# Occurrence Import

Occurrence Import is used to "take" the Occurrence data from a file and store them back into the database. The whole operation must be performed very carefully so that the integrity of data that are already in the database is not violated.

Occurrence Import data can be stored in XML format only, but the framework allows simple creation of new Parsers.

## *Introduction*

The Occurrence import must be capable of parsing files of virtually any size. That's why DOM4J is not a suitable choice. The SAX is used instead. SAX is based on a different principle - it uses callbacks to process the recognized XML elements and the contents between them. SAX is used because it is the only other option, it is not a wise choice to write another XML parser.

The most important difference between SAX and DOM is, that SAX does not construct any tree at all. It merely reports events that happened while parsing the file.
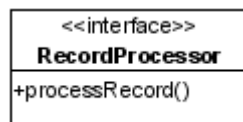
Because SAX is based on callbacks, the architecture must have been adapted to this behaviour. The input is parsed using SAX events and the record is gradually reconstructed from the file. If the Occurrence record is completely reconstructed, it is passed to the Record Processor, that will insert it back into the database.

## *The Participants*

### Record Processor

The Record Processor must be able to process the supplied Occurrence record. Processing the Occurrence data requires special measures. The Occurrence data consist of several subrecords, and some of them must be well-defined. It is not possible to process an Occurrence record whose plant, territory, nearest bigger seat or phytochorion is not in the database already. The contents of these tables cannot be altered by the Occurrence Import. The User must use the Table Import in order to make changes to one of those tables. The record processor must also always make an effort to find a match in the database for every subrecord of the Occurrence - and reuse existing subrecords in the database whenever possible.
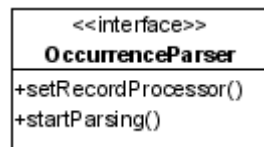
The Record Processor is in the file net.sf.plantlore.client.occurrenceimport.RecordProcessor.java.



### Occurrence Parser

The Occurrence Parser parses the input and reconstructs Occurrence records from it. Every time a record is fully reconstructed, the Record Processor must be invoked on this record.

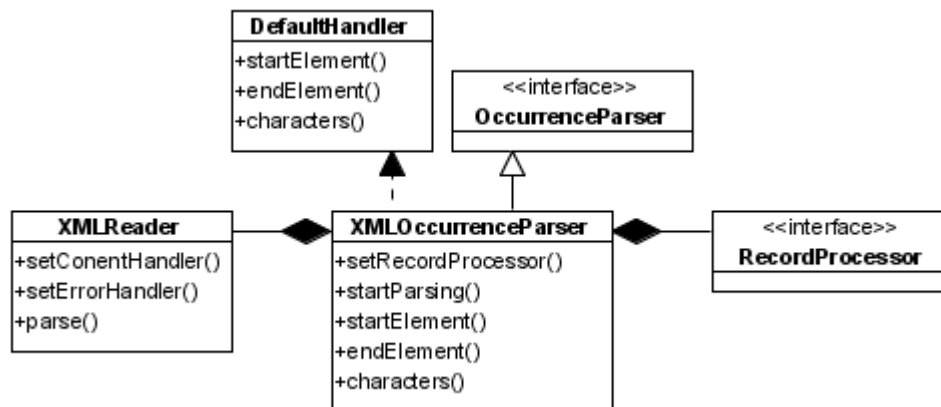The interface is in the file net.sf.plantlore.client.occurrenceimport.OccurrenceParser.java.

```
        <<interface>>
      OccurrenceParser
  +setRecordProcessor()
  +startParsing()
```

## XML Occurrence Parser

The XML Occurrence Parser utitilizes the SAX framework. In order to do it, the XML Occurrence Parser must extend the Default Handler and override its methods. These methods are those callbacks called by SAX. There the Occurrence record is gradually reconstructed.

After the record is complete, it is handed over to the Record Processor.

The XML Reader is part of the SAX interface where our implementation of the DefaultHandler must be registered and where the parsing should start.

The source is in the file net.sf.plantlore.client.occurrenceimport.parsers.XMLOccurrenceParser.java

```
         DefaultHandler                        <<interface>>
  +startElement()                           OccurrenceParser
  +endElement()
  +characters()

  XMLReader              XMLOccurrenceParser              <<interface>>
  +setConentHandler()    +setRecordProcessor()           RecordProcessor
  +setErrorHandler()     +startParsing()
  +parse()               +startElement()
                         +endElement()
                         +characters()
```
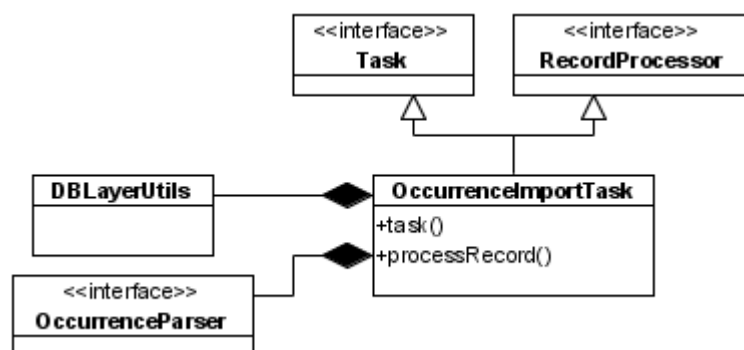
## Occurrence Import Task

The Occurrence Import Task is a Plantlore Task that must be dispatched accordingly. To learn more about tasks see the Section **Tasks and Dispatcher**.

The Occurrence Import task initiates the parsing routine. It has no other purpose. This is why in the implementation the Occurrence Import task and the Record Processor are merged together.
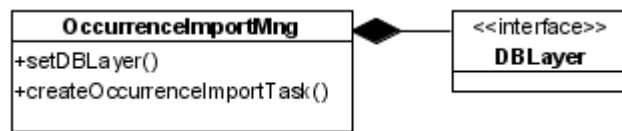
The Occurrence Import Task is in the file net.sf.plantlore.client.occurrenceimport.OccurrenceImportTask.java.

```
        <<interface>>            <<interface>>
           Task                 RecordProcessor


  DBLayerUtils             OccurrenceImportTask
                           +task()
                           +processRecord()
        <<interface>>
      OccurrenceParser
```
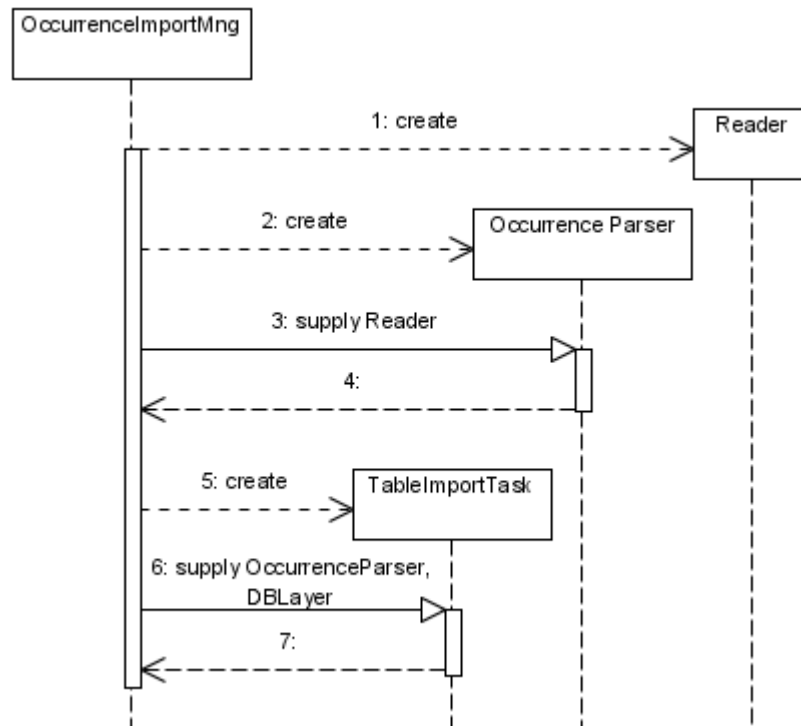
## The Import Manager

The Import Manager is the Import Task factory. It is used to store the partial information about the future Occurrence Import Task and once all the information required are available, it can create a new Record Processor, Parser and finally the Occurrence Import Task that can start the Parser.

net.sf.plantlore.client.occurrenceimport.OccurrenceImportMng.java.

## The Interaction

# Creating a New Occurrence Import Task



# Starting the Occurrence Import Task

The start of work of the Occurrence Import Task is quite complicated because several interfaces are merged together.

First, the XML Occurrence Parser is supplied a Record Processor. The XML Occurrence Parser creates a new XMLReader and registers itself as the Content Handler. The XMLReader will parse the input and send the SAX events back to the ContentHandler.

Second, start of the parsing is initiated from the Occurrence Import Task and comes to the XML Occurrence Parser who starts the XMLReader. The XMLReader starts parsing the input and reports events back to the XML Occurrence Parser which is also the Content Handler. The XML Occurrence Parser reconstructs the record and once the reconstruction is finished, the Occurrence Import task is called to process the record, because the Occurrence Import task also happens to be the Record Processor.