# Using metadata quality to filter datasets in data portals

Susanne Feistel[1], Romina Ihde[4], Ulrike Kleeberg[2], Jörn Kohlus[3], Rainer Lehfeldt[4], Carsten Schirnick[5], Stefanie Schumacher[6], Susanne Tamm[7]

## Good Practice

Quality information may be stored in a practical number of specified recommended fields of the metadata standard. Therefore data portals with specific filter options can increase dissemination and availability of data with documented quality information.

## Creating accessible quality information

| **Document your workflow:** | **Put quality information in ISO metadata:** |
|---|---|
| 1. Purpose of your research | → MD_Identification.purpose |
| 2. Subject of your research | → MD_Identification.abstract |
| 3. Data sampling or sources | → LI_Source, LI_Lineage |
| 4. Geo-reference your data | → EX_Extent, MD_Identification.spatialResolution |
| 5. Data processing steps | → LI_ProcessSteps |
| 6. Quality Control | → DQ_QualityElement, DQ_EvaluationMethod |
| 7. Availability of your data | → MD_Distribution, MD_Constraints |

## Creating a quality flag

**Create standardized string as quality flag:**

Batch xml analysis via software

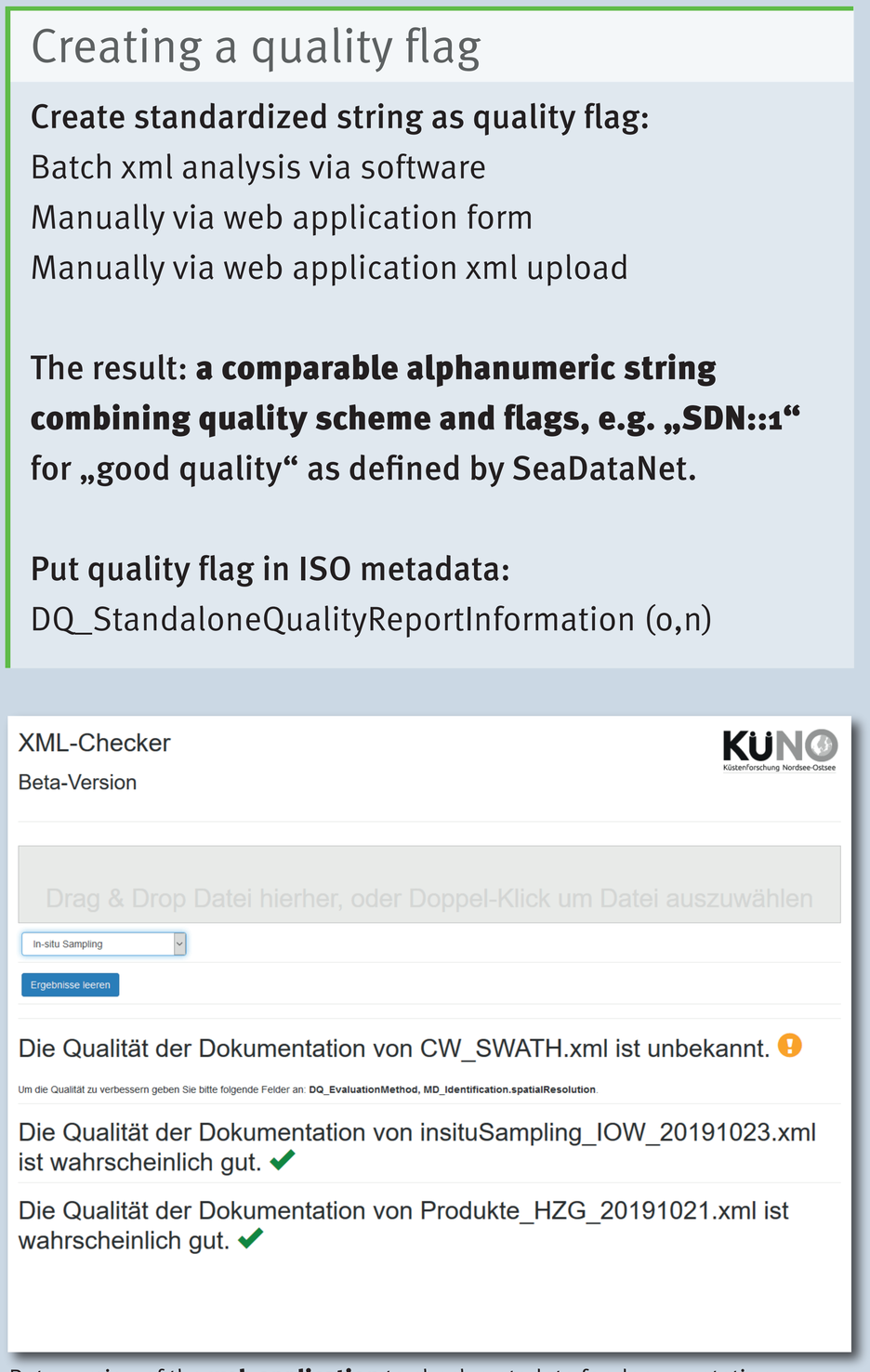
Manually via web application form

Manually via web application xml upload

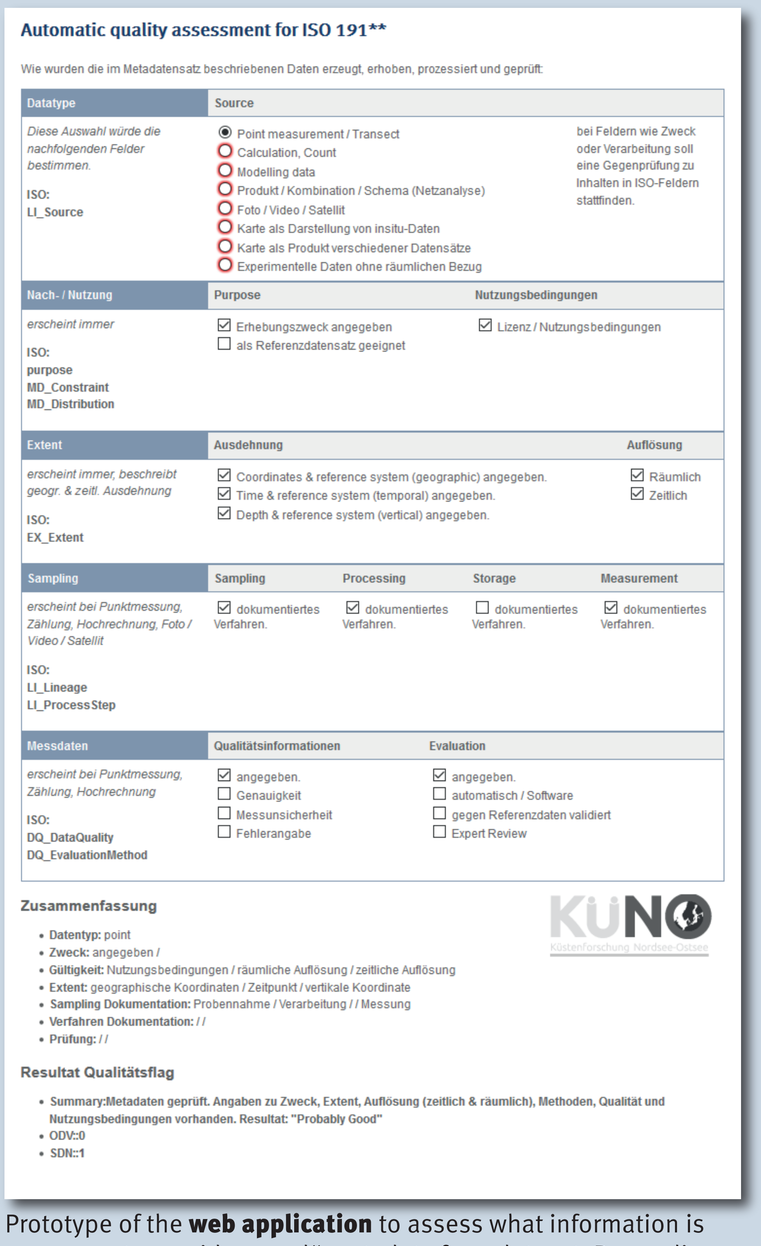
The result: **a comparable alphanumeric string combining quality scheme and flags, e.g. „SDN::1“** for „good quality“ as defined by SeaDataNet.

**Put quality flag in ISO metadata:**

DQ_StandaloneQualityReportInformation (o,n)




Beta version of the **web application** to check metadata for documentation quality. A drag & drop field allows the **direct upload** of a file in XML-format (ISO 19139). According to the selected datatype, the metadata is then checked for the necessary information (described in the web form to the right and in the table below). A **quality rating** and possibly information about missing elements are the results. (Result examples from NCEI, IOW and HZG)




Prototype of the **web application** to assess what information is necessary to provide „good“ metadata for a dataset. Depending on the type of data the web form compiles the given information and displays the probable quality flag. (Example: point data)

## Discussion

**Pro:**

- applicable for all data types
- minimal quality information
- motivation to document
- self assessment of quality
- objective assessment
- easy implementation

**Contra:**

- very generic
- no content analysis possible

**Next:**

- data type specific quality information

**Documenting data with standardized metadata for quality.**

## Outlook: Rating in data portals

**Quality checked:** „(Probably) Good“
**„SDN::1“ || „ODV::0“ || „QUARTOD::3“**

**Quality checked:** „Questionable“
**„SDN::3“ || „ODV::4“ || „QUARTOD::2“**

**Quality checked:** „Unknown“
**„SDN::0“ || „ODV::1“ || „QUARTOD::0“**

Quality unchecked

| | Good Practice Documentation | | | | | | | | ISO 191** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | combining multiple source data & run algorithms | **Modelling** | combining multiple source data & create visualization, maps, etc. | **Products** | photo, audio, video | **Media** | in the broadest sense | **In-situ Sampling** | occurence | obligation | |
| **ABOUT THE DATA** | | | | | | | | | | | |
| **WHY** 1. Purpose of your research | | X | | X | | X | | X | 1 | o | **MD_Identification.purpose** |
| | | | | | | | | | | | |
| **WHAT** 2. Subject of your research 3. Data sources | | X | | X | | | | | n | o | **LI_Source** |
| | | X | | X | | | | | 1 | o | **CI_Citation** |
| | | X | | X | | X | | X | 1 | m | **MD_Identification.abstract** |
| | | X | | X | | X | | X | n | o | **MD_Identification.descriptiveKeywords** |
| **WHERE & WHEN** 4. Geo-reference your data | | X | | X | | X | | X | n | o | **MD_Identification.spatialResolution** |
| | | X | | X | | X | | X | n | c | **EX_Extent** |
| **HOW** 3. Data sampling 5. Data processing steps 6. Quality Control | | X | | X | | X | | X | n | o | **LI_Lineage** |
| | | X | | X | | X | | X | n | c | **LI_ProcessStep** |
| | | X | | X | | | | | n | m | **DQ_Element** |
| | | X | | X | | X | | X | 1 | o | **DQ_EvaluationMethod** |
| **WHO** | | | | | | | | | n | o | **CI_Contact** |
| **AVAILABILITY** 7. Availability of your data | | X | | X | | X | | X | n | o | **MD_Distribution** |
| | | X | | X | | X | | X | n | c | **MD_Constraint.accessConstraint** |

** ISO 19107, ISO 19111, ISO 19115 (2014), ISO 19139, ISO 19156, ISO 19157

[1] susanne.feistel@io-warnemuende.de
[2] ulrike.kleeberg@hzg.de
[3] joern.kohlus@lkn.landsh.de
[4] rainer.lehfeldt@baw.de
[4] romina.ihde@baw.de
[5] cschirnick@geomar.de
[6] stefanie.schumacher@awi.de
[7] susanne.tamm@bsh.de