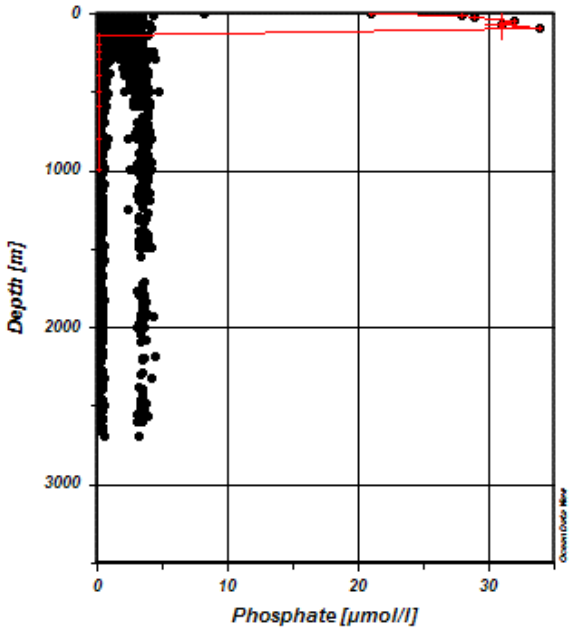
ODV Development during SeaDataNet2

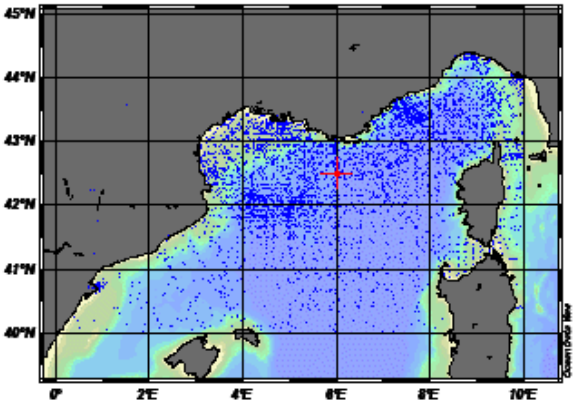
Reiner Schlitzer, Alfred Wegener Institute, Bremerhaven, Germany

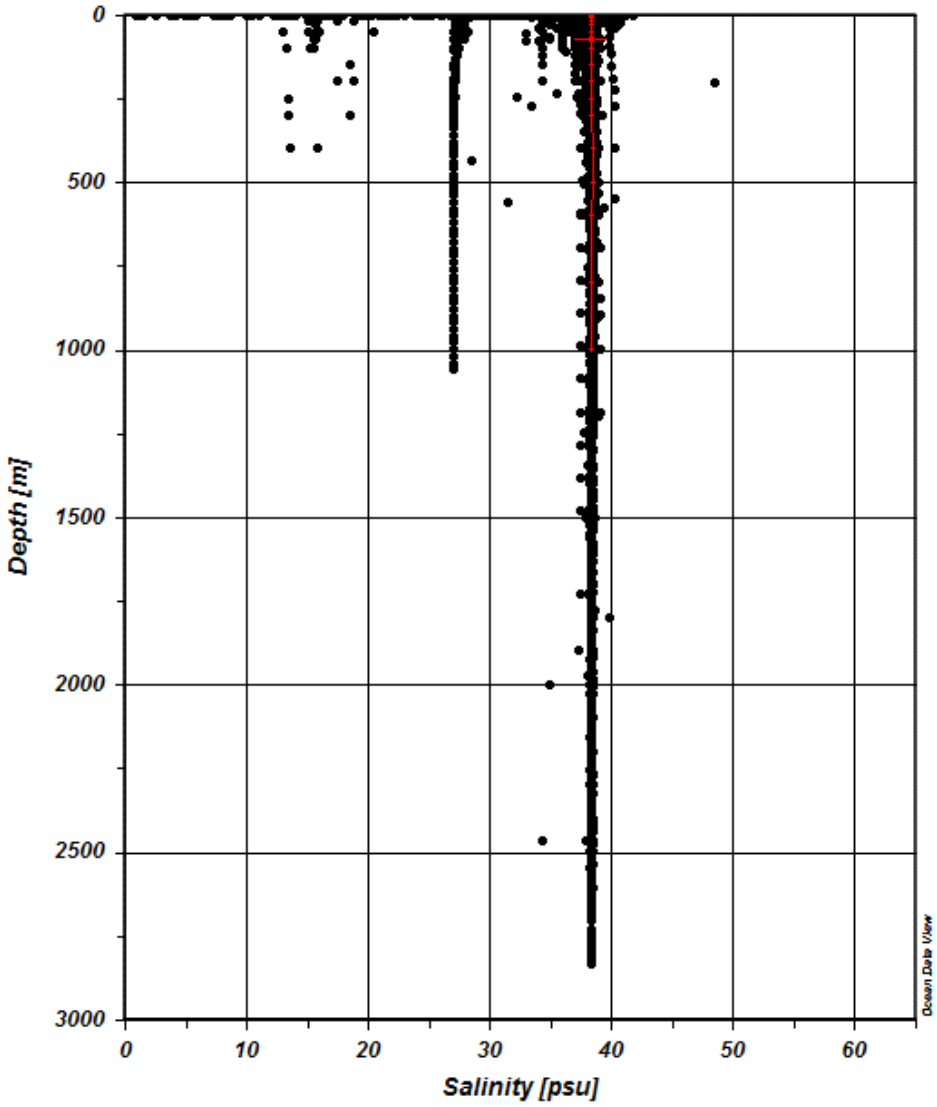
1. **Enhancing and facilitating QC/QA procedures.**
2. **Enhancing SDN data file aggregation and parameter merging.**
3. **Developing tools for duplicate station detection and processing.**
4. **Facilitating access to local or remote netCDF datasets.**

Wish: All data are of highest quality.

Reality:







- **Need to quality control data.**
- **Flag dubious data. Don‘t delete data!**
- **ODV maintains a quality flag value for every data value.**
- **Has many ways of assigning quality flags.**

1. Enhancing and facilitating QC/QA procedures

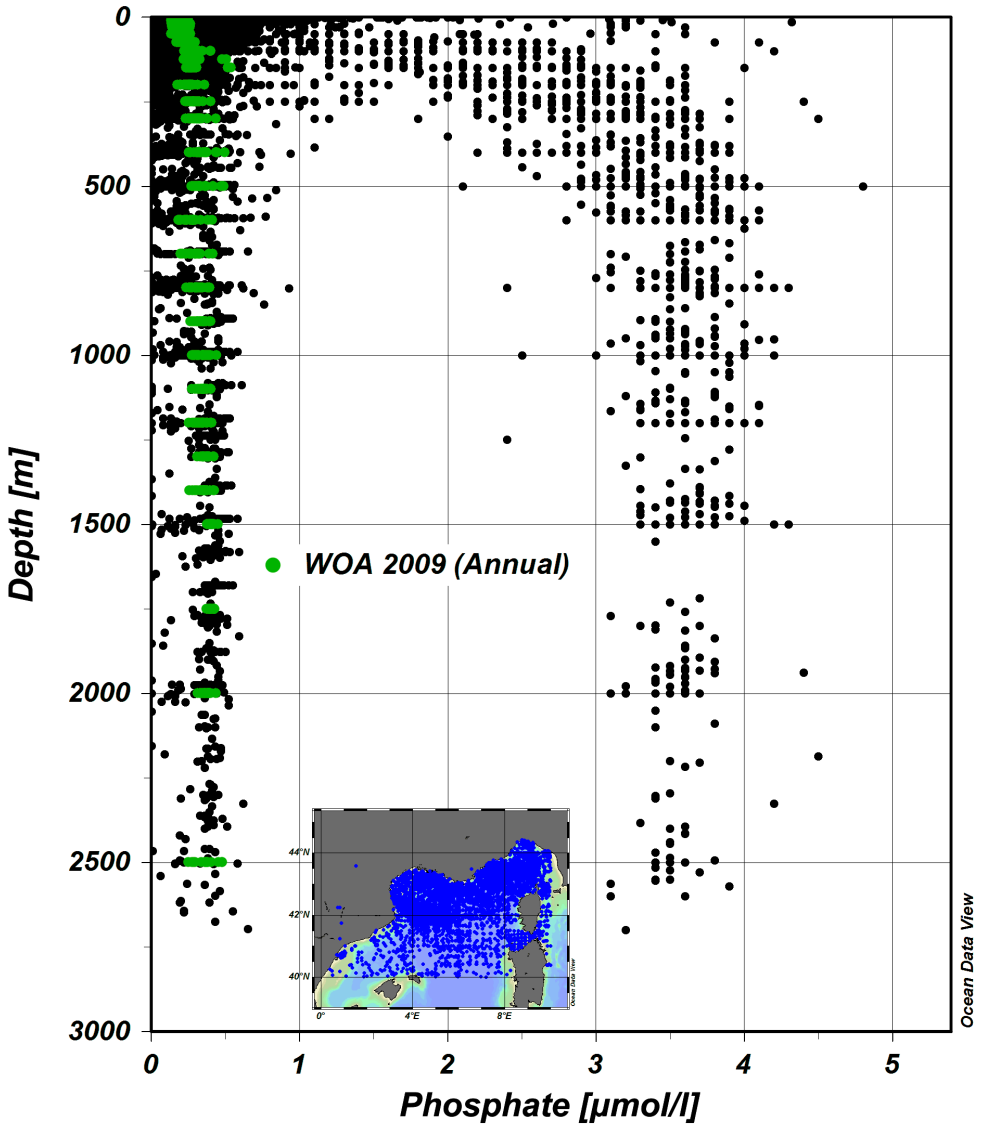
- Implement standard test suites, such as US NODC World Ocean Database and Atlas tests, EMODNET recommended tests, etc.

- Provide plugin interface supporting of user-defined specific quality tests.

- Attach reference datasets such as World Ocean Atlas, MedAtlas II, HELCOM, etc. for visual or algorithmic comparisons.

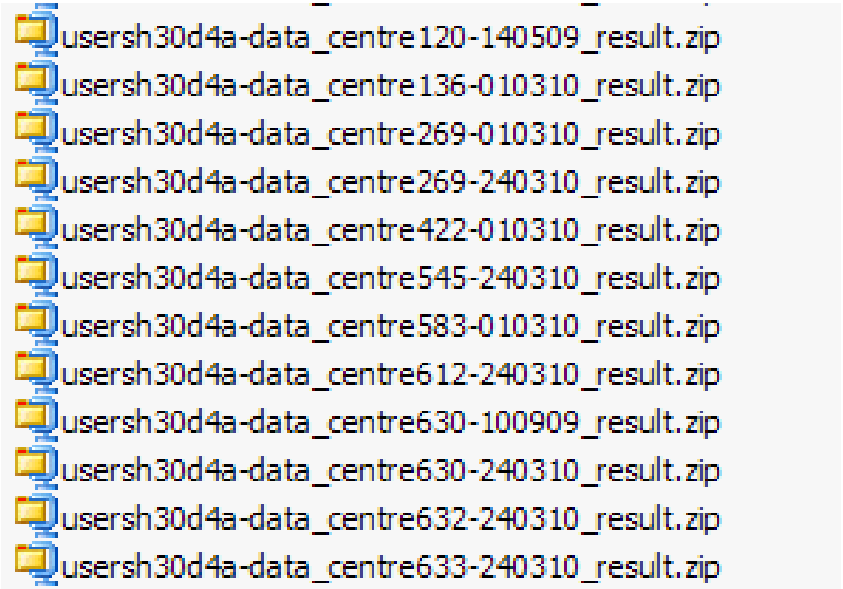
- Participate in training courses for SDN2 partners.

Using reference data for QC/QA:



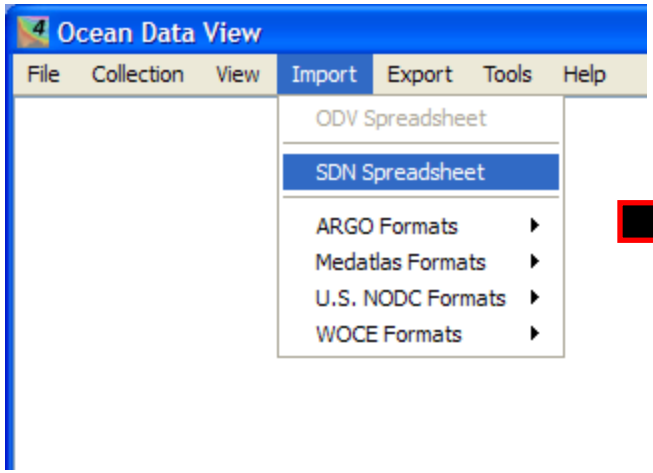
SDN Data Aggregation…

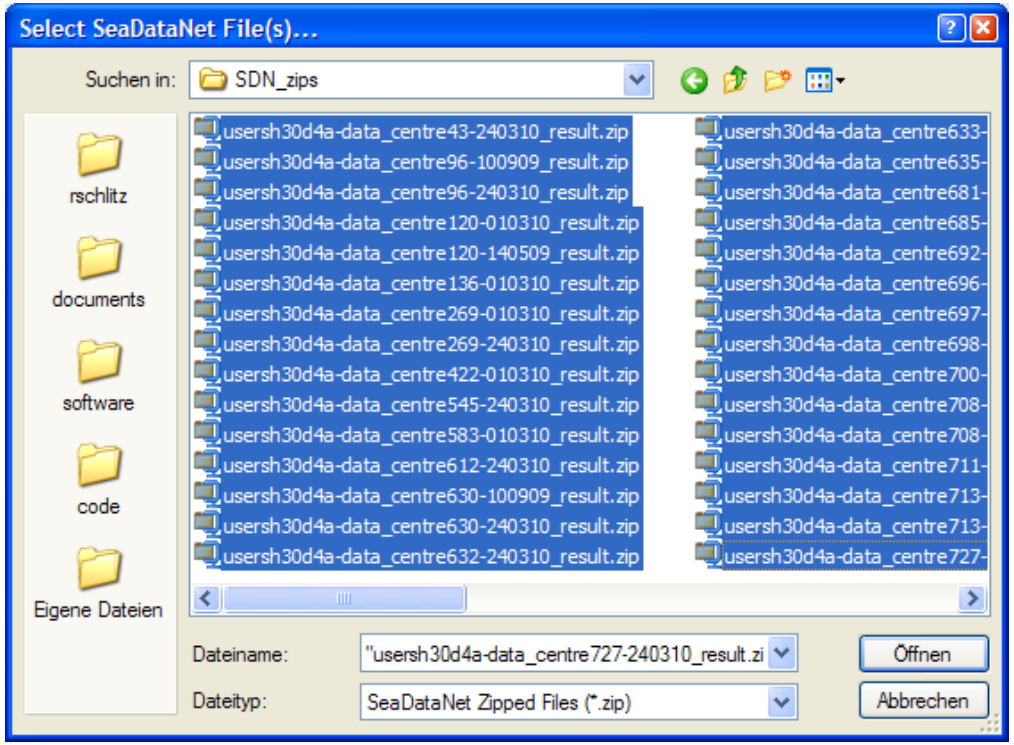
- ***data retrievals yield large number of zip files***
- ***each zip file may contain a potentially large number of individual data files***
- ***the type of data and set of variables may differ from file to file***



Challenge: ***While all individual data files are in ODV spreadsheet format the large number of files and their diversity make normal (single file) ODV spreadsheet import impractical.***

Solution: *SDN Spreadsheet importer consisting of file analysis and aggregation.*





File Analysis…

- For every file determine the type of data (profile, time-series, trajectory).
- Uses the //SDN_parameter_mapping section

```
//SDN_parameter_mapping
//<subject>SDN:LOCAL:PRESSURE</subject><object>SDN:P011::PRESPS01</object><units>SDN:P061::UPDB</units>
//<subject>SDN:LOCAL:T90</subject><object>SDN:P011::TEMPS901</object><units>SDN:P061::UPAA</units>
//<subject>SDN:LOCAL:Salinity</subject><object>SDN:P011::PSALPR02</object><units>SDN:P061::UUUU</units>
//<subject>SDN:LOCAL:fluorescence</subject><object>SDN:P011::CPHLPM01</object><units>SDN:P061::UGPL</units>
//<subject>SDN:LOCAL:Trans_red_25cm</subject><object>SDN:P011::POPTDR01</object><units>SDN:P061::UPCT</units
//<subject>SDN:LOCAL:POTM</subject><object>SDN:P011::POTMCV01</object><units>SDN:P061::UPAA</units>
//<subject>SDN:LOCAL:Density</subject><object>SDN:P011::SIGTPR01</object><units>SDN:P061::UKMC</units>
```

Aggregation…

- **ODV collections are created for every data type found.**
- **Data of a given file are imported into the corresponding ODV collection.**
- **Process is fully automatic.**
- **Results in one or more ODV collections containing all data of a given type.**

2. Enhancing SDN retrieval data file aggregation

- **Consolidate the file analysis and parameter identification (robustness and tolerance).**

- **Identify and process duplicate stations.**

- **Offer built-in as well as user-defined conversions to allow a change of variables (e.g., pressure vs. depth) and the merging of variables using different units (e.g., O_2 ml/l, µmol/l, µmol/kg).**

Merging of Parameters for Scientific Usage

```
Oxygen [ml/l] =
        1 *       (DOX1 DISSOLVED OXYGEN ml/l)                    + 0
        0.02297 * (DOX2 DISSOLVED OXYGEN micromole/kg)            + 0
        0.02241 * (DOXY DISSOLVED OXYGEN millimole/m3)            + 0
```

Merging Principles:

- Parameter priority
- Conversion formulas
- Merging Type (exclusive use of single parameter, average over all available parameter values)
- Quality flag inheritence

3. Developing tools for duplicate station detection and processing

Problem: ***Many datasets are held by more than one data center. SDN data retrieval will return the same data more than once: Duplicates.***

Action: ***Duplicates may bias scientific products and therefore must be identified and removed.***

Development:

- ***Increase flexibility and efficiency of duplicates search***
- ***Implement procedures for duplicates removal (delete all but one; merge data of duplicates into single, new station; others?)***

4. Strengthening support for netCDF datasets

Increasing interoperability with remote sensing and modeling communities by…

- **Supporting netCDF4 and HDF formats**
- **Allowing remote data access via OpenDAP**
- **Providing netCDF export from ODV collections**