|  |
| --- |
| **CDDA v16**  **Data quality assessment results**  **Luxembourg** |

**Prepared by / compiled by: ETC/BD (philipsen@space4environment.com / loehnertz@space4environment.com)**

**Organisation: ETC/BD / space4environment**

**EEA project manager: Mette Palitzsch Lund (Mette.Lund@eea.europa.eu)**

**Task manager: Sabine Roscher (sabine.roscher@mnhn.fr)**

Contents

[1 Background and Introduction 3](#_Toc532978201)

[1.1 CDDA 3](#_Toc532978202)

[1.2 Definition of terms 5](#_Toc532978203)

[2 Overview quality checks on the delivery 6](#_Toc532978204)

[2.1 The automatic QC 8](#_Toc532978205)

[2.2 ETC- QC 10](#_Toc532978206)

[3 QC results for the delivery 11](#_Toc532978207)

[3.1 CDDA delivery on CDR and results of automated QC 11](#_Toc532978208)

[3.2 Results from ETC/BD evaluation 12](#_Toc532978209)

# Background and Introduction

The present document describes the activities and procedures for the verification of the CDDA version 16 (as reported in 2018). The document provides quality information and comments on the delivered spatial and tabular data.

## CDDA

The Nationally designated areas inventory (CDDA, formerly known as “Common Database on Designated Areas”) holds information on protected areas and the national legislative instruments, which directly or indirectly create protected areas. The CDDA dataset contains data on individual nationally designated sites and types of designations in EEA member- and collaborating countries. The CDDA data are delivered by each country as tabular dataset and as spatial dataset. Using the CDDA template and the specifications, the countries update the tables with the national CDDA information and deliver the database and the associated spatial datasets created or adapted by the country.

The CDDA reporting can be divided into two different types of delivery:

* Type 1 – which comes from and are defined by the INSPIRE Protected Sites
* Type 2 – which comes from the revised CDDA tabular data

**Type 1** includes the spatial data and overlaps with certain CDDA reporting elements.

**Type 2** includes the remaining tabular CDDA reporting elements. The Type 2 part delivery is a table with considerable similarity to the current CDDA database specifications. Field names and code lists are however all revised and new. Both Type 1 and Type 2 data files hold mandatory CDDA information.

The two parts of the CDDA reporting (Type 1 and Type 2) are linked by use of a common identifier at data set level as well as on object level:

Figure 1‑1 The two parts of the CDDA reporting

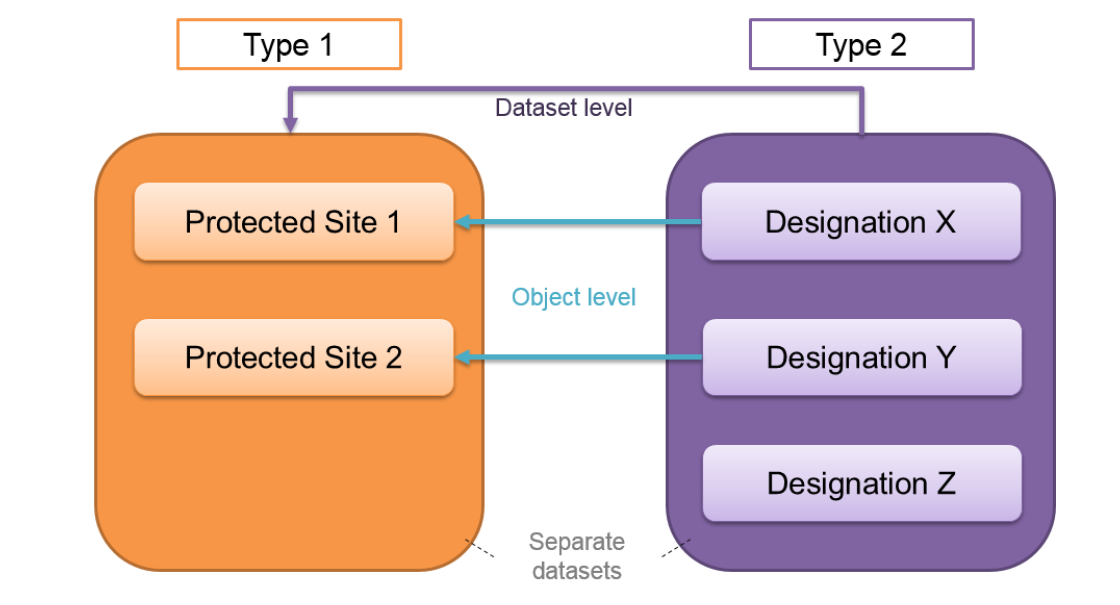
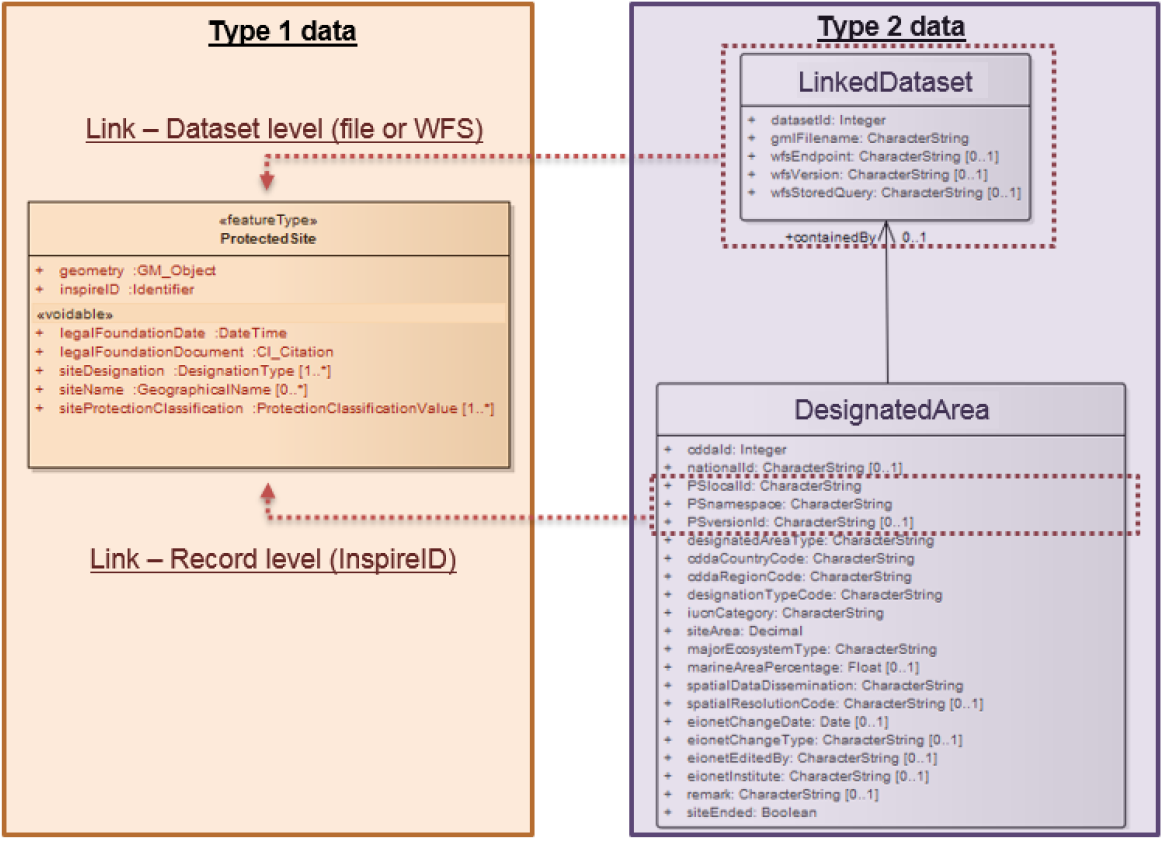


Figure 1‑2 CDDA model in UML INSPIRE Protected Sites (left) and the DesignatedArea and the LinkedDataset tables (right)



Additional specifications about the CDDA delivery can be found here:

<http://cdr.eionet.europa.eu/help/cdda>

## Definition of terms

Before entering into the details of the QA/QC process of the CDDA database, it is important to understand the definitions of these terms.

Table 1‑1 Definition of terms

|  |  |
| --- | --- |
| **Validation / Quality control (QC)** | Validation is the process by which the accuracy and consistency of products are evaluated and the associated uncertainties are quantified (Justice et al., 2000).  Product accuracy is assessed by a comparison with independent data sources such as ground-based measurements, more detailed data or well-calibrated models.  Inter-comparison with other equivalent products is also part of the validation process allowing building up a community reference product when no or not enough independent data are available.  Quality control, or QC for short, is normally carried out after the end of the production and aims at providing the user with measurable / quantitative information how well the product meets the pre-defined specifications. |
| **Verification / Quality assurance (QA)** | The act of reviewing, inspecting, testing, checking, auditing, or otherwise establishing and documenting whether items, processes, services, or documents conform to specified requirements.  Verification is a qualitative process in which intermediate or final results of the production process are commented and potential deviations from the specifications are highlighted. The verification will be done during the course of production and is meant to increase data and production quality.  Quality Assurance (QA) is a way of preventing mistakes or defects in products and avoiding problems when delivering solutions or services to customers.  QA is applied to physical products in pre-production to verify what will be made meets specifications and requirements, and during manufacturing production by validating whether lot samples meet specified quality controls.  QA is also applied to software to verify that features and functionality meet business objectives, and that code is relatively bug free prior to shipping or releasing new software products and versions. |

# Overview quality checks on the delivery

The spatial and tabular data were checked by **EEA** and **ETC/BD** at two points during the CDDA production workflow:

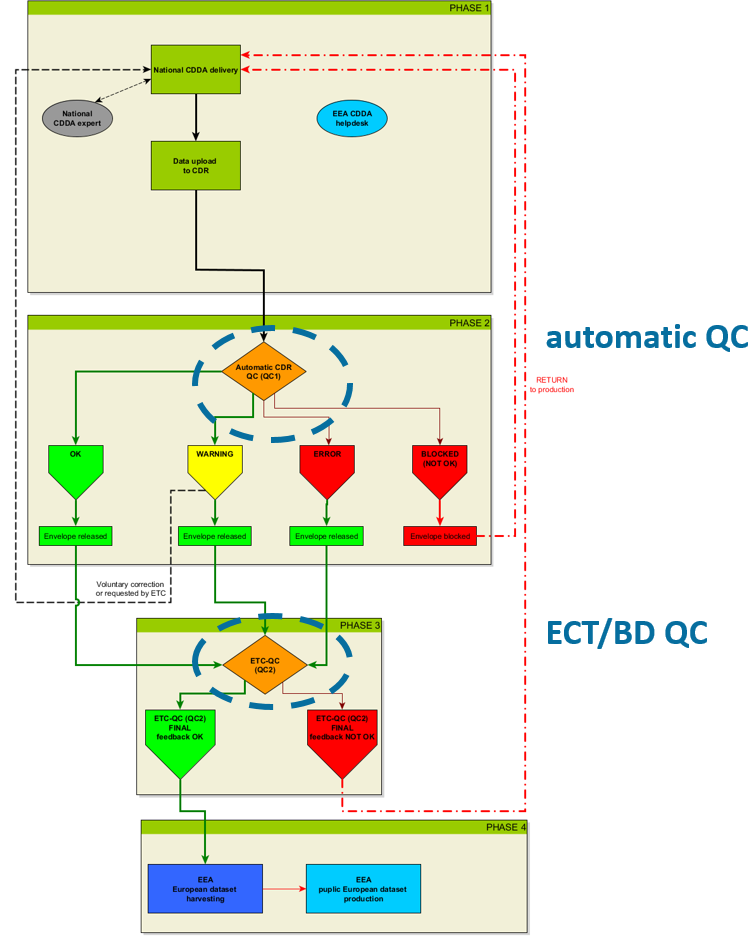
* Automatic QC
* ETC/BD QC

Automatic quality control can be initiated after uploading the data to the CDR. After completion of the automatic QC, a QC report is published on the CDR. If the report does not contain any blockers (i.e. QC checks that indicate missing or wrong elements), the next step can be started – the ETC/BD QC.

Figure 5.1 shows a simplified representation of CDDA production. The workflow is divided into four phases:

* phase 1 - national CDDA production and upload to CDR
* phase 2 - automatic QC
* phase 3 – ETC/BD QC
* phase 4 - European CDDA dataset production

Figure 2‑1 CDDA workflow (the two QC steps are marked with a blue dotted circle)



## The automatic QC

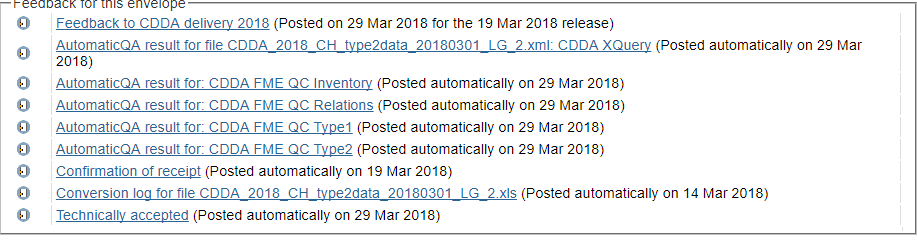
After uploading the data and releasing the envelope, automatic QC can be started.

The following checks were carried out during the automated checks in CDR:

* TEST 1 on Type 2 data inventory (Designated Area (a) and Linked Dataset (b))
  + 1a./ 1b. Mandatory values test
  + 2a./2b. Record uniqueness test
  + 3a./3b. Data types test
  + 4a. Valid codes test
* TEST 2 on Type 1 data inventory
  + 1. Inventory test – ProtectedSite
  + 2. Inventory test – DesignatedArea
  + 3. InventoryTest – LinkedDataset
* TEST 3 relation test between type 1 and type 2 data
  + 1.a/b. Presence test - gml file
  + 2.a/b. Relational test - DesignatedArea link to ProtectedSite
  + 2.c Relational test - DesignatedArea link to ProtectedSite through LinkedDataset
  + 3. Geometry coherence test
  + 4. Relational test - ProtectedSite link to ended DesignatedArea
* TEST 4 Type 1 data QC
  + 1. Mandatory values test - ProtectedSite
  + 2. Uniqueness test - ProtectedSite
  + 3. Coordinate reference system test
  + 4. Geometry validity test
  + 5. Positional check
  + 6. Format test - legalFoundationDate
  + 7. Range test – legalFoundationDate
* TEST 5 Type 2 data QC
  + 1.a Conditional mandatory value test - containedBy
  + 2.a Uniqueness test - DesignatedArea records - cddaId
  + 2.b Uniqueness test - DesignatedArea records - PSlocald, PSnamespace
  + 3. Reference test - cddaId
  + 4. Missing data test - cddaId
  + 5. Reference test - cddaCountryCode and cddaRegionCode
  + 6. Reference test - designationTypeCode
  + 7. Logical coherence test - designationTypeCode, cddaCountryCode and cddaRegionCode
  + 8. Logical coherence test - majorEcosystemType and marineAreaPercentage
  + 9.a Relational test - DesignatedArea link to LinkedDataset
  + 9.b Relational test - LinkedDataset link to DesignatedArea

After the tests an automatic feedback on EIONET with the QC results is published.

Figure 2‑2 QC feedback on the CDR (overview of the QC feedback links)



Four different types of feedback are returned:

* OK – Data check has been executed without generating any warning or error messages.
* WARNING – Warning messages indicate issues that may be an error. Data reporters are expected to double-check relevant records. Some information for future deliveries or questionable data content.
* ERROR – Errors are major issues which are not blocking the release of the envelope this year.
* BLOCKER- Blocker messages indicate that the detected error will prevent data submission (envelope release is not possible).

More information about the single error levels can be found here:

<http://www.eionet.europa.eu/dataflows/2016/criteria>

The following figure shows an example of the “OK” type feedback:

Figure 2‑3 QC feedback on CDR (type 2)



If no blocker appears in CDR, the directory is passed to the ETC/BD QC procedure.

## ETC- QC

The ETC/BD QC is a combination of automatic and manual quality checks.

The first part of the manual check by the ETC contains the checking of issues returned and highlighted in the automatic QC. Depending on the issue the causes for these are analysed and the need for a potential correction is evaluated. The results of this recheck of the returned issues (WARNINGS / ERRORS) is communicated to the country for further clarification and potential correction via the CDDA Helpdesk. For specific significant issues identified already in the automatic QC routines, which influence the data quality, a correction is coordinated with the responsible institution in the country to re-upload revised data. For minor issues and issues which result from elements only mandatory starting with 2019, the country is only informed. It is up to the country in such cases to decide if they will deliver revised datasets or leave the reported datasets as they are.

In the second part of the ETC/BD QC further checks are performed on the delivered data mainly concerning the spatial part (type 1) and detailed assessment of issues returned in the automatic QC. This part consists of comparing the delivered data to the previous reporting, reference information such as country borders or marine territories and further checks for logical consistency (see also figure 2.1).

In detail the following specifications are checked:

1. Examination of check results from automatic QC
2. Logical consistency of the coordinate reference system used
3. Examination of geometry issues returned in the automatic QC
4. Control of protected area positions found to be outside the country territory (comparison to European country border dataset and Marine regions EEZ outlines)
5. Significant differences in site areas defined in type 2 data and the actual area described by the geometries in the type 1 data (buffer of 10% acceptable differences applied)
6. Logical consistency of defined major ecosystems for protected areas
7. Spatial consistency compared to previous reporting identifying partial or systematic shifts in geometries. Check for differences in delineation such as decreased resolution….
8. Check of type 2 content for logical consistency
9. Comparison of sites names between current and previous reporting to identify potential erroneous links or use of siteCode ↔ cddaId relations
10. Rerun of automatic QC

# QC results for the delivery

In this chapter the major issues of the delivery are listed and discussed.

## CDDA delivery on CDR and results of automated QC

The colouring in the table below is based on the colours of the error messages on CDR.

|  |  |
| --- | --- |
| **Name** | **LU - Luxembourg** |
| Delivery date: | 26 Apr 2018 |
| CDR link | http://cdr.eionet.europa.eu/lu/eea/cdda1/envwuhj8q/ |
| autom.QC\_envelope |  |
| CDDA XQuery | **OK – All tests passed successfully** |
| CDDA FME QC Inventory | **OK – All tests passed successfully** |
| CDDA FME QC Relations | **OK – All tests passed successfully** |
| CDDA FME QC Type1 | **OK – All tests passed successfully** |
| CDDA FME QC Type2 | **OK – All tests passed successfully** |

## Results from ETC/BD evaluation

The tables below list the main results derived from the QC checks performed subsequently to the initial CDR based QC.

Points written in *brown colour* *and italic font* indicate issues which have been found to be minor weaknesses in the datasets. In case major issues were found the errors are highlighted in **red and bold** colour.

|  |  |
| --- | --- |
| Type 1 - spatial data: |  |
| Delivery OK? | **OK** |
| Basic information | Projection: EPSG:3035  Geometry: no errors found  Count of sites: 125 |
| Points & poly mapping the same sites? | No points delivered |
| Type 1 – SRS check | OK |
| Type 1 – Area check | *For 10 out of 125 sites, the calculated spatial area differs by more than 10 % from the declared area.* |
| Type 1 – major ecosystem check | OK (only terrestrial sites) |
| Type 1 – additional checks based on Reportnet QC | No errors found in CDR QC checks |
| Type 1 – consistency against previous delivery | OK (no shifts or systematic errors found) |
| Type 1 – logical consistency polygon vs. point delivery | NA |
| Type 2 - tabular data: |  |
| Delivery OK? | Overall result: **OK**   * XML format OK? **OK**   Comments: see comments below |
| [DesignatedArea] | Count of designated areas: 125 Count of designated boundaries: 0  Count of sites flagged “to be deleted”: 0  Comments: None |
| [LinkedDataset] | Count of records: 1  Comments: |
| Name check for consistency with previous reporting’s | OK |

|  |
| --- |
| **Concluding remarks:**  **In summary both the automatic QC as well as our checks have not identified any issues to be corrected. In consequence the reporting is accepted.** |

|  |
| --- |
| CDDA v16 - Map |
|  |

In case of further questions please do not hesitate to contact us.

**The CDDA Helpdesk Team / ETC-BD**

Kongens Nytorv 6, Copenhagen, Denmark

Email: [cdda.helpdesk@eionet.europa.eu](mailto:cdda.helpdesk@eionet.europa.eu) - Web: <http://www.eionet.europa.eu>