JÜLICH
FORSCHUNGSZENTRUM

A Data Framework for Assessing Heterogeneous Observation Data

30th September 2014 | TERENO International Conference 2014
Anusuriya Devaraju, Ralf Kunkel, Juergen Sorg, Heye Bogena, Harry Vereecken
Email : a.devaraju@fz-juelich.de

Presentation Outline

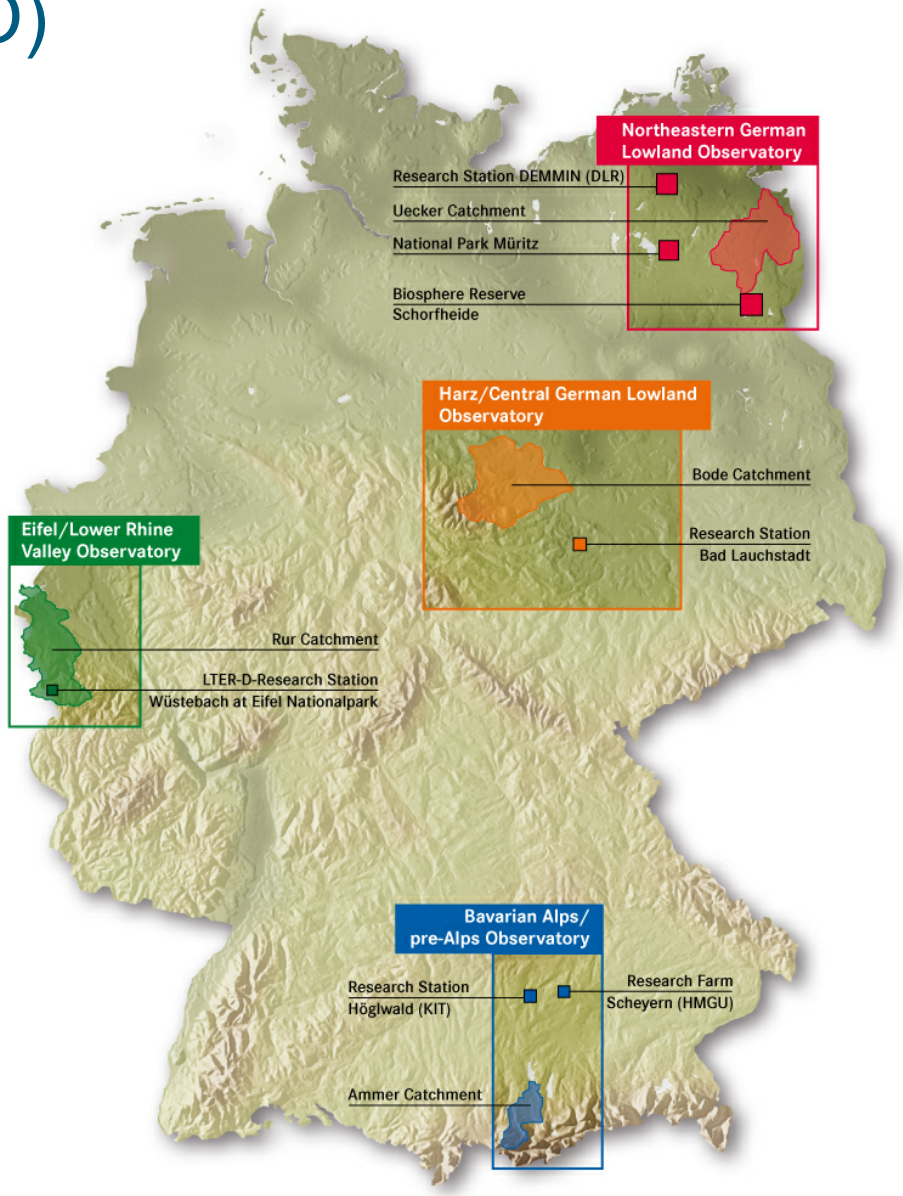
1. • Introduction to TERENO Data Infrastructure
2. • Motivation & Challenges
3. • Quality Assessment (or Quality Control) Framework
3. • Conclusions

Presentation Outline

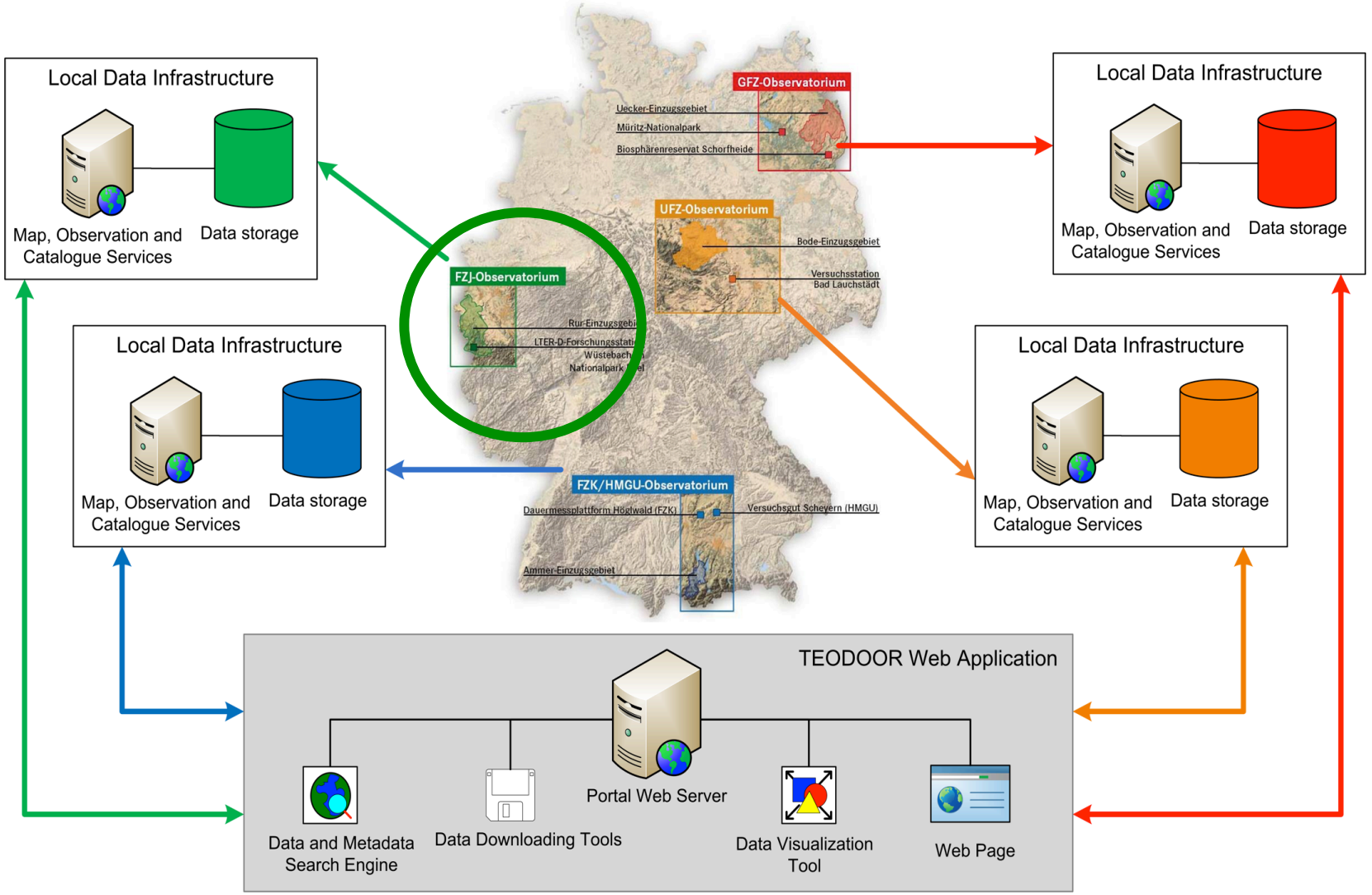
1. Introduction to TERENO Data Infrastructure
2. Motivation & Challenges
3. Quality Assessment (or Quality Control) Framework
3. Conclusions

Introducing Terrestrial Environmental Observatories (TERENO)

- Northeastern German Lowland Observatory
- Harz / Central German Lowland Observatory
- Eifel / Lower Rhine Valley Observatory
- Bavarian Alps / pre-Alps Observatory



TEODOOR Spatial Data Infrastructure



Presentation Outline

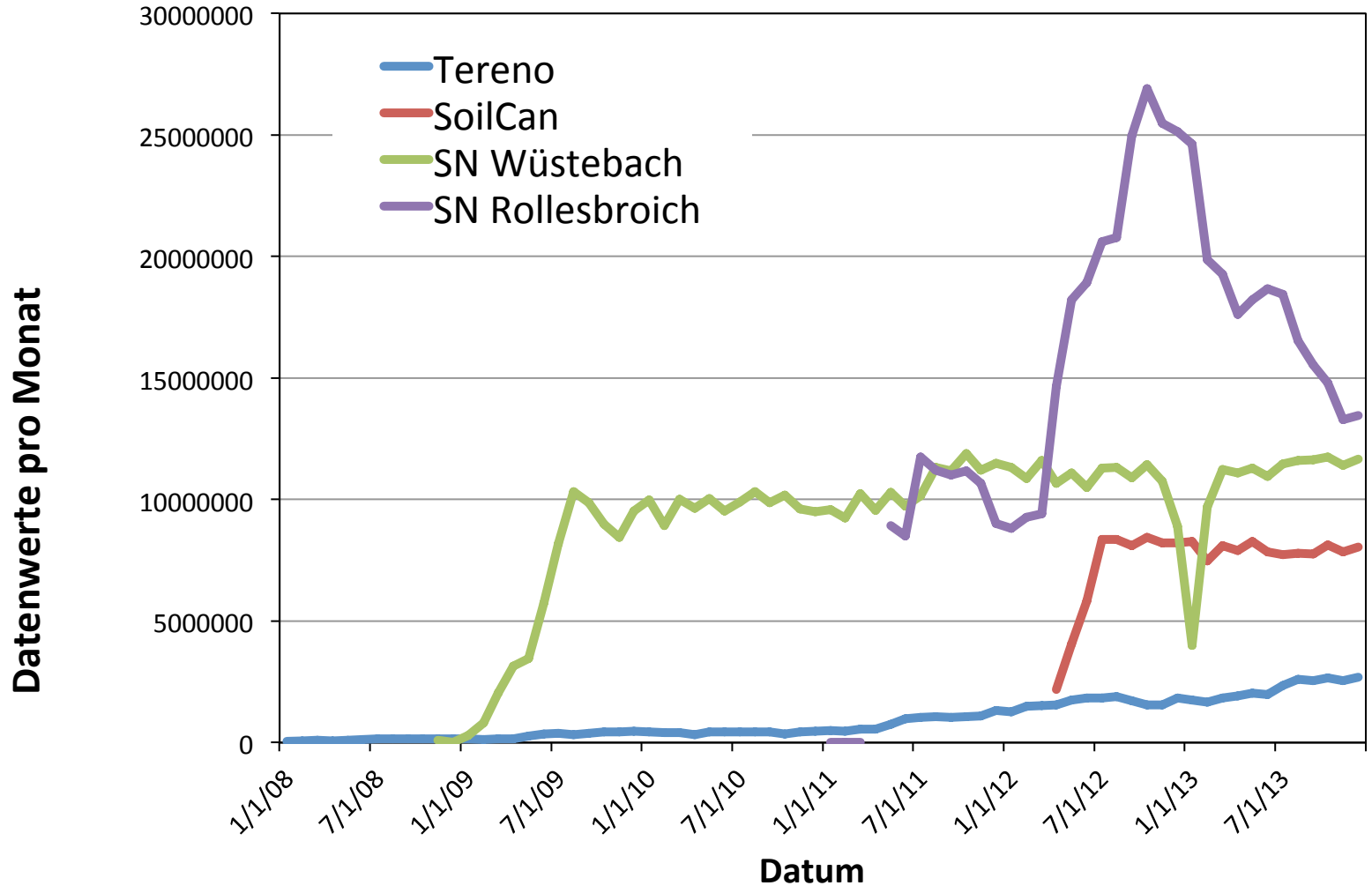
1. Introduction to TERENO Data Infrastructure
2. Motivation & Challenges
3. Quality Assessment (or Quality Control) Framework
3. Conclusions

Sensors Deployed at Eifel/Lower Rhine Observatory

Sensor Types	Count
Climate, soil, water	589 stations
EC flux data	7 stations
Weather radar	2 devices
SoilCan	36 lysimeters

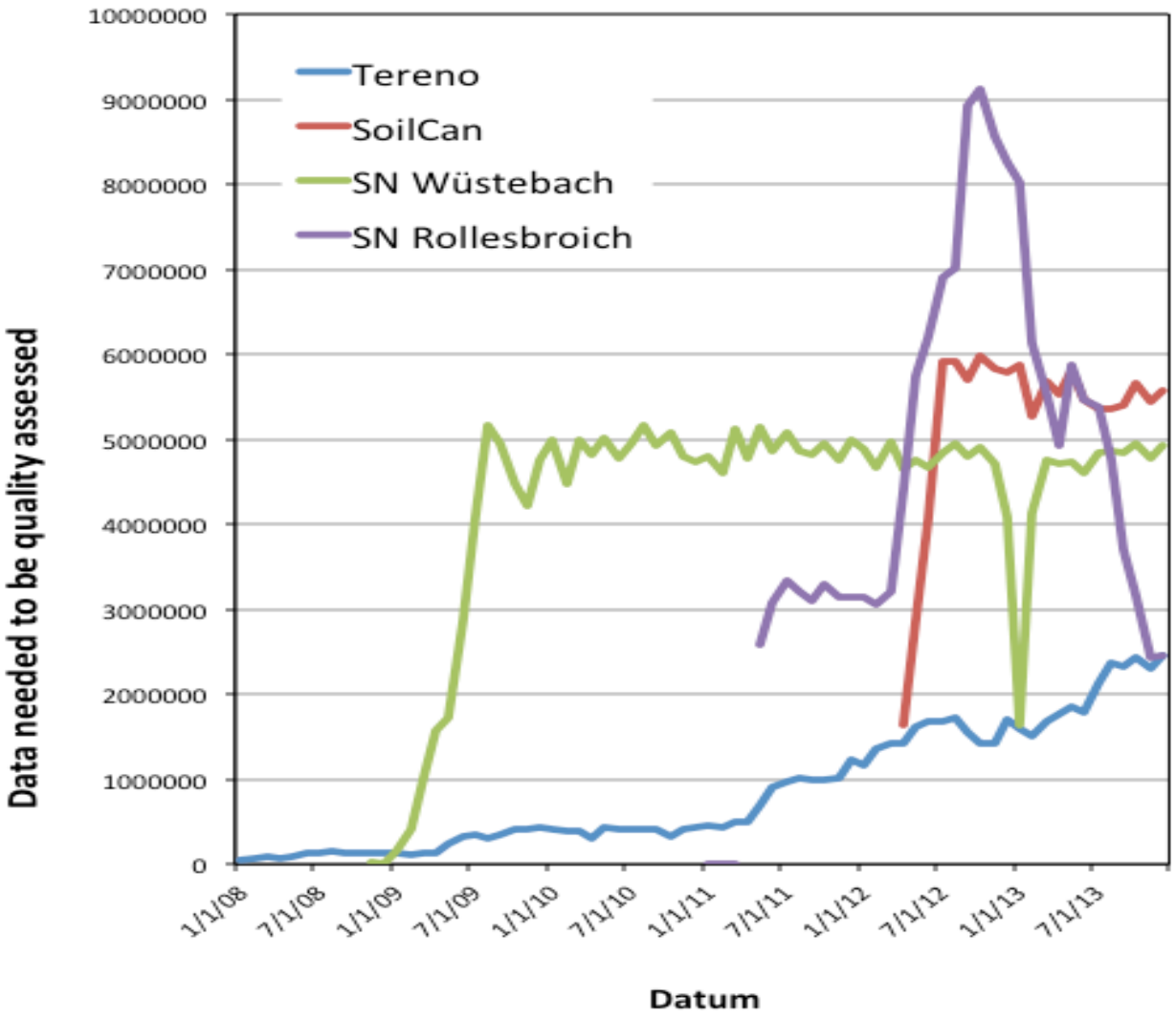
Heterogeneous Observation Data

Huge Amounts of Observation Data



Observation data processed at FZJ

Quality Assessment – Not Too Easy!



We are buried in data!!!

The **challenges** are to (a) process and assess heterogeneous observation data in a systematic way; (b) make these descriptions available to data consumers.

Presentation Outline

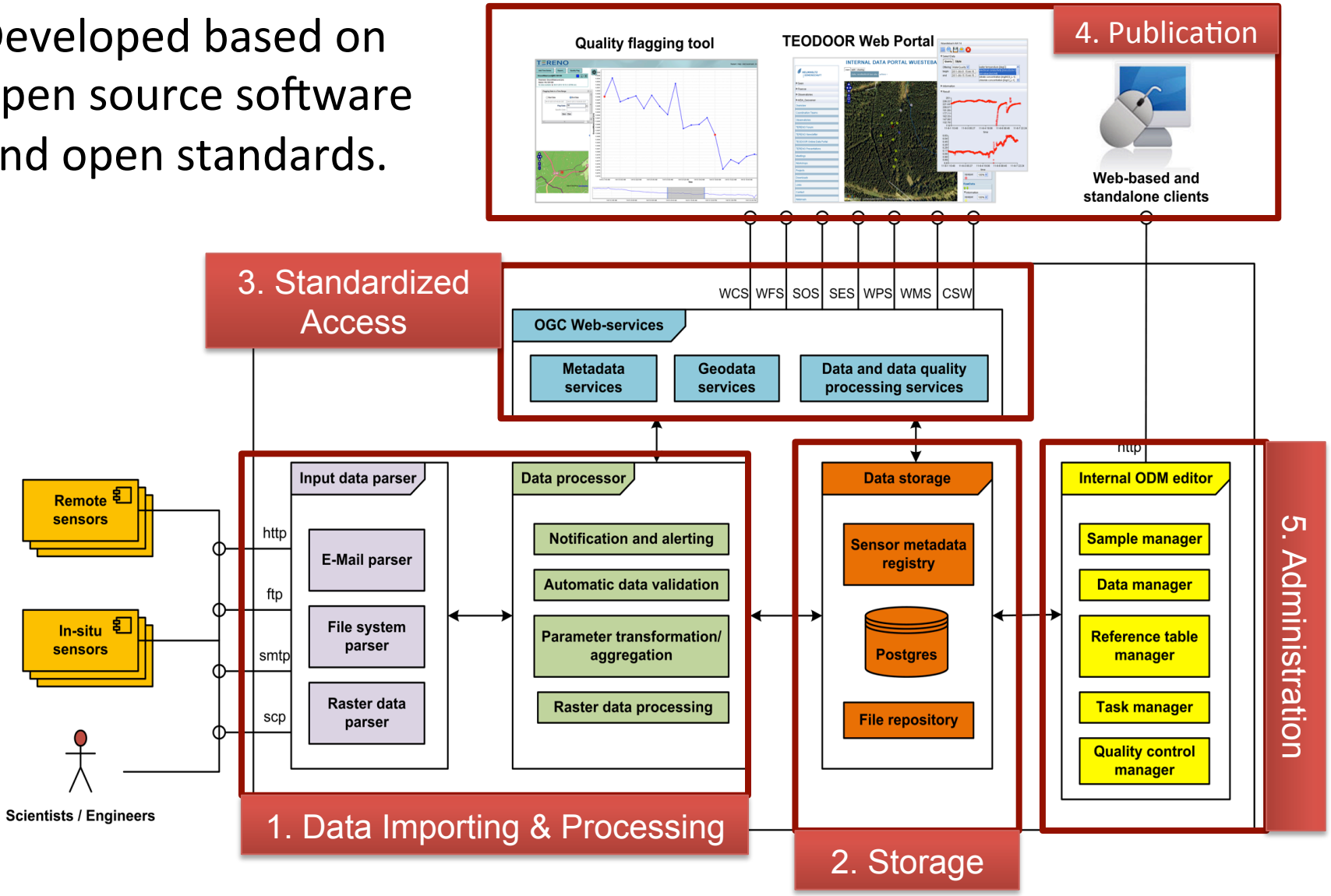
1. • Introduction to TERENO Data Infrastructure
2. • Motivation & Challenges
3. • Quality Assessment (or Quality Control) Framework
3. • Conclusions

Why Data Quality Assessment Framework?

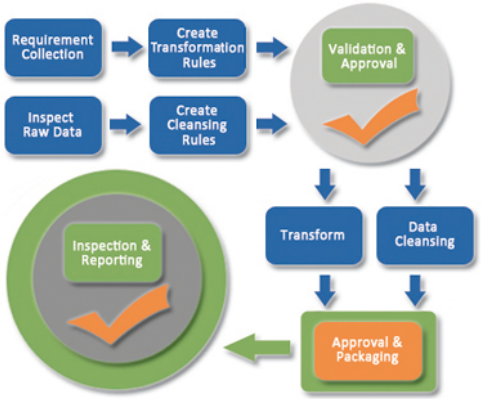
1. Data Consumer
   - Select datasets best suited their applications
   - Avoid potential errors
2. Data Provider
   - Eases data maintenance
   - Enhances research visibility

TERENO Data Infrastructure

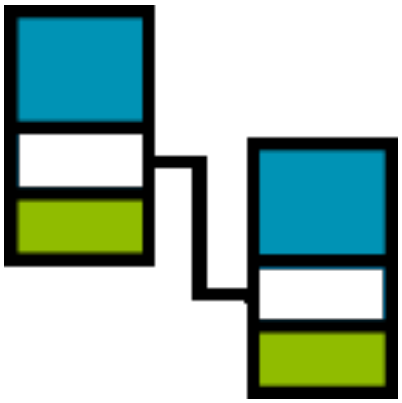
Developed based on open source software and open standards.



TERENO Data Assessment Framework



1. Data Workflows



2. Data Model

3. Data Service

Data Assessment Framework

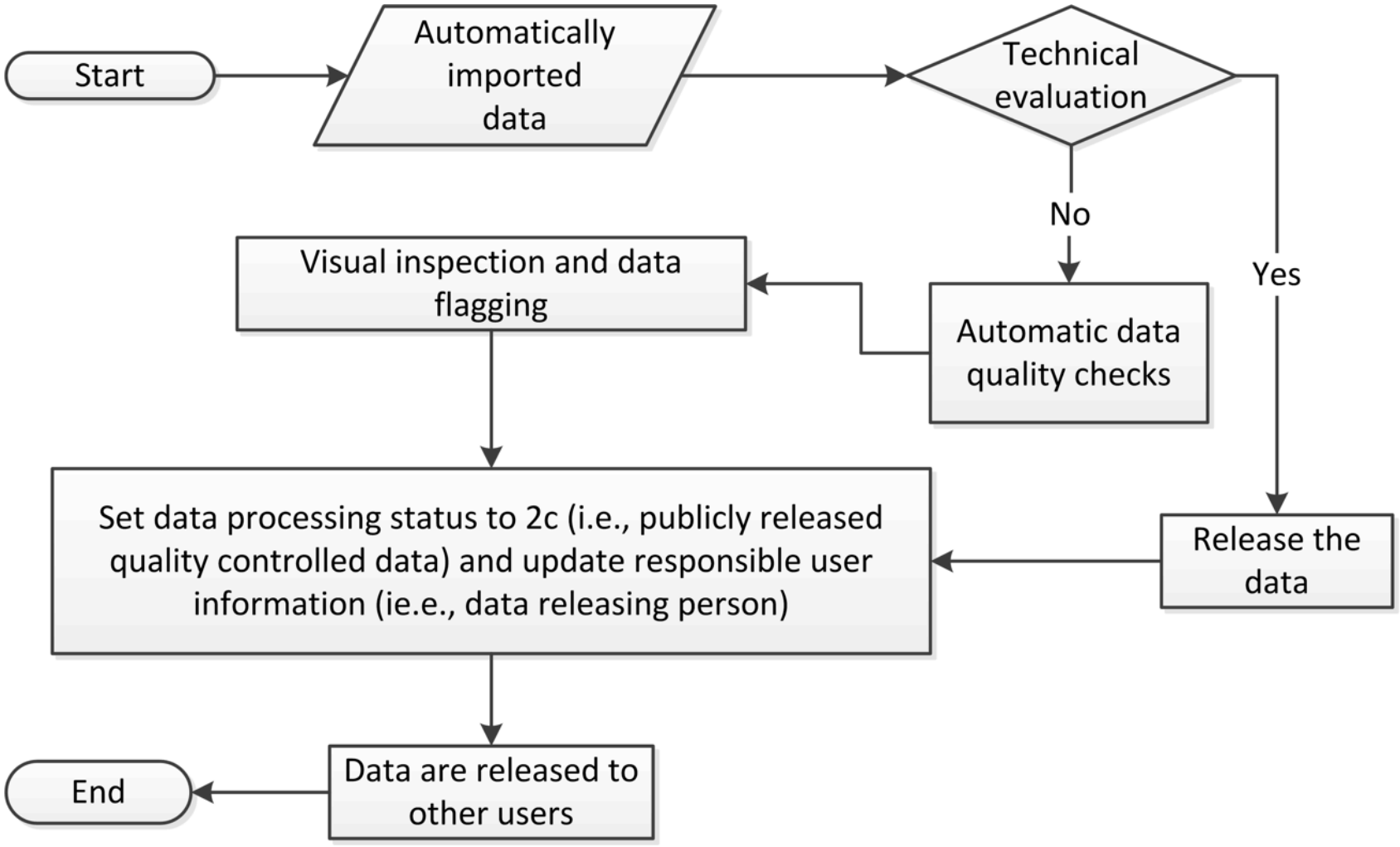
4. Client Applications

1. Data Workflows Development

1. Data series are **quality assessed (aka. quality controlled) externally** via proprietary tools and then imported into the data infrastructure

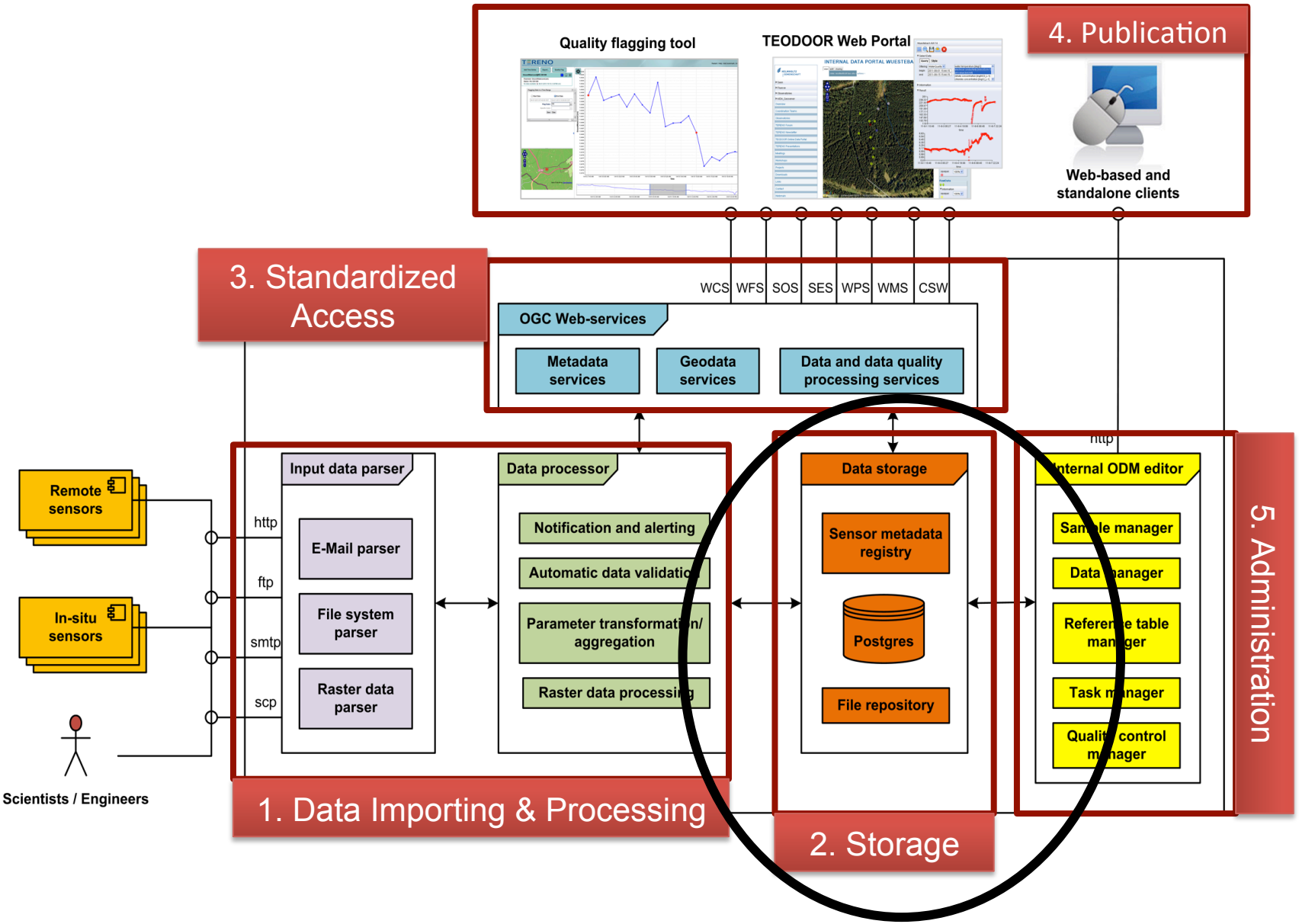
2. Data series are imported automatically from sensors and then **quality assessed internally** (within the TERENO data infrastructure).

1. Data Workflows Development



Example of the data workflow of automatically imported data

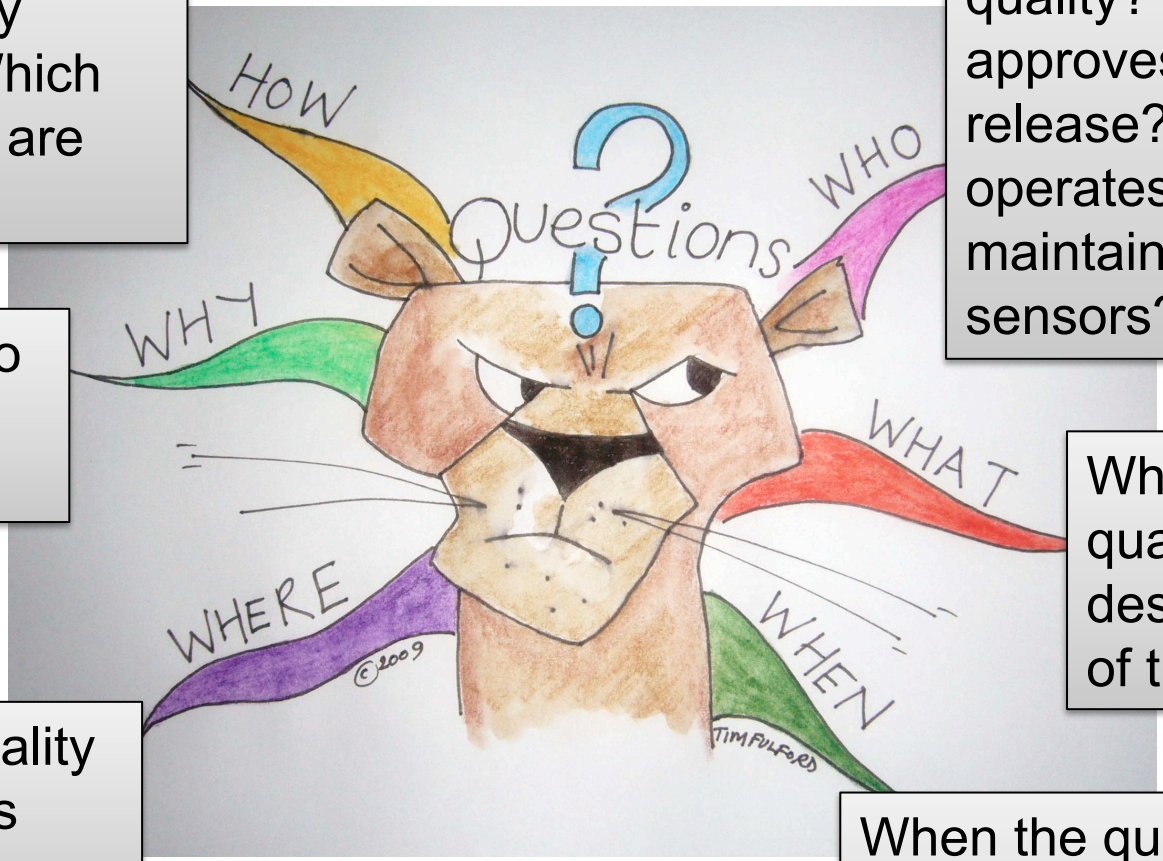
2. Data Model and Implementation



2. Data Model and Implementation

How are data series quality checked? Which quality tests are applied?

Who checks the data quality? Who approves the data release? Who operates and maintains the sensors?



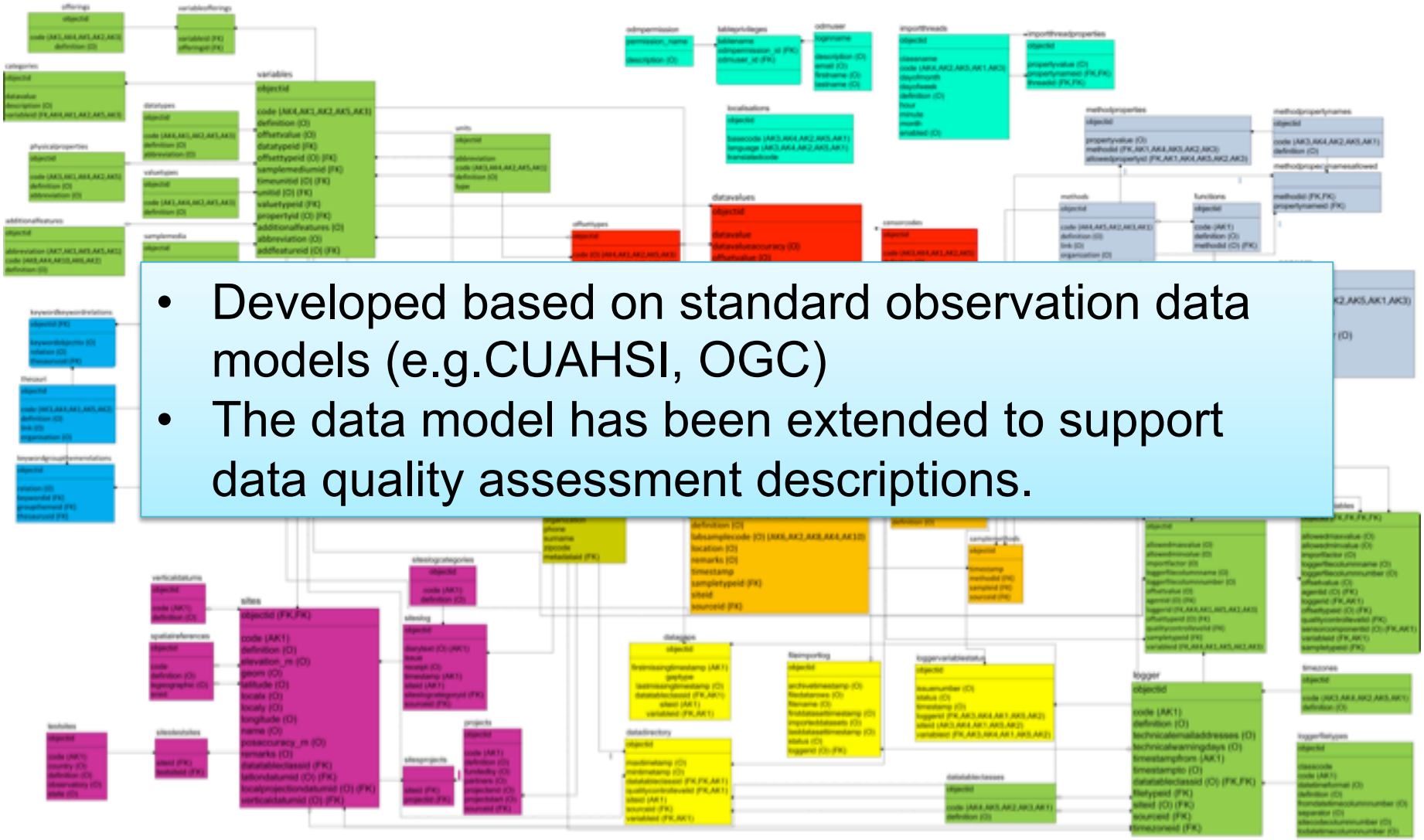
What leads to problems within data?

What are the quality descriptions of the data?

Where the quality assessment is performed?

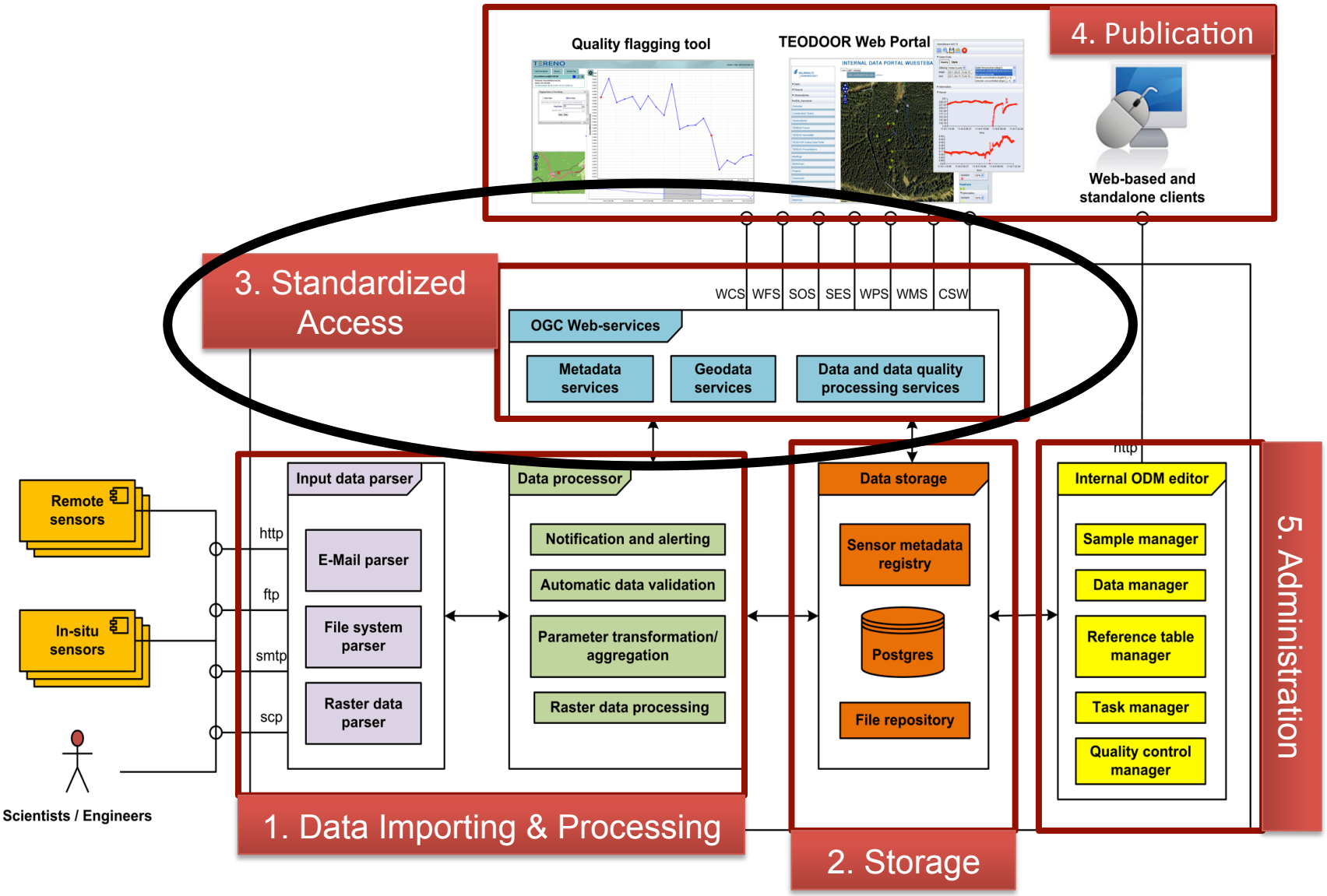
When the quality control procedure is performed? When data is released?

2. Data Model and Implementation

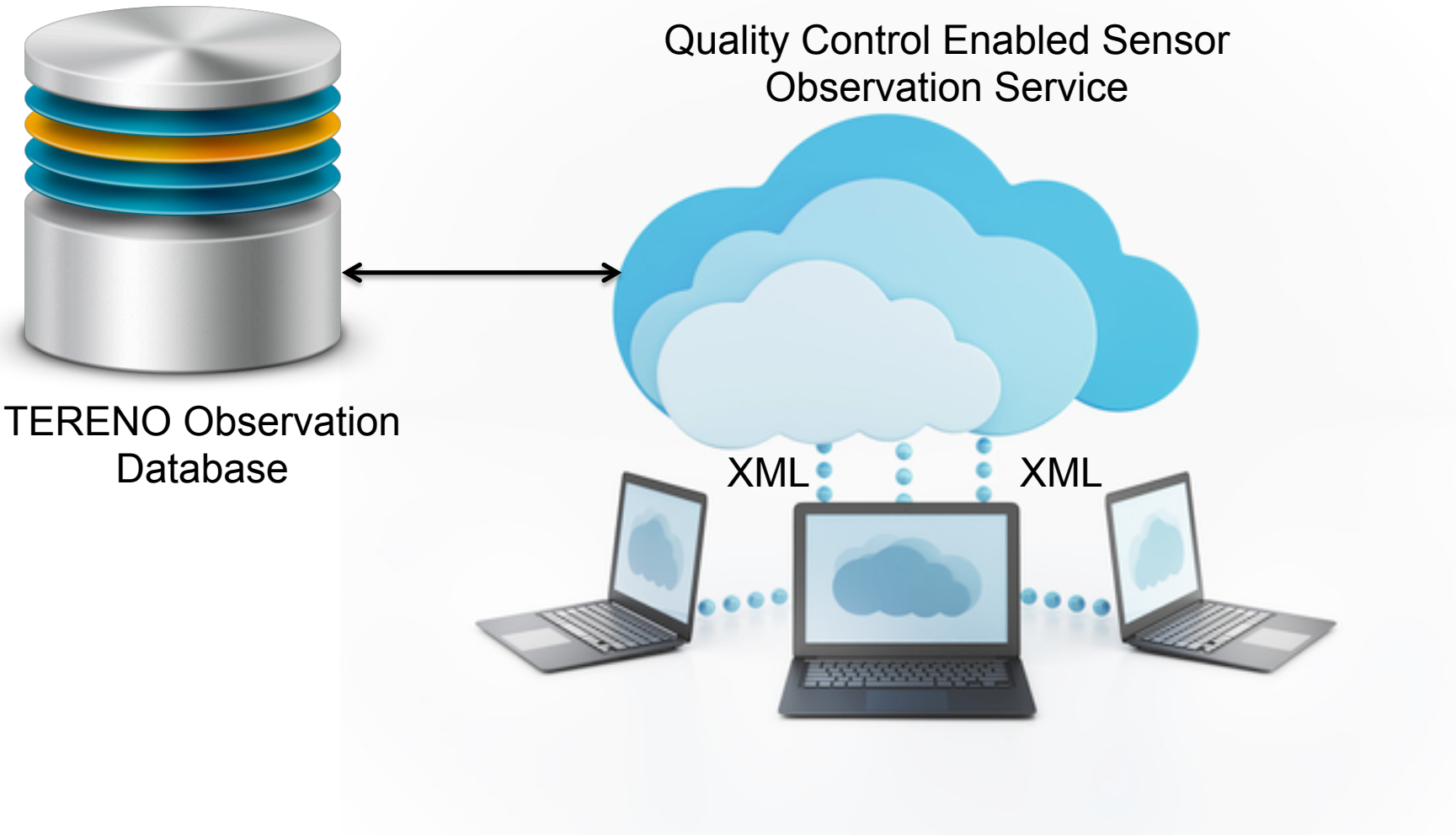


- Developed based on standard observation data models (e.g.CUAHSI, OGC)
- The data model has been extended to support data quality assessment descriptions.

3. Observation Data Service



3. Observation Data Service



3. Observation Data Service

```
<?xml version="1.0" encoding="UTF-8"?>
<GetObservation xmlns="http://www.opengis.net/sos/1.0"
  xmlns:ows="http://www.opengis.net/ows/1.1"
  xmlns:gml="http://www.opengis.net/gml"
  xmlns:ogc="http://www.opengis.net/ogc"
  xmlns:om="http://www.opengis.net/om/1.0"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.opengis.net/sos/1.0
  http://schemas.opengis.net/sos/1.0.0/sosGetObservation.xsd"
  service="SOS" version="1.0.0" srsName="urn:ogc:def:crs:EPSG:4326">

  <offering>Quality</offering>
  <eventTime>
    <ogc:TM_During>
      <ogc:PropertyName>om:samplingTime</ogc:PropertyName>
      <gml:TimePeriod>
        <gml:beginPosition>2008-03-01T17:44:15+00:00</gml:beginPosition>
        <gml:endPosition>2008-05-01T17:44:15+00:00</gml:endPosition>
      </gml:TimePeriod>
    </ogc:TM_During>
  </eventTime>
    <procedure>WU_GW_001</procedure>
  <observedProperty>GroundWaterLevel</observedProperty>
  <responseFormat>text/xml;subtype=&quot;om/1.0.0&quot;</responseFormat>

</GetObservation>
```

OGC's Sensor Observation Service (Data Request)

Quality Flags

```
<om:ObservationCollection xmlns:om="http://www.opengis.net/om/1.0" xmlns:gml
  <gml:metaDataProperty>
    <swe:DataArray>
    ...
      <swe:elementType name="Components">
        <swe:SimpleDataRecord>
          <swe:field name="Id"/>
          <swe:field name="GenericQualifier"/>
          <swe:field name="SpecificQualifier"/>
        </swe:SimpleDataRecord>
      </swe:elementType>
      <swe:encoding>
        <swe:TextBlock decimalSeparator="." tokenSeparator="," blockSeparato
      </swe:encoding>
      <swe:values>13,baddata,autosampler;22,baddata,badquality;24,baddata,fr
    </swe:DataArray>
  </gml:metaDataProperty>
```

Data Processing Status

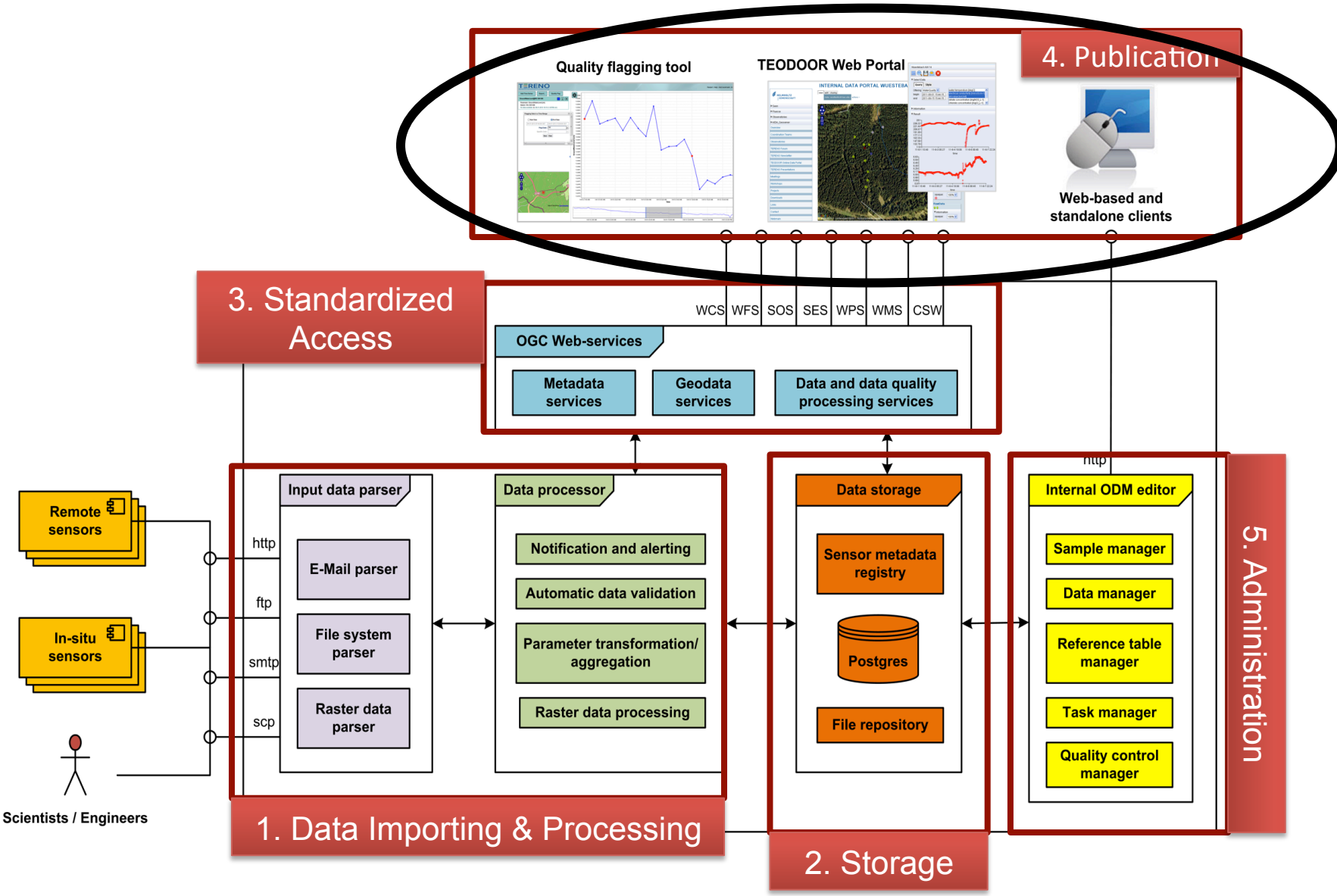
```
  <gml:metaDataProperty>
    <swe:DataArray>
    ...
      <swe:elementType name="Components">
        <swe:SimpleDataRecord>
          <swe:field name="Id"/>
          <swe:field name="ProcessingStatus"/>
        </swe:SimpleDataRecord>
      </swe:elementType>
      <swe:encoding>
        <swe:TextBlock decimalSeparator="." tokenSeparator="," blockSeparato
      </swe:encoding>
      <swe:values>1,1;2,2a;3,2b;4,2c;5,3</swe:values>
    </swe:DataArray>
  </gml:metaDataProperty>
  <om:member>
    <om:Observation gml:id="ot_749712098">
      <om:samplingTime>
    ...
      </om:samplingTime>
      <om:procedure xlink:href="WU_GW_001"/>
      <om:observedProperty>
        <swe:CompositePhenomenon gml:id="cpid0" dimension="2">
          <gml:name>resultComponents</gml:name>
          <swe:component xlink:href="urn:ogc:data:time:iso8601"/>
          <swe:component xlink:href="GroundWaterLevel"/>
          <swe:component xlink:href="GroundWaterLevelQualityFlag"/>
        </swe:CompositePhenomenon>
      </om:observedProperty>
      <om:featureOfInterest>
        <gml:FeatureCollection>
    ...
      </om:featureOfInterest>
      <om:result>
        <swe:DataArray>
          <swe:elementCount>
            <swe:Count>
              <swe:value>288</swe:value>
            </swe:Count>
          </swe:elementCount>
```

Observation Values

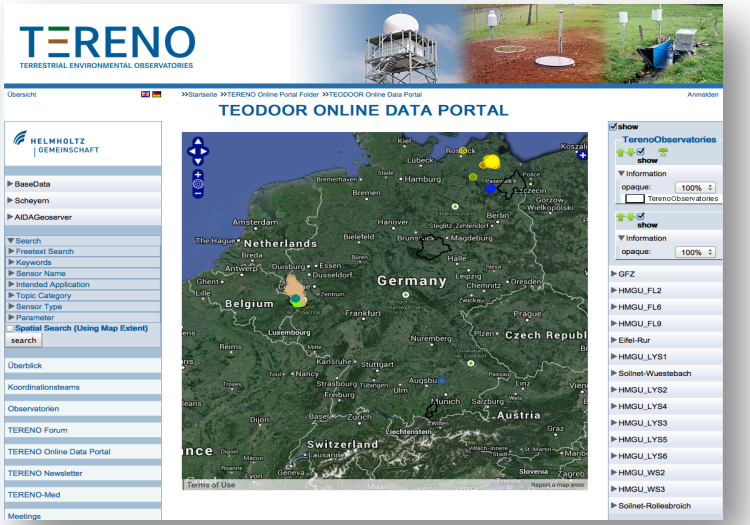
```
          <swe:elementType name="Components">
            <swe:SimpleDataRecord>
              <swe:field name="Time">
                <swe:Time definition="urn:ogc:data:time:iso8601"/>
              </swe:field>
              <swe:field name="feature">
                <swe:Text definition="urn:ogc:data:feature"/>
              </swe:field>
              <swe:field name="GroundWaterLevel">
                <swe:Quantity definition="GroundWaterLevel">
                  <swe:uom code="cm"/>
                </swe:Quantity>
              </swe:field>
              <swe:field name="GroundWaterLevelQualityFlag">
                <swe:Category definition="GroundWaterLevelQualityFlag"/>
              </swe:field>
            </swe:SimpleDataRecord>
          </swe:elementType>
          <swe:encoding>
            <swe:TextBlock decimalSeparator="." tokenSeparator="," blockSepa
          </swe:encoding>
          <swe:values>2013-03-01T18:50:00.000+01:00,WU_GW_001,-2.54618000000
          2013-03-01T19:00:00.000+01:00,WU_GW_001,-2.54541,3_1;2013-03-01T19
          WU_GW_001,-2.54542,3_1;2013-03-01T19:20:00.000+01:00,WU_GW_001,-2.
        </swe:DataArray>
      </om:result>
```

Each value is accompanied with a reference combining quality flag id and data processing status id

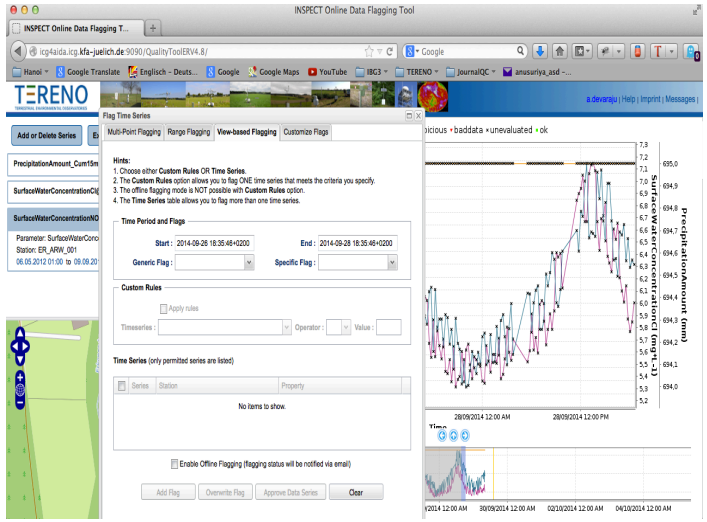
4. Client Applications



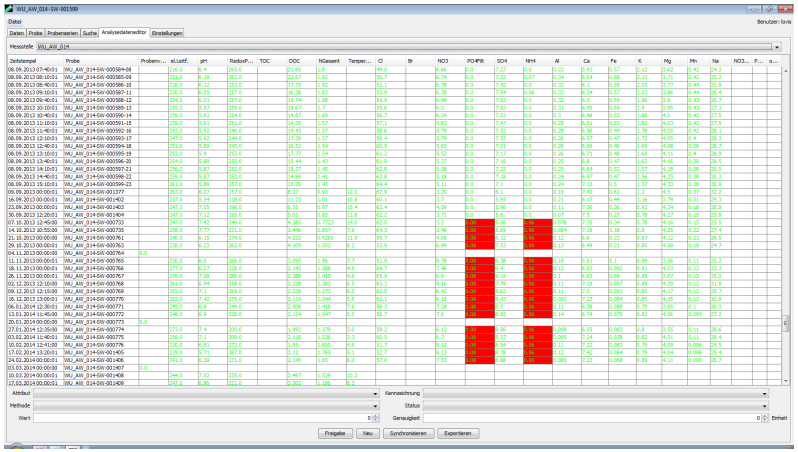
4. Client Applications



TERENO Data Portal



INSPECT Data Inspection Tool

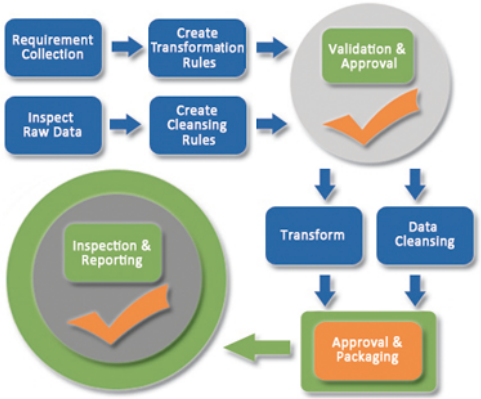


Sample Data Management Tool

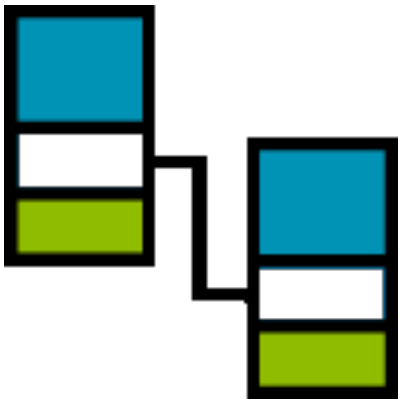
Presentation Outline

1. Introduction to TERENO Data Infrastructure
2. Motivation & Challenges
3. Quality Assessment (or Quality Control) Framework
3. Conclusions

TERENO Data Assessment Framework



1. Data Workflows



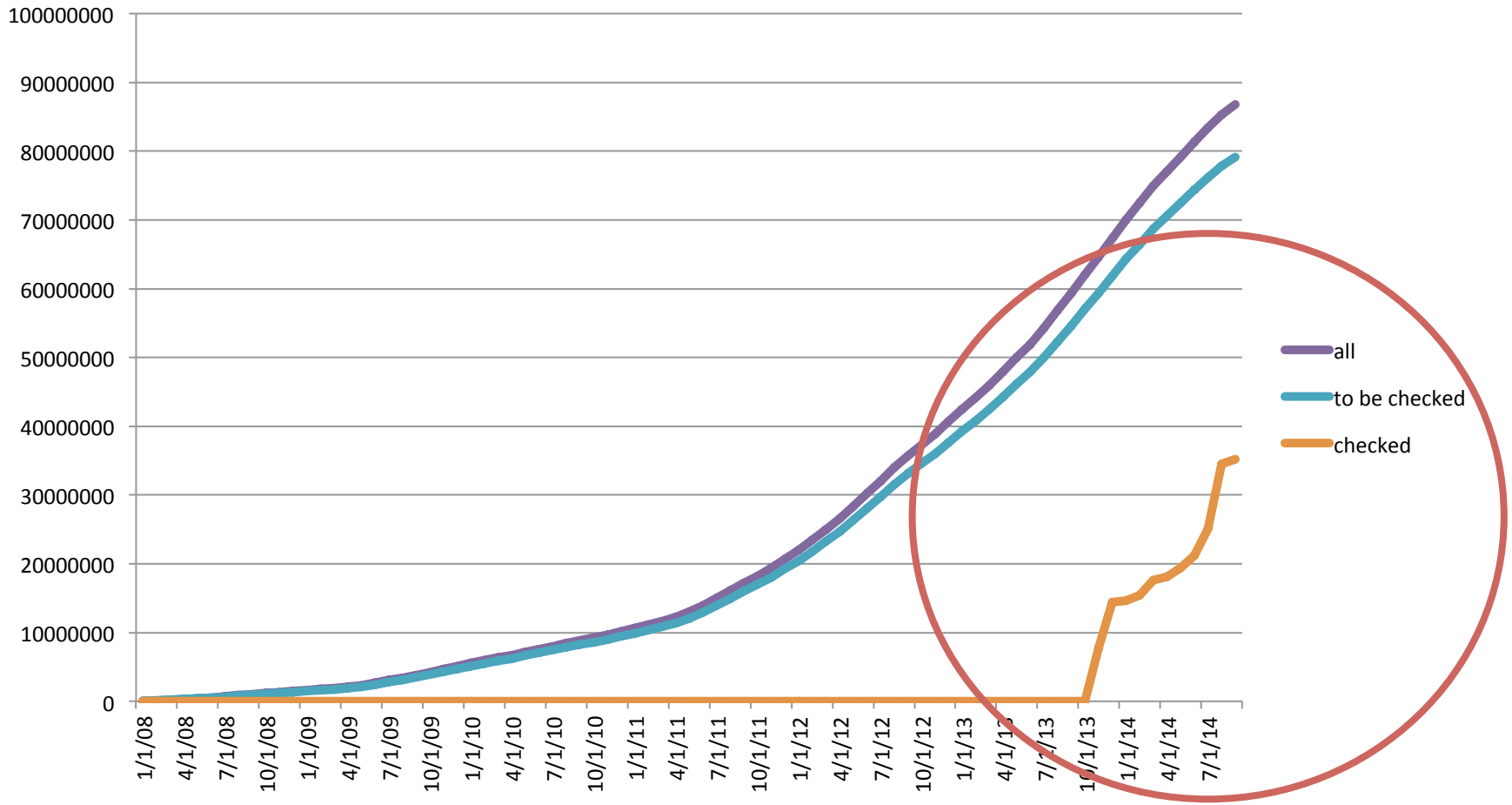
2. Data Model

3. Data Service

Data Assessment Framework

4. Client Applications

A Long Way To Go…



Quality Controlled TERENO Data
(excluding soilnet and soilcan data)

What's Next?

1. Facilitates data quality assessment by means of provenance information (data origin and derivation), e.g., offline records of sensor maintenance and operation
2. User feedback on data usability


THANK YOU
FOR
YOUR
ATTENTION!
ANY QUESTIONS?