban Atlas can be split in to five levels of classes, which correspond to the CORINE classes.

The Level 1 class includes all the urban classes. These have a minimum mapping are of 0.25ha, a thematic accuracy of >=85% and a positional pixel accuracy (of the imagery used to derive the UA polygons) of <+/- 5 m.

Level 2 - 5 classes are Agricultural areas, Forests and Water. These have a minimum mapping are of 1ha, a thematic accuracy of >=80% and a positional pixel accuracy (of the imagery used to derive the UA polygons) of <+/- 5 m.

2.4 GEOLOGICAL DATA

The input geological data used for the interpretation will vary in origin from town to PanGeo town. Since the Geological Survey, who own the data, are carrying out the interpretation it is expected that the most up to date, highest quality and finest scale geological data available are used.

In many cases, such as geological maps or models, the geological data represents a trained individual's interpretation of the observed facts. It is therefore impossible to state the absolute accuracy of a geological map. As such maps produced by the British Geological Survey have the following statement:

"This maps gives and interpretation of the data available at the last date of survey"

Some geological data might be considered less accurate; such as secondary data derived from the primary data by means of the application of assumptions and models.

2.5 ANCILLARY DATA

Once again the input ancillary data used for the interpretation will vary in origin, it is expected that the interpreters will use their judgement and discount any data which they perceive to be of poor quality.

3 QUALITY OF THE INTERPRETATION PROCESS

It is difficult to assign a measure of accuracy to the interpretation process since it is carried out by at least 27 individuals; however measures have been put in place to ensure a consistent, high quality result:

3.1 RIGOROUS SERVICE DESIGN

The PanGeo service was designed over a period of one year by individuals with many years experience of interpreting ground motion and geological data. The design process involved the compilation of the needs and requirements of all geological surveys involved. The service and product was designed with these in mind to ensure that the surveys had the resources to produce the products. The design process was then iterated a number of times and has resulted in a tightly specified product and a standardised procedure.

3.2 DETAILED, STEP BY STEP INSTRUCTIONS

The procedure is documented in the PanGeo Production Manual; a series of step by step instructions followed by the geological surveys to create the products. The manual has been tested by geological survey staff not involved in the product design and specification, testing feedback led to an iteration of the manual.

3.3 TRAINING

Two training courses have been arranged for the participating geological surveys to attend, the first focused on the PSI data, its characteristics and nuances. This has ensured that all geological surveys are aware of how best to interpret the data. The second training course focuses on the procedure followed to create the products. Importantly it also provides the opportunity for the geological surveys to ask questions and ensure they have a full understanding of what is required.

3.4 VARIABLE ACCURACY AND CONFIDENCE

The PanGeo interpretation is to be carried out at 1:5 000 scale to ensure it is viewable at 1:10 000 scale. Although the creation scale does not guarantee the accuracy of the product it does allow a nominal accuracy of 200m to be given to the placement of a boundary line (which is assumed to be 1mm thick). i.e. in a clear cut case, where no interpretation is necessary, the boundary can be drawn within 200m of its actual location.

Within the GSL the accuracy will vary from polygon to polygon depending on the quality of the input data. To allow this to be reflected in the product the interpreter assigns a confidence level to the polygon. This reflects the interpreters overall confidence in the location of the polygon and its geological interpretation.

GSL polygons can be either potential or observed, observed polygons are those where instability has been observed on a dataset; they are therefore likely to have a higher accuracy than potential polygons. Polygons representing areas which have the potential to be unstable are, by their nature, less accurate.

3.5 GEOLOGICAL INTERPRETATION

The accuracy of the geological interpretation will also vary from polygon to polygon. In some cases the interpreter will be certain that the reason given is the actual reason for the motion, however in other cases the interpreter might be less certain that the reason given is the actual reason for motion. In such as case it is still valuable to have the interpreters thoughts recorded as the reason for motion. The confidence level attribute once again reflects the interpreter’s confidence in their interpretation.

4 CONFORMANCY FILTER

Before any PanGeo products will be made available on the portal they will be checked by the British Geological Survey. This 'conformancy filter' ensures that all products meet the specification as set out in D3.3 the service and product specification. While it is not possible for the BGS to validate the position of each polygon or check that the Geological Interpretation is correct it is possible to ensure that the description is sensible and well founded. Automated tools will check the GSL polygons to ensure they meet the basic GIS rules such as closed polygons and full attributes. It is only once these checks have been passed that the GSL will appear on the portal.

5 VALIDATION TOOL

Manually validating each GSL was foreseen as a potentially time consuming task (some cities are expected to have over 200 polygons), so it was decided to automate as much as possible.

A user uploads a zipfile (containing the GSL shapefile) to the PanGeo website. Scripts are then automatically initiated which run checks as outlined below. If the tool fails one of these checks then the user is shown what has failed and why. This process is repeated until all the checks have been passed, at which point BGS are sent an automated email informing them that user x has successfully passed the validation test (the GSL data is attached to the email).

5.1 AUTOMATED VALIDATION CHECKS:

- Essential files are in zip file (.shp, .dbf,.prj).
- Shapefile is of type polygon, with records and geometry having a one-to-one relationship (not multi-polygons).
- Projection is Geographic (EPSG:4326).
- Any date entries are between 1800 and 2050.
- Column names and data types are correct (as specified in spec & template shapefile).
- Records in “inspireid” column begin with “PGGH_”
- Records in “HazCat” column contain one of the following values: "1_DeepGroundMotions","2_NaturalGroundInstability","3_NaturalGroundMotion","4_AnthropogenicGroundInstability","5_Other","6_Unknown"
- Records in “HazType” column contain one of the following values: "1_1Earthquake", "1_2TectonicMovements", "1_1Earthquake", "1_2TectonicMovements", "1_3SaltTectonics", "1_4VolcanicInflationDeflation"
- Records in “Determin” column contain one of the following: "1_ObservedPSI", "2_ObservedOtherDefMeasurement", "3_ObservedGeologyFieldCampaigns", "4_PotentialInstability"
- Records in “Confidence” column contain one of the following: "Low","Medium","High","External"
- Records in “Area” column are greater than 0 and to three decimal places.
- All inspireIDs are unique.
- Geometry validity checked (closed polygons etc).
- No overlapping potential polygons with the same geohazard type.
- No overlapping observed polygons of the same geohazard type.

If for example, a record in the “HazCat” column has a value of “2_N aturalGroundInstability” then it will fail the validation test due to the extra space after the ”2_N”. Checking that these values are exactly the same ensures that future styling and querying is consistent across all the PanGeo GSL datasets.

As well as the tool validating data, it additionally computes various statistics, incorporating other data sets such as the Urban Atlas. This automated process will allow statistics such as “ Show me the land cover type, area and affected population for any single/multiple GSL polygon(s)”. This information will be accessible on the PanGeo website.

5.2 VALIDATION PROCEDURE

1. Create a Zip file containing **all** component files of the Ground Stability Layer shapefile/ GeoDataBase. This Zip file should have the name *" <town_name>_ground_stability_layer.zip".*
2. The Survey will log onto www.pangeoproject.eu/validate-shp (using the username & password sent out for the initial questionnaire, a reminder will be sent).
3. After logging on a "Validating Tool" link will be visible above the "Logout" link on the top left corner of the web page.
4. Clicking on this link opens a page with more detailed instructions than given here.
5. The page allows the Survey to upload a Zip file containing the GSL shapefile and PDF of the Geohazard Description.
6. The Zip file and shapefile names should contain **no spaces or special characters and be lowercase.**
7. The server performs the data-validation checks.
8. If it passes it will be sent on to BGS with an email stating who uploaded it and when. The up-loader is also copied in on the email so they know it has passed.
9. If it does not pass then the reasons for not passing are displayed on the web page. No email is sent to BGS.
10. The user then rectifies any problems until the GSL passes the checks.
11. Once the GSL has passed the automated checks the GS uploads the MA word document version of the GHD to the relevant town folder in the PanGeo document Library and notifies BGS that it has been uploaded (lbateson@bgs.ac.uk).
12. BGS take further steps as necessary to complete the compliancy-checking of the GSL and GHD.
13. BGS email the GS with the results of the further checks and grants them permission to go ahead and make the products accessible to the portal.
14. In no event should the GS make the GSL and GHD accessible through the portal until BGS has provided the confirmation.

6 CONCLUSIONS

Due to the nature of the PanGeo service and its products it is not possible to assign a definitive number for the accuracy. However it has been possible to consider the quality and accuracy of the inputs to the service, and also the design and specification of the service to provide an assessment of its quality. The following table provides a summary of the quality elements.

Input	Accuracy	Reliability	Quality
PSI	>2mm/yr	High	High
Other deformation techniques	Typically > PSI	Typically high	High
Urban Atlas	>=85%	High	High
Geological Data	NA	High	High
PanGeo interpretation	NA	High	High