Open Data Standards v2.0.0 - Summary

1.0 Proposed Updates for ODS v2.0.0

This document outlines the proposed updates for ODS v 2.0.0. All changes will be released onto the 'master' branch of the ODS repo (https://github.com/OasisLMF/OpenDataStandards).

2.0 'Major' Updates

These updates could cause 'breaking' changes for current users; however, the impact is considered low (see Governance doc and GitHub repo for more details on versioning semantics). The likely issues would occur within OED validation processes which will need to be updated to incorporate the changes to the data headers in each input file.

2.1. Updates to the OED 'Loc' file

- Added *'CondTag'* varchar(20) field:
- Removed *'CondNumber'* int field
- Added *'OEDVersion'* varchar(10) field
- Change *'NumberOfEmployees'* int field to *'NumberOfOccupants'* int field
- Add *'OccupantPeriod'* tiny int field
- Addition of an aggregate field *'IsAggregate'* to flag whether the data is aggregated.
- Removing the limit of five *'GeogScheme'/'GeogName'* pairs

2.2 Updates to the OED 'Acc' file

- Added '*CondTag'* varchar(20) field
- Changed data type of '*CondNumber'* from int to varchar(20)
- Added '*OEDVersion'* varchar(10) field

2.3 Updates to the OED 'ReinsInfo' file

- Added '*RiskLevel'* char(3) field (from ReinsScope file)
- Added '*OEDVersion'* varchar(10) field

2.4 Updates to the OED 'ReinsScope' file

- Removed '*RiskLevel'* char(3) field (moved to ReinsInfo file)
- Added '*OEDVersion'* varchar(10) field

2.5 Rationale for 'Major' Updates

Its key to keep the ODS community updated with changes to ODS and their rationale and benefits.

2.5.1 Introduce a new field *'CondTag'*

The rationale behind adding a 'condition tag' and 'condition number' to the OED input files is to ease the data entry for multiple special conditions and to make the capture of that data less ambiguous and decrease the risk of misinterpretation.

Special condition terms can vary by policy, even though they apply to the same subset of locations. To support this, the user must enter a duplicate entry in the location file relating to each *'CondNumber'* that applies to that location. This enables the entry of a whole range of potentially unsupported conditions (any location could belong to unlimited '*CondNumbers'*, in any combination and at any priority) which is difficult to validate.

In recognition that the grouping of locations to which each special condition applies tends to be constant, the proposal is to (re)introduce the '*CondTag'* field in the location file **and** the account file so that locations subject to a condition can be tagged once, and the *'CondNumber'* (relating to the specific terms) associated to the '*CondTag'* in the account file. This prevents unnecessary duplication in the location file and makes it easier to validate files and flag (or fail) unsupported conditions.
Note that hierarchal conditions (genuine multiple conditions on the same location with different priorities) could still be supported by allowing one location file entry per condition priority.

2.5.2 Move *'RiskLevel'* from 'ReinsScope' to 'ReinsInfo' file

The reason for adding *'RiskLevel'* to the reinsurance files is to restrict the reinsurance contract to a single definition of risk. This would simplify the entry of reinsurance risk level by reducing the possibility of entering unsupported terms which would fail or cause incorrect results.

For some types of reinsurance, there is a definition of the 'risk level' to which risk level terms apply. This may be location, account, policy or location group.
Currently the 'risk level' is defined in the '**ReinsScope'** file so that each individual risk under the scope of the treaty may be defined as a different 'risk level'. This makes it possible to define a treaty that applies risk level terms to any mixture of locations, policies and accounts

Moving the '*RiskLevel* 'to the **'ReinsInfo'** file would mean that only one risk level can be defined for each reinsurance contract. This prevents users entering mixed risk level treaties.
The rationale is that it is very rare or unusual to have mixed risk level treaties. Where there is complexity or ambiguity in the definition of risk in the contract, in OED there is the option to use the *Location Group* RiskLevel, where any set of locations can be tagged together (by the '*LocGroup'* field) and defined as a risk.

2.5.3 Changing *'NumberOfEmployees'* to *'NumberOfOccupants'* in the 'Loc' file

This makes it less ambiguous when capturing the occupants of a building, regardless of whether it's a commercial building, place of work or a residential property. This also supports the capture of additional (demographic) information of the occupants for humanitarian use cases (see 3.9).

2.5.4 Adding '*OccupantPeriod'* to the 'Loc' file

In conjunction with 2.5.3, including a field for '*OccupantPeriod'* will help define whether the occupants are daytime employees or night-time residents in the same location file.

2.5.5 Including a version number in all input files

This is beneficial when differentiating between multiple versions of ODS following the release of a new version.

2.5.6 Adding an *'IsAggregate'* field to the 'Loc' file

As some of the less detailed models (that use the OED format) use aggregate data, its beneficial to introduce a permanent aggregate flag field rather than using the 'flexi-fields'. It would be advantageous when disaggregating data.

'IsAggregate' would be used in conjunction with the *'NumberOfBuildings'* field as follows:

- *IsAggregate* = 1, *NumberOfBuildings* > 1 means aggregated data where number of locations is known
- *IsAggregate* = 1, *NumberOfBuildings* = 0 or blank, means aggregated data where number of locations is unknown
- *IsAggregate* = 0, *NumberOfBuildings* >1 means a campus

2.5.7. Removing the limit of five *'GeogScheme'* and *'GeogName'* pairs

Some users have stated that the current limit of five geography schemes (i.e. postcode sector, postcode unit, CRESTA high res, CRESTA low res, etc.) is not enough for their requirements, so OED can now support up to a possible ninety-nine pairs. It's worth noting that using multiple geogschemes could have a detrimental impact on performance during the 'lookup' phase.

The data field header will be changed to *GeogSchemeXX* where the 'XX' will need to be changed to an integer (i.e. *GeogScheme1* > *GeogScheme20* etc).

3.0 Minor Updates

The following updates are considered 'minor' and will have an immaterial impact on current users. The updates to scope of ORD will not affect any users at this stage as ORD has not been fully implemented into ODS or Oasis at the time of publishing this document.

3.1 Updates to the Open Results Data (ORD) Outputs

As ORD is still in the process of being implemented into Oasis, the updates proposed below are changes to the ORD schema developed in the original scope.

3.2 Report name changed from Period Average Loss Table (PALT) to Average Loss Table (ALT)

3.3 Reindexing of *EPCalc* and *EPType* are as follows:

EPCalc:
1 = Mean Damage Ratio
2 = Full Uncertainty
3 = Per Sample Mean
4 = Sample Mean
EPType:
1 = OEP
2 = OEP TVaR
3 = AEP
4 = AEP TVaR

3.4 Removing Stand-Alone Financial Perspectives in ORD

The proposed changes are to remove the following RI perspectives as standalone:

- Reinsurance FAC

- Reinsurance Quota Share
- Reinsurance Surplus Share
- Reinsurance PR
- Reinsurance CXL

These perspectives are all covered by the 'Reinsurance Loss' perspective when broken down by 'reinsurance type'. Each type of reinsurance can be represented by a different *'SummaryID'*

3.5 Additional RI Perspectives in ORD

A new perspective of *'Net Loss Post Cat;'* has been added for those platforms that support 'Occurrence' and 'Aggregate Cat XL', and 'Stop Loss' reinsurance.
A 'user defined net' perspectives has been added for platforms that support use defined reinsurance net perspectives. Further information about these perspectives has been added to the data specification.

3.6 Removed *'version'* from ORD spec filename

Now version number is included in all OED input files (see 2.5.5), having it in the filename is unnecessary.

3.7 Introducing *'What3Words'* to the Geogscheme

What3Words ("W3W") will be included as a *Geogscheme* in OED to support the use of the global grid address system represented by three random words (https://what3words.com).

There will be no access to *What3Words* data in ODS, this will need to be licensed directed from them.

3.8 Inclusion of the following additional perils codes in OED:

New peril codes:
s = subsidence
g = agriculture
p = pandemic

New group perils:
'PP1' = All Pandemic perils
'GG1' = All Crop perils
'CC1' = All Cyber perils
New perils:
'SSD' = Drought induced subsidence
'XCH' = Crop Hail (from Convective Storm)
'CSB' = Cyber security data and privacy breach

'CPD' = Cyber security property damage
'PNF' = Pandemic Flu

3.9 Inclusion of an additional 'LocPopulation' input file

This will be a separate OED input file to capture detailed information for population such as age ranges, sex, ethnicity, and refugee status. These data fields have been proposed by the *Insurance Development Forum Risk Modelling Steering Group (IDF RMSG)* who considers them key for humanitarian use cases of this data.
This input file will be located on the ODS GitHub repository alongside all other OED files.

4.0. New Conversion Tools

Converting data to and from ODS formats is key for interoperability and lowering barriers for adoption. ODS v2.0.0 includes two new python conversion scripts.

4.1. OED to Parquet Conversion

This is a Python conversion script that converts all data in csv format in any OED input file to parquet format. The conversion of ORD will follow in a later release.
Parquet is an open-source, columnar data storage format that has superior compression capabilities which will significantly improve efficiency over using csv. It's worth noting that Parquet is not human-readable and so a UI is required to query the data directly. Plans to develop a UI are under review. More information about Parquet can be found here (https://databricks.com/glossary/what-is-parquet)

4.2 OED to PERILS Conversion

Oasis have been working with PERILS AG (https://www.perils.org) to create a conversion tool for their aggregate industry exposure data to be available in OED format. There is now a python conversion script located in the ODS repo that will convert PERILS formatted data into OED-ready input files. This will allow users to perform industry-level exposure analytics using the OED data format.
No PERILS data will be accessible through ODS and will still need to be licensed directly from PERILS AG.

5.0 Notice and User Acceptance Testing (UAT)

The updates outlined in this document were presented to the ODS steering committee (SC) on 19th May 2021 with general agreement. They were discussed

again at the SC on 20th September 2021 as an update and, subject to SC signoff, will be implemented into Oasis LMF by the end of October 2021. There were no further comments on these changes from the SC, however, it was suggested that the SC members perform their own internal UAT of the changes (if necessary) to ensure they cause no issues to their internal processes. The SC agreed that sufficient time has been given to do any UAT before the release on **1st November 2021.**

As the release of ODS v2.0.0 is not dependant on the updates being implemented into Oasis LMF, the release, subject to SC signoff, will go ahead on 1st November 2021 regardless.

6.0 Signoff

This document will be circulated via email to the SC for written signoff. A quorum for signoff will be achieved once confirmations are received from **a majority of SC members.**