Pivot Tables

Background:

What we see in Excel as a pivot table is actually the Pivot Table View or SXVIEW record. Usually there is one SXSTREAM associated with one PivotCache Storage & Stream, and one or more SXVIEW set of records that is based off of the data identified in the stream/cache set of structures.

I. Pivot tables contain 3 main components:

1- a SXSTREAM set of records, in the workbook global substream, directly after Style records.

- the sid species the stream in the pivot cache storage [see below]
  "The stream specified is the one that has its name equal to the hexadecimal representation of this field. The four-digit hexadecimal string representation of this field, where each hexadecimal letter digit is a capital letter, MUST be equal to the name of a stream in the PivotCache storage."
  in other words sid=01 ==> stream 0001 (or is it sid==00 ==> 0001)?
- identifies the stream/cache/data that a pivot table/sxview is associated with
- sub-records defines the data source (range, named range ...) + some versioning/option-type records
- SxStream records are stored in WorkBook.ptstream arraylist for ease of lookup.
- see SxStream. addInitialRecords for details

2- an SXVIEW set of records, in the sheet global substream, after the DIMENSIONS record

- SxView records are stored in WorkBook.ptLookup for ease of lookup.
- immediately following the SxView record are records which define which pivot cache fields go on which axis, what type, summary functions, etc. Most of this is not completely implemented. Certainly most SxView and associated record options are not handled.

3- a Pivot Cache Storage, named `_SX_DB_CUR`, and it's children, streams named 0001, 0002 ... one for each pivot cache. Each pivot cache contains data which represents a set of records based upon a sheet range. The range is equivalent to the data range selected for a pivot table. It is parsed and each cell type (string, numeric, formula[not handled] ...) is represented via a different record type. The current value of each cell is recorded as well.

II. Important Concepts

1. The first row of the data range of a pivot cache defines the Pivot Fields- each column in the data range is a Pivot Field.
2- Pivot items are distinct (unique??) instances of a pivot field. **This is a little bit murky, because there are things like summaries and aggregate functions ...**

3- For a Pivot Field to be used in a Pivot Table, it is placed on an Axis. A Pivot Table can have one or more: ROW, COL, PAGE or DATA axes. PAGE axis governs Pivot Table filtering. DATA axes govern summary info usually
4- Each time a pivot table is put on an axis, a whole slew of records are added + the associated pivot item indexes are aggregated into an array ...
5- Pivot lines are very unclear ...

III. Important Methods

1. Any existing Pivot Cache is initialized in WBH.postLoad.initPivotCache().
2. PivotCache.java.createPivotCache creates the storage and stream for a pivot cache
3. Boundsheet.addPivotTable(reference, wbh, streamid, tablename)
4. WorkBook.addPivotCache -- adds a pivot cache storage and substreams if not present.
   TODO: How to know to reuse a pivot cache stream or to create a new one?
5. WorkBook.addPivotStream(reference, sheetname, streamid)
   Only Sheet-type data sources are supported at this time. TODO: should create pivot cache at the same time, but it's separated out due to OOXML parsing [which contains several xml files which work in concert to define pivot tables]

IV. Incomplete Functionality [Major]:

1-Finish Basic Creation -- putting fields on axes, a basic summary and some basic filters (no Formats
2-Need output to OOXML
3- Need complete Create interface in a logical and easy to use
4- Pivot Charts ...??? Have to have special handling??

V. To Deal with/To figure out:

As it stands now, a pivot table which defines a pivot cache and pivot fields can be created, but when a field is put on a ROW or COL axis, Open Errors occur. It is probably 2 things:

1- need to create an MsoDrawing record
2- pivot lines/items - have to decipher and add correct sub-records

***** NOTE -- recent changes have made even basic pivot tables fail ... it should be something minor -- do a comparison of input records and output *****

1. have to reconcile OOXML iCache # vs Storage iCache
2. pivot lines -- figure out!!!! - if change cDimCol or cDimRw have to adjust items
3. data field name for SXVIEW
4. SXPI-- idObj = object record id --> must add a MsoDrawing item for arrow?
5. SXDI-- data display functions
6. SXLI - have to figure out
   # pivot items, refers to SXVI [refers to pivot cache record]
   cSic means line is same as a prev line ...
7. SXVIEW cCol = 1 when there are no col axis fields?
8. OOXML formats ...
9. Pivot Cache/Storage: have to handle SxDtr, SxFormula, SxDB ...

10. Is TableStyles record necessary??
11. SxView.addRow/Col/Data/PageField - ensure required records are added correctly