Beginner's Guide to Nipype v0.4

Release 0.2

Michael Notter

August 03, 2011

CONTENTS

CHAPTER

ONE

INTRODUCTION

1.1 Welcome to Nipype

The Beginner's Guide will introduce you to the main aspects of **Nipype**, a user-friendly software written in Python that provides a uniform interface to existing neuroimaging softwares like SPM, FSL, FreeSurfer, Camino, AFNI, Slicer, etc.

This guide will teach you all you need to know to get started. It will explain all you need to know about the basics of Nipype, how a pipeline is structured and how you can create your own pipeline. It will show you in a step by step example how to set up a first and a second level analysis on the volume or the surface. It will introduce you to some additional knowledge about how to use Nipype to extract anatomical or functional regions of interests (ROIs) and how to create some basic pictures of functional slices.

Note that this guide is ment as a general introduction. The implementation of Nipype is nearly unlimited and there is alot of advanced knowledge that won't be covered by this guide but can be looked up at various places on the nipype homepage.

Note: There are also alot of very good tutorials specific for different usages of Nipype like using FreeSurfer for smoothing, using FSL for DTI analysis, for fMRI analysis with FEEDS data, for fMRI analysis or using SPM for analysis on auditory dataset etc. They all can be found in the tutorial section of the Nipype homepage.

1.2 What if you're lost?

If you have any questions about Nipype and can't find the answer on the nipype homepage feel free to contact the Nipype mailinglist or to search its archive. You can sign up here.

If you have any questions about this Beginner's Guide or want to give its author any kind of feedback or suggestions, please contact me at mnotter@mit.edu. It's highly appreciated.

1.3 But first... Preparation!

Befor you can start, make sure that you have all the necessary modules and software installed on your system. For more information go to the download and install section.

Befor you can run a nipype python script make also sure to set up the necessary environment variables, meaning to source to your corresponding Nipype and FreeSurfer and to export the corresponding FreeSurfer and Matlab paths. With the current Nipype v0.4 and FreeSurfer v5.1.0 the terminal commands can look like that:

```
source /software/python/setup-nipype-0.4.sh
source /software/Freesurfer/5.1.0/SetUpFreeSurfer.sh

export MATLABCMD=$pathtomatlabdir/bin/$platform/MATLAB

export FREESURFER_HOME=/software/Freesurfer/5.1.0
```

```
export FSFAST_HOME=/software/Freesurfer/5.1.0/fsfast
export SUBJECTS_DIR=/software/Freesurfer/5.1.0/subjects
export MNI_DIR=/software/Freesurfer/5.1.0/mni
```

After you've installed everything and set up all the necessary environment variables you can start iPython (command: `ipython`), a strongly recommended interactive python environment. Now you're good to go!

CHAPTER

TWO

HOW TO BUILD A PIPELINE

This section is meant as a step by step introduction to building your own pipeline. At the end you should know what the important characteristics of a pipeline is, how it is constructed, how its parts are connected, so that you are ready to implement your own pipeline.

2.1 What is a pipeline?

A pipeline in the Nipype sense is a sequence of procedures to automate the analysis of fMRI-data. A pipeline or also called workflow is built by connecting specific nodes to each other. In the context of nipype, nodes contain specific functions or algorithms of interfaces such as SPM, FSL, FreeSurfer etc. All those nodes have defined inputs and outputs. Creating a workflow then is a matter of connecting appropriate outputs to inputs. The main advantage of the pipeline is that it can use different modules from different packages (e.g. SPM, FSL, FreeSurfer, Camino, AFNI, Slicer) and is able to exchange the data between them.

2.2 Create the framework of your own pipeline

There are many ways to construct a pipeline but in the end it comes down to the following steps:

1. Import appropriate modules
2. Define nodes
3. Define pipeline(s)
4. Create connections
5. Visualize pipeline
6. Execute pipeline

2.2.1 Import appropriate modules

The first thing you'll have to do is to import the interfaces you want to use. That depends on the nodes and algorithms you want to use in your pipeline. You can either import an interface and give it a specific name or import only a desired algorithm of it.

```
# imports the engine interface as 'pe'
import nipype.pipeline.engine as pe

# imports only the function Bunch from the base interface
from nipype.interfaces.base import Bunch
```

Important: If you use the freesurfer interface, please make sure to tell freesurfer where the subjects directory is by using the following command:

```
import nipype.interfaces.freesurfer as fs
freesurfer_dir = '~SOMEPATH/freesurfer'
fs.FSCommand.set_default_subjects_dir(freesurfer_dir)
```

2.2.2 Define nodes

Node Initiation

Before the parameters of a node can be specified they first have to be initiated. The initiation is quite simple and done as follows:

```
nodename = pe.Node(interface=interface.algorithm(), name='visibleName')
```

- **nodename**: name of the variable which identifies the node in the code
- **pe.Node**: defines the characteristic of the node, which can be a Node, a MapNode or a Workflow
- **interface**: name of the imported interface you want to use (e.g. SPM, FSL, FreeSurfer,...)
- **algorithm**: name of the algorithm you want the node to execute
- **visibleName**: name which is used for the naming of the folder the node output is stored in and the name of a node in the pipeline graph (recommended to be the same as nodename)

```
#Example of an initiation of a spm-realignment node
import nipype.interfaces.spm as spm
realign = pe.Node(interface=spm.Realign(), name='realign')
```

Note: The difference between a Node and a MapNode is explained here

Node Parameters

Depending on the purpose of a node and its underlying algorithm, different parameters can be specified. They can be distinguished into:

1. **mandatory inputs**: inputs that have to be given
2. **optional inputs**: inputs to get the node to behave in a specific way
3. **outputs**: the possible outputs that a node creates

But how can you find out what the possible inputs and outputs of a node are?

- check the section Interfaces and Algorithms on the nipype homepage by clicking on the node you're interested in
- use the help method of a module in iPython (e.g. `fsl.Smooth.help()`)
- view the docstring of a module in iPython (e.g. `fsl.MCFLIRT?`) which showes you the documentation and an example of an implementation in the command window.

Node Specification

The specification of a node can be done in three ways.

```
#1. specify parameters in the during initiation
mybet = fsl.BET(in_file='foo.nii', out_file='bar.nii')

#2. specify parameters after initiation
mybet = fsl.BET()
mybet.inputs.in_file = 'foo.nii'
mybet.inputs.out_file = 'bar.nii'
```

```
#3. specify parameters when running a node
mybet = fsl.BET()
mybet.run(in_file='foo.nii', out_file='bar.nii')
```

Iterables (optional)

If you want a node to be executed over different sets of data (e.g. different subjects, different conditions, different smoothing kernels,...) you have to use iterables.

```
nodename.iterables = ('input_to_iterate_over',[conditions_to_iterate_over])
```

E.g. If you want to execute a pipeline for subject1 and subject2 and you want to run the pipeline with a smoothing kernel of 4 and one of 8 you do the following:

```
startnode.iterables = ('subject_id', ['subject1','subject2'])
smoothnode.iterables = ('fwhm', [4, 8])
```

Iterfields

If you'll use MapNodes you'll also have to use an iterfield. This enables running the underlying interface over a set of inputs and is particularly useful when the interface can only operate on a single input. A good tutorial to iterables and iterfields can be found at MapNode, iterfield, and iterables explained

stand-alone node

If you want to run an algorithm without being a node or being a part of a pipeline you only have to define the nodename, interface and algorithm.

```
smooth = spm.Smooth()
```

This is most of the time used if you want to test a node or if you want to use the nipype environment to run the different kind of algorithms without using them inside a pipeline. For example, it isn't recommended to use the `recon-all` algorithm inside a pipeline or more extrem in parallel mode because of its computational and time costs. Nonetheless you can use Nipype to execute the recon-all process.

Individual nodes

If you want to create an individual node by yourself that doesn't use an algorithm of an interface already specified and you want to use the advantage of input and output fileds you can build your own node with the `IdentityInterface` method of the `utility` interface.

```
#import the utility interface
import nipype.interfaces.utility as util

#define the fields you want to use
individualnode = pe.Node(interface=util.IdentityInterface(fields=['field1','field2']),
                         name='individualnodename')
```

Now we have designed our own node with the in- and output field 'field1' and 'field2'. If you now want the node to execute a specific kind of algorithm, you'll have to add a function to the connections to the inputnode. How this can be done is described in **4. Create connections - Modifying inputs to nodes**.

2.2.3 Define pipeline(s)

The initiation of a pipeline is quite the same as the one of a node. Except that you don't need to declare an interface.

```
workflow = pe.Workflow(name='preproc')
```

Cloning

If you've already created a pipeline with all its connections and want to reuse it in another part of the workflow you can simply clone it with the `clone` method.

For example, if you've already created an analysis pipeline for the workflow after a volume preprocess and want now to reuse this algorithm for the analysis of preprocessed sufacedata you can clone the volanalysis like this:

```
surfanalysis = volanalysis.clone(name='surfanalysis')
```

This cloneing has to be done, because if you would use the volanalysis again, the data of the volanalysis would be overwritten. Because of that and because of the ambiguity of the connections in the workflow, the pipeline would not run. To solve this problem, every node and pipeline has to have its unique name.

If you want to change some parameters of the pipeline after cloning you just have to specify the exact pipeline, node and parameter you want to change:

```
surfanalysis.inputs.level1design.timing_units = 'secs'
```

Which now would set the input field `timing_units` of the level1design node which is part of the surfanalysis pipeline to 'secs'.

2.2.4 Create connections

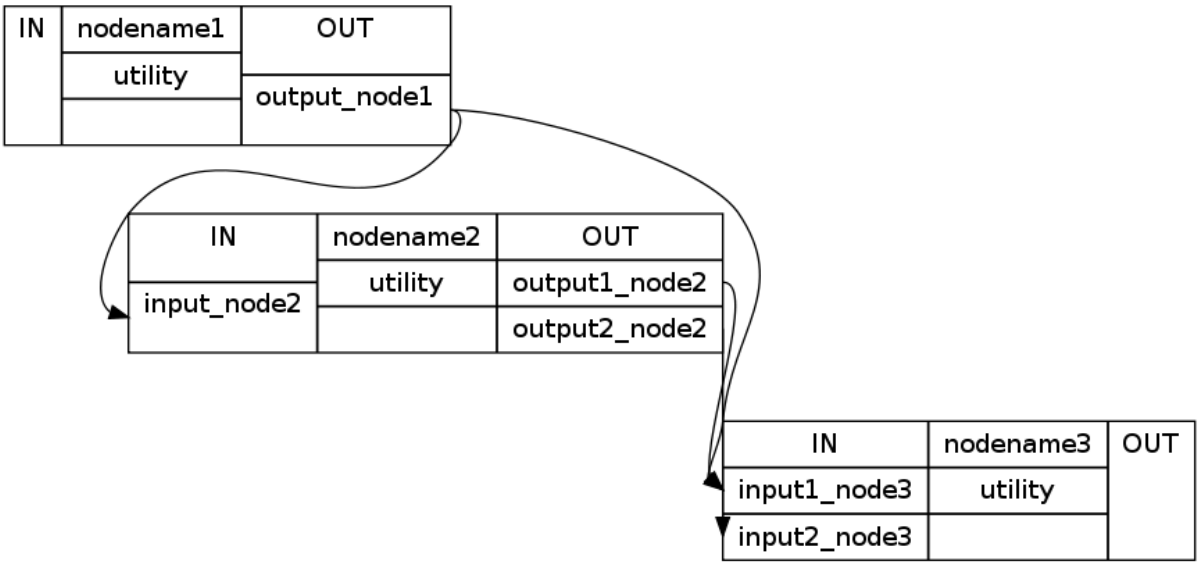
This is the essential part of creating a pipeline and leads to the main advantage of the pipeline which is to execute everything autonomous, in one workflow and if you want to in parallel.

Connect nodes to each other

There is a basic and an advanced way to create connections between two nodes. The basic way allows only to connect two nodes at a time whereas the advanced one can establish multiple connections at once.

```
#basic way to connect two nodes
workflowname.connect(nodename1, 'out_files_node1', nodename2, 'in_files_node2')

#advanced way to connect multiple nodes
workflowname.connect([(nodename1, nodename2, [('output_node1', 'input_node2')]),
                      (nodename1, nodename3, [('output_node1', 'input1_node3')]),
                      (nodename2, nodename3, [('output1_node2', 'input1_node3'),
                                              ('output2_node2', 'input2_node3')
                                              ])
                      ])
```

The advanced connection example would as a detailed graph look like this:



It is important to point out that you don't just have to connect the nodes, but rather to connect the output and input fields of each node.

Connect pipelines to each other (necessary if you have multiple pipelines)

If you have multiple pipelines, like one for preprocessing, one for modelestimation and one for volume analysis, you can't just connect the nodes to each other. You have to connect the pipelines to each other instead.

Assumed that we have a node "realign" which is part of a pipeline called "preprocess" and that we have a node called "modelspec" which is part of a pipeline called "modelestimation". If we now want to connect does to pipelines at those particular points we first have to create a kind of meta-pipeline which contains those to pipelines. This initiation and the following connections would look like that:

```
frameflow = pe.Workflow(name='frameflow')
frameflow.connect([(preprocess, modelestimation,[('realign.out_files', 'modelspec.in_files')
                                                 ])
                   ])
```

You see that the main difference to the connections between nodes is that you connect the pipelines, but have to specify which nodes with which output should be connected to which nodes with which input.

Add node(s) to pipeline (optional)

If you want to run a node by itself without connecting it to any other node, you can do that with the add_nodes method.

```
#adds node smooth and node realign to the pipeline
workflow.add_nodes([smoother, realign])
```

Modifying inputs to nodes

If you want to modify the output of a node before sending it to the next one you can do that by adding a function into the connection process.

First you have to define your function that modifies the data and returns the new output. If you have done this, than you can insert the function into the connection process.

```
#your function that does something
def myfunction(input_from_node):
```

```
    #changes the data as you defined
    output_for_node_2 = input_from_node * 2

    #return the output
    return output_for_node_2


#connection of two nodes with a function in between
workflowname.connect([(nodename1, nodename2,[(('out_file_node1', myfunction),
                                              'in_file_node2')]),
                      ])
```

This will take the output of 'out_file_node1' and give it as an argument to the function `myfunction`. The return value that will be returned by `myfunction` then will be forwarded as input to 'in_file_node2'.

If you want to insert more than one parameter into the function do as follows:

```
def myfunction(input_from_node, additional_input):

    output_for_node_2 = input_from_node + additional_input
    return output_for_node_2


#connection of two nodes with a function in between which takes two arguments
workflowname.connect([(nodename1, nodename2,[(('out_file_node1', myfunction, additional_input),
                                              'in_file_node2')]),
                      ])
```

2.2.5 Visualize pipeline

To visualize the flow of a pipeline you can use the method `write_graph()`.

```
#Example of using write_graph
workflowname.write_graph(graph2use='flat')
```

This method will create two files:

- **graph.dot:** which contains the general connections between nodes
- **graph_detailed.dot:** which contains the detailed connections between nodes with the individual output and input fileds.

If graphviz is installed the dot files will automatically be converted into png-files. Otherwise you can take the dot files and load them in a graphviz visualizer elsewhere.

You can also specify the deepness you want the graph to show by changing the argument `graph2use`:

- **'orig':** only the highest level of the workflow will be visualized (e.g. this would show you only the frame-flow and leave the contained pipelines out)
- **'flat':** all levels of the workflow are visualized once
- **'exec':** all levels of the workflow are visualized and an iteration of a field will lead to a splitting of the graph (this is a very good way to see what the benefit of executing the pipeline in parallel mode is)

2.2.6 Execute pipeline

After setting everything up, the pipeline can be executed with `run()`. You can run a pipeline either in serial or in parallel mode.

```
#To run the pipeline serial
workflow.run(plugin='Linear')
```

```
#To run the pipeline paralle using 2 processes
workflow.run(plugin='MultiProc', plugin_args={'n_procs' : 2})
```

Hint: A good tutorial how to set up the parallel mode can be found under Distributed processing with nipype.

2.3 How to get data into and out of your pipeline

After constructing the framework of the pipeline in the previous section, we're almost ready to execute the pipeline. But first we have to define the data we want to run the pipeline on and the results we want to get out of it. To do that we have to concider the following points:

1. **Infosource:** to define the list of subjects on which the pipeline will be executed
2. **Datasource:** to grab the data you want to use and insert it into the pipeline
3. **Inputnode:** to distribute the data and experiment specific parameters to the pipeline (optional)
4. **Datasink:** to store specific outputs at a given place (optional but recommended)
5. **Model Specification:** to feed the pipeline with model specific components like contrast, conditions, onset times etc.
6. **Connect new nodes to your pipeline:** to connect the infosource, datasource, inputnode and datasink with the framework of your pipeline

2.3.1 Infosource

The best way to tell a pipeline on which subjects it should be executed on is to build an infosource node. The only thing that this node contains is a list of the subjects and the instructions to execute the pipeline on each of this subjects. This is done with the iterables method.

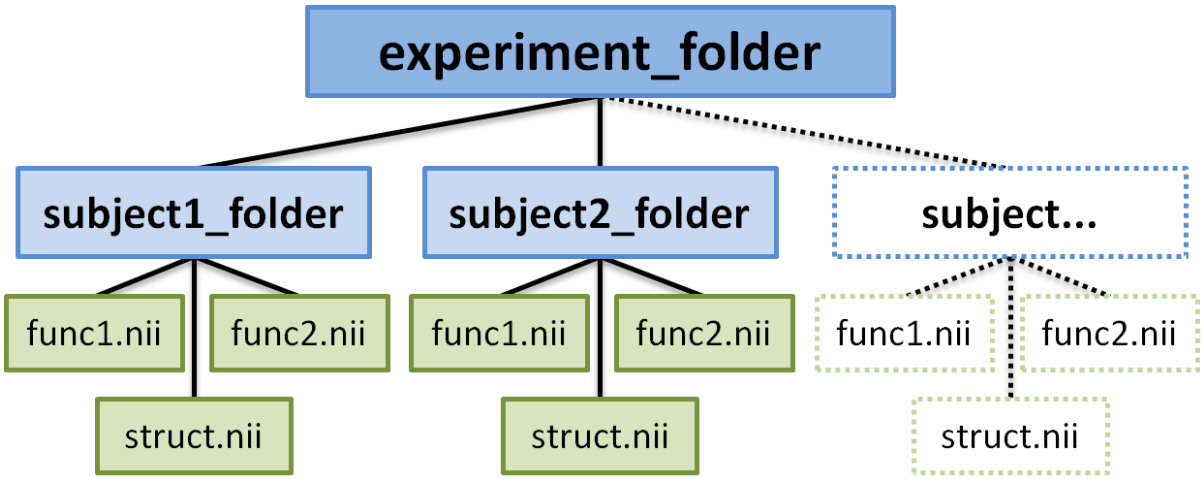
```
#import the utility interface
import nipype.interfaces.utility as util

#initiate the infosource node
infosource = pe.Node(interface=util.IdentityInterface(fields=['subject_id']),
                     name="infosource")

#define the list of subjects your pipeline should be executed on
infosource.iterables = ('subject_id', ['subject1','subject2','subject3'])
```

2.3.2 Datasource

To get the subject specific data into the pipeline we need the datasource node. As the name of the algorithm implies the DataGrabber grabs the data from a specified folder and stores it in the specified output fields.

```
#initiate the DataGrabber node with the infield: 'subject_id'
#and the outfield: 'func' and 'struct'
datasource = pe.Node(interface=nio.DataGrabber(infields=['subject_id'],
                                               outfields=['func', 'struct']),
                     name = 'datasource')
```

To use the datagrabber it is important to know what the exact structure of your folders is and where the data is stored at. In this example we assume that the layout of our data is as following:



As you can see all the necessary data is stored in the experiment folder. The data of each subject is stored in its individual subject_folder. The name of this folder changes with each subject. There are two ways how you can define the structure of data you want to grab.

```
#to specify the location of the experiment folder
datasource.inputs.base_directory = '~/experiment_folder'

#define the structure of the data folders and files.
#Each '%s' will later be filled by a template argument.
datasource.inputs.template = '%s/%s.nii'

#First way: define the arguments for the template '%s/%s.nii' for each field individual
datasource.inputs.template_args['func'] = [['subject_id', ['func1','func2']]]
datasource.inputs.template_args['struct'] = [['subject_id','struct']]

#Second way: store all the arguments for the template in a dictionairy and ...
info = dict(func=[['subject_id', ['func1','func2']]],
            struct=[['subject_id','struct']])
#... pass it to template_args.
datasource.inputs.template_args = info
```

Note: The values defined in `template_args` will be filled into the placeholders of `template`. Because 'subject_id' is defined as ['subject1','subject2','subject3'] (see definition of infosource node), the outfield 'func' of the datagrabber node will store 'subject1/func1.nii', 'subject1/func2.nii' and 'subject1/struct.nii' in the 'struct' outfield for subject1.

2.3.3 Inputnode (optional)

If you want to keep a clearly arranged distribution of the input data it is suggested to create an inputnode that serves that purpose. This inputnode specifies and collects all the inputs that are needed for the workflow and distributes them to specific places in the pipeline.

```
#import the utility interface
import nipype.interfaces.utility as util

#define the inputnode with the fields you want to distribute
inputnode = pe.Node(interface=util.IdentityInterface(fields=['func',
                                                             'subject_id',
                                                             'session_info',
                                                             'contrasts']),
                    name='inputnode')
```

2.3.4 Datasink (optional)

Sometimes you have some output you want to store at an easy accessible place so that you don't have to search in the depth of your workingdir where all in- and outputs of every node is stored. For this purpose the datasink node was created:

```
#import i/o routines
import nipype.interfaces.io as nio

#initiate node
datasink = pe.Node(interface=nio.DataSink(), name="datasink")

#specify the name and location of the datasink folder
datasink.inputs.container = 'name_of_datasink_folder'
datasink.inputs.base_directory = '~/experiment_folder'

#define the outputs you want to store by connecting them to the datasink node
metaflow.connect([(frameflow,datasink,[('preproc.bbregister.out_reg_file',
                                        'bbregister'),
                                       ('volanalysis.contrastestimate.spm_mat_file',
                                        'spm_mat_file'),
                                       ])
                  ])
```

Note: The name that you give the input-filed of the datasink node (here 'bbregister' and 'spm_mat_file') will be taken as a name giver for the subfolder where those specific files will be stored in the datasink folder.

The datasink node is really usefull to keep controll over your storage capacity. If you store all important files that you'll need for further analysis in this folder you can delete the `workingdir` of the pipeline after executing and counteract storage shortage. You can even set up the configuration of the pipeline so that it will not creat a workingdir at all. For more information go to Configuration File.

2.3.5 Model Specification

Before we can run our pipeline we have to feed it with model specific components as the name of the conditions, the contrasts and onset times. We might also want to add some parametric modulators or regressors etc.

Contrasts

To insert all the contrast specific values into the pipeline we first have to save them into a variable, in this case called `contrasts`. The structure of this variable is a list of lists. The inner list specifies the contrasts and has the following format - [Name, Stat, [list of condition names], [weights on those conditions]. The condition names must be the same we later feed into `subjectinfo` function described below.

```
#Names of different conditions
namesOfConditions = ['basic','condition1','condition2','condition3']

#contrasts for all sessions
contrast_1 = ('basic vs. conditions','T', namesOfConditions, [3,-1,-1,-1])
contrast_2 = ('all vs. condition1',  'T', namesOfConditions, [0,1,0,0])
contrast_3 = ('all vs. condition2',  'T', namesOfConditions, [0,0,1,0])
contrast_4 = ('all vs. condition3',  'T', namesOfConditions, [0,0,0,1])

#contrasts for e.g. session 1 and 3 out of ['session1','session2','session3']
contrast_5 = ('1+3 vs. condition1',  'T', namesOfConditions, [0,1,0,0],[1,0,1])
contrast_6 = ('1+3 vs. condition2',  'T', namesOfConditions, [0,0,1,0],[1,0,1])
contrast_7 = ('1+3 vs. condition3',  'T', namesOfConditions, [0,0,0,1],[1,0,1])

#store all contrasts into a list
contrasts = [contrast_1,contrast_2,contrast_3,contrast_4,
             contrast_5,contrast_6,contrast_7,contrast_8]
```

```
#feed those contrasts to the inputnode filed 'contrasts'
frameflow.inputs.inputnode.contrasts = contrasts
```

Session info

Here we create a function that returns session specific information about the experimental paradigm. This is needed by the SpecifyModel function to create the information necessary to generate an SPM design matrix. This function `subjectinfo` is used to feed the inputnode `session_info` for each subject with the paradigm conditions.

```
def subjectinfo(subject_id):

    #import Bunch from interface base
    from nipype.interfaces.base import Bunch

    #restate the names of
    namesOfConditions = ['basic','condition1','condition2','condition3']

    #Onset Times in seconds for condition ['basic','condition1','condition2','condition3'
    onsetTimes = [[1,10,42,49.6,66.1,74.1,97.6,113.6,122.2,130.2,137.2,153.7,169.2,
                   185.7,201.8,290.4,313.4,321.4,377.5,401.5,410,418.6,442.1,473.6],
                  [17.5,82.1,89.6,145.2,225.3,242.3,281.4,426.6],
                  [26,162.2,209.3,249.3,265.9,205.4,450.1,386],
                  [34,273.4,329.5,338.5,354,362,370,466.4]
                  ]

    #to define two parametric modulators for 'condition1','condition2','condition3'
    para_modu = [None,
                 base.Bunch(name=['target2','target3'], poly=[[1],[1]],
                            param = [[0,0,1,0,0,0,0,0],[0,0,0,0,1,0,0,1]]),
                 base.Bunch(name=['target2','target3'], poly=[[1],[1]],
                            param = [[0,0,0,1,1,1,0,0],[1,0,0,0,0,0,1,1]]),
                 base.Bunch(name=['target2','target3'], poly=[[1],[1]],
                            param = [[0,1,0,0,0,1,0,1],[0,0,0,0,1,0,1,0]]),
                 ]

    #to feed this information to the inputnode we have to store the information
    #in a list 'output' which we will return later
    output = []

    #the parameters get added three times if we have three sessions like in this example.
    #if you need to, you would be able here to specify the session specific parameters
    #for each session differently
    for r in range(3):
        output.append(Bunch(conditions=namesOfConditions,
                            onsets=onsetTimes,
                            durations=[[8] for s in namesOfConditions],
                            amplitudes=None,
                            tmod=None,
                            pmod=para_modu,
                            regressor_names=None,
                            regressors=None))
    return output
```

Note: A detailed instruction on how to set the model specific parameters can be found in the in the Model Specification section.

2.3.6 Connect new nodes to your pipeline

Before you can run your pipeline you will have to connect infosource, datasource, inputnode and datasink to each other and to the pipelines of your framework workflow, here called frameflow. For this purpose you will have to create a meta pipeline.

```
#initiate the meta workflow
metaflow = pe.Workflow(name='metaflow')

#connect infosource, datasource and inputnode to each other
metaflow.connect([(infosource, datasource,[('subject_id','subject_id')]),
                  (datasource,inputnode,[('func','func'),
                                         (('subject_id', subjectinfo),'session_info'),
                                         ]),
                  #connect the inputnode to your workflow
                  (inputnode, frameflow,[('func','surfsmooth.in_file'),
                                         #...etc...
                                         ])
                  #connect output you want to be stored into datasink
                  (frameflow,datasink,[('preproc.bbregister.out_reg_file',
                                        'bbregister'),
                                       ('volanalysis.contrastestimate.spm_mat_file',
                                        'spm_mat_file'),
                                       ])
                  ])
```

Now you're done and can run your pipeline.

```
metaflow.run(plugin='Linear')
```

CHAPTER

THREE

PREPARE DATA FOR NIPYPE

The first step after data acquisition is as expected: **Preparation!** Before we can run the data threw a pipeline we have to convert the dicoms into niftis and execute the recon-all process for each subject. Otherwise we can't use the benefits of FreeSurfer. Because this can be done in rather little code it is a very good example to get started.

3.1 Convert Dicoms into Niftis

Let us start with the conversion of the 3D dicoms into 4D nifti files.

3.1.1 Import modules

The first step of every pipeline is always to import all the necessary modules. In this case we only need the basic modules `os`, `util`, `pipeline.engine` and the `freesurfer interfase`.

```
import os                                    # system functions
import nipype.interfaces.freesurfer as fs    # freesurfer
import nipype.interfaces.utility as util     # utility
import nipype.pipeline.engine as pe          # pypeline engine
```

3.1.2 Define experiment specific parameters

After importing the modules, it is recommended to specify all the necessary variables like the path to the experiment folder, a list of the subject identifiers etc.

```
#Specification of the folder where the dicom-files are located at
experiment_dir = '~SOMEPATH/experiment'

#Specification of a list containing the identifier of each subject
subjects_list = ['subject1','subject2','subject3','subject4']

#Specification of the name of the dicom and output folder
dicom_dir_name = 'dicom' #if the path to the dicoms is: '~SOMEPATH/experiment/dicom'
data_dir_name = 'data'   #if the path to the data should be: '~SOMEPATH/experiment/data'
```

3.1.3 Define nodes to use

Let us now construct the three nodes we want to use to create our preparation pipeline. First we need to define the infosource-node which specifies on which subjects the workflow is run on.

```
#Node: Infosource - we use IdentityInterface to creat our own node, to specify
#                   the list of subjects the pipeline should be executed on
infosource = pe.Node(interface=util.IdentityInterface(fields=['subject_id']),
```

```
                                                        name="infosource")
infosource.iterables = ('subject_id', subjects_list)
```

Now we define the main node for the data conversion, DICOMConvert.

```
#Node: DICOMConvert - converts the .dcm files into .nii and moves them into
#                     the folder "data" with a subject specific subfolder
dicom2nifti = pe.Node(interface=fs.DICOMConvert(), name="dicom2nifti")
```

If we now look at the secion nipype.interfaces.freesurfer.preprocess about the DICOMConvert node we see the following:

[Mandatory]

```
base_output_dir : (a directory name)
                  directory in which subject directories are created
dicom_dir : (an existing directory name)
            dicom directory from which to convert dicom files
```

Remember that the path to the dicoms changes from subject to subject. Therefore we can't declare the varible `dicom_dir` right away. But because we already have an `infosource` node, we will be able to overcome this problem when we establish the connections between the nodes. Let us first define the path to the output dir:

```
dicom2nifti.inputs.base_output_dir = experiment_dir + '/' + data_dir_name
#This will store the output to '~SOMEPATH/experiment/data'
```

Now we have to specify the **optional** inputs `file_mapping`, `out_type` and `subject_dir`.

```
dicom2nifti.inputs.file_mapping = [('nifti','*.nii'),('info','dicom.txt'),('dti','*dti.bv*')]
dicom2nifti.inputs.out_type = 'nii'
dicom2nifti.inputs.subject_dir_template = '%s'
```

This `dicom2nifti` node will convert the dicoms into niftis and create a summary text file calles **shortinfo.txt** which contains most of the important informations about the dicoms:

```
dcmdir ~SOMEPATH/experiment/dicom/subject1
PatientName subject1
StudyDate 20150403
StudyTime 083819.578000
1 fl2d1 localizer_BC 578000-1-1.dcm
2 fl2d1 localizer_32 578000-2-1.dcm
3 fl3d1_ns AAScout 578000-3-100.dcm
4 tfl3d1_ns T1_MPRAGE_1mm_iso 578000-4-100.dcm
5 epfid2d1_90 ge_functionals_2meas 578000-5-1.dcm
6 fm2d2r field_mapping 578000-6-10.dcm
7 fm2d2r field_mapping 578000-7-10.dcm
8 epir2d1_96 ep2d_t1w 578000-8-10.dcm
9 epfid2d1_96 ge_functionals_125 578000-9-10.dcm
10 epfid2d1_96 ge_functionals_125 578000-10-10.dcm
```

But because we also want to know how many timepoints were acquired and what the resolution of the dicoms are we'll have to run a node called `ParseDICOMDIR` which gives us an output text file with the following content:

```
  1 578000-1-2.dcm    1 0     2 0  512 512   3   1    0.0086   4.0000 localizer_BC
  4 578000-2-2.dcm    2 0     2 0  512 512   3   1    0.0086   4.0000 localizer_32
  7 578000-3-1.dcm    3 0     1 0  128 128 128   2    0.0024   1.1300 AAScout
263 578000-4-1.dcm    4 0     1 0  256 256 176   1    2.5300   3.4800 T1_MPRAGE_1mm_iso
439 578000-5-1.dcm    5 0     1 1   96  90  28   2    9.5000  30.0000 ge_functionals_2meas
441 578000-6-1.dcm    6 0     1 0   96  96  28   1    0.5000   2.8300 field_mapping
469 578000-7-1.dcm    7 0     1 0   96  96  28   1    0.5000   5.2900 field_mapping
497 578000-8-1.dcm    8 0     1 0   96  96  28   1   10.0000  56.0000 ep2d_t1w
525 578000-9-1.dcm    9 1     2 1   96  96  28  60    8.0000  30.0000 ge_functionals_125
585 578000-10-1.dcm  10 1     2 1   96  96  28  60    8.0000  30.0000 ge_functionals_125
```

This output shows us that the T1-file has resoluiton of 256x256x176 and that the two functional runs, file 9 and 10, have 60 timepoints. Now, let us implement this `ParseDICOMDir` node.

```
#Node ParseDICOMDIR - for creating a nicer nifti overview textfile
dcminfo = pe.Node(interface=fs.ParseDICOMDir(), name="dcminfo")
dcminfo.inputs.sortbyrun = True
dcminfo.inputs.summarize = True
dcminfo.inputs.dicom_info_file = 'nifti_overview.txt'
```

As before with the dicom2nifti node, we have the problem with the dcminfo node, that the input variable `dicom_dir` changes for each subject. And as before, we will see how to handle this issue during the connection of the nodes.

3.1.4 Define pipeline

After we've defined all the nodes we want to use, we are ready to implement the preparation pipeline:

```
#Initiation of the preparation pipeline
prepareflow = pe.Workflow(name="prepareflow")

#Define where the workingdir of the all_consuming_workflow should be stored at
prepareflow.base_dir = experiment_dir + '/workingdir_prepareflow'
```

3.1.5 Specify node connections

Now we've defined all the nodes and implemented the preparation pipeline. The only thing missing is the connection of the different nodes in the pipeline. But before we can do this, it is important to be aware about outputs, that get created and the inputs that have to be provided and more importantly how they look like.
infosource:
This node provides an output called `subject_id` that is either
`'subject1'`,
`'subject2'`,
`'subject3'`,
`'subject4'`.

dicom2nifti:
This node needs `dicom_dir` as an input that is either:
`'~SOMEPATH/experiment/dicom/subject1/'`,
`'~SOMEPATH/experiment/dicom/subject2/'`,
`'~SOMEPATH/experiment/dicom/subject3/'`,
`'~SOMEPATH/experiment/dicom/subject4/'`.

dcminfo:
This node needs `dicom_dir` as an input that is either:
`'~SOMEPATH/experiment/data/subject1/'`,
`'~SOMEPATH/experiment/data/subject2/'`,
`'~SOMEPATH/experiment/data/subject3/'`,
`'~SOMEPATH/experiment/data/subject4/'`.

As you can see, for both nodes `dicom2nifti` and `dcminfo` the input from dicom_dir does change for each subject. But so also do the outputs of `infosource`. The only thing we have to do is to take the output `'subject1'` from `infosource` and change it into `'~SOMEPATH/experiment/dicom/subject1/'` for `dicom2nifti` and into `'~SOMEPATH/experiment/data/subject1/'` for `dcminfo`. This can be accomplished by inserting a function into the connection process.

Let's call this function `pathfinder`, which takes as arguments the subjectname and the foldername (either `dicom` or `data`) and returns `'~SOMEPATH/experiment/foldername/subjec_name/'`.

```
def pathfinder(subjectname, foldername):
    return os.path.join(experiment_dir, foldername, subjectname)
```

Now we can start with the connection of the nodes.

```
#Connect all components
prepareflow.connect([(infosource, dicom2nifti,[('subject_id', 'subject_id')]),
                     (infosource, dicom2nifti,[(('subject_id', pathfinder, 'dicom_dir_name'),
                                                'dicom_dir')]),
                     (infosource, dcminfo,[(('subject_id', pathfinder, 'dicom_dir_name'),
                                                'dicom_dir')]),
                     ])
```

Perhaps it's best if we take a closer look at the second connection. As it is written, the function `pathfinder` takes `'subject_id'` as its first argument which will represent `subjectname` in the function. `'dicom_dir_name'` will be given as the second argument `foldername` to the `pathfinder` function. The return value of the `pathfinder` function will than be sent as input `'dicom_dir'` to the `dicom2nifti` node.

3.1.6 Important for Nipype 0.4 users

There's an important issue about functions you have to consider in Nipype Version 0.4. If you would want to run the pathfinder function above you would encounter following error:

NameError: `("global name 'os' is not defined", 'Due to engine constraints all imports have to be done inside each function definition')`

This occurs because all values that you are using in a function have to be specified or imported within it's boundaries. Therefore we have to extend the function a bit by importing the `os` module and specifying the experiment_dir variable.

```
def pathfinder(subject, foldername):
    import os
    experiment_dir = '~SOMEPATH/experiment'
    return os.path.join(experiment_dir, foldername, subject)
```

3.1.7 Run pipeline

After the connection of the nodes there is only one last thing to do. To actually run the pipeline.

```
prepareflow.run(plugin='MultiProc', plugin_args={'n_procs' : 2})
```

Important: After running the prepareflow a folder called **data** gets created that contains all nifti files for each subject. For further work it is recommended to rename those files. For example:
- `578000-4.nii` into `struct.nii` because it is the T1 file
- `578000-9.nii` into `func1.nii` because it is the first functional run file
- `578000-10.nii` into `func2.nii` because it is the second functional run file

3.1.8 Clean up (optional)

After running the prepareflow we now have a folder called **data** that contains the converted nifti files. Additionally the pipeline has created a folder **workingdir_prepareflow** which contains alot of unnecessary outputs. But within this pile of data lie our `nifti_overview.txt` files for each subject, created by the `dcminfo node`. That's why I would recommend to add some additional python code after the run command to save those outputs in the corresponding subject subfolders in the data folder.

In our case the data we are interested in is in the folder `'~SOMEPATH/experiment/workingdir_prepareflow/prepareflow/_subject_id_subject1/dcminfo'` and we want to move it into

'/mindhive/gablab/u/mnotter/Desktop/TEST/data/subject1'. This can be accomplished with the following code:

```
#to run the loop for each subject
for subject in subjects_list:

    #specify where the nifti_overview.txt file is stored at
    from_path = os.path.join(prepareflow.base_dir,prepareflow.name,'_subject_id_%s'%subject,
                             dcminfo.name,dcminfo.inputs.dicom_info_file)

    #specify where to store the  nifti_overview.txt file at
    to_path = os.path.join(dicom2nifti.inputs.base_output_dir,subject)

    #with os.system('text') you're able to state the command 'text' in your terminal
    #therefore we use mv to move the date
    os.system('mv %s %s'%(from_path, to_path))
```

Note:
Following are the values of the variables we are using

```
prepareflow.base_dir                = '~SOMEPATH/experiment/workingdir_prepareflow'
prepareflow.name                    = 'prepareflow'
'_subject_id_%s'%subject            = '_subject_id_subject1' for the first run
dcminfo.name                        = 'dcminfo'
dcminfo.inputs.dicom_info_file      = 'nifti_overview.txt'
dicom2nifti.inputs.base_output_dir = '~SOMEPATH/experiment/data'
```

To finally delete the workingdirectory of the prepare pipeline with all it's content we can use the command `rm -rf ~SOMEPATH/experiment/workingdir_prepareflow` and we're done.

```
os.system('rm -rf %s'%prepareflow.base_dir)
```

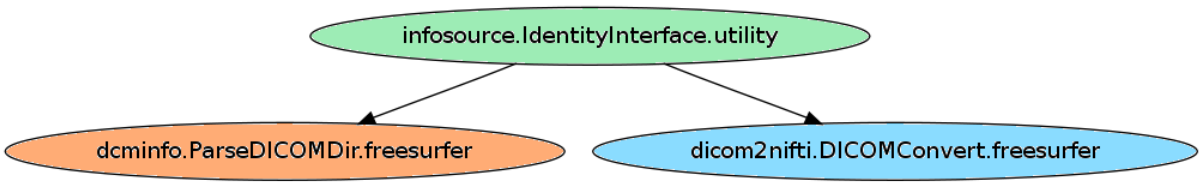
3.2 How to create and read graph.dot files

After the connction of all nodes is established we are able to creat a graph.dot file which gives us a very good overview over our pipeline. For the example above, the command would simply be:

```
prepareflow.write_graph(graph2use='flat')
```

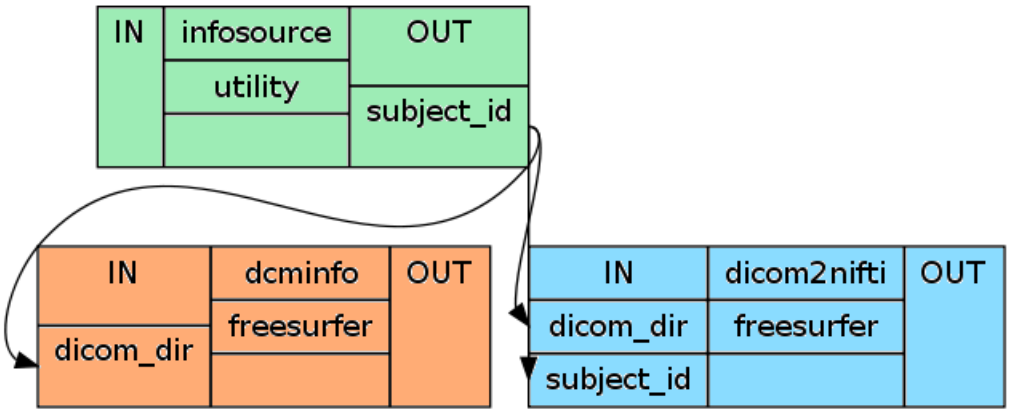
There are different kinds of graphs you can creat. But basically there is a simple overview graph called graph.dot that shows you the basic connections between nodes. In the more detailed version which creates a graph_detailed.dot file you can also see which outputs and inputs are connected together.

3.2.1 graph2use='flat' - graph.dot



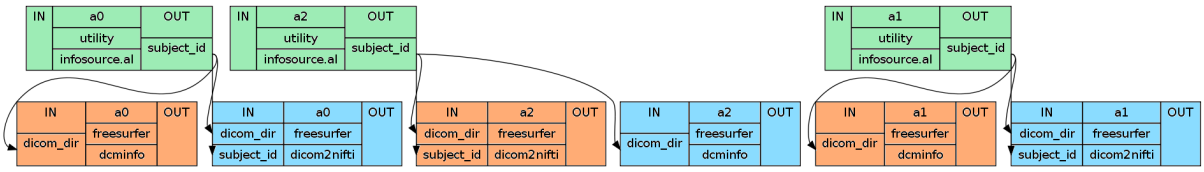
In the normal-dot file you can see how the nodes are connected generally. The name is composed in the format of `nodename.algorithm.interface`

3.2.2 graph2use='flat' - graph_detailed.dot



The detailed version of the graph gives you better informations about which input is connected to which output.

3.2.3 graph2use='exec' - graph_detailed.dot



If you set `graph2use` to `'exec'` you get the same level of detail as if you would run it with `'flat'` but all the runs that would be running in parallel are shown at once.

3.3 Run Recon-All with Nipype

Now that we are able to build a preparation pipeline, it will be a piece of cake to create a recon-all pipeline. Only after executing the recon-all algorithm will we be able to use all the benefits of the FreeSurfer interface.

3.3.1 Prepare modules and important variables

First let us again import all necessary modules and specify the experiment specific parameters.

```
import os                                    # system functions
import nipype.interfaces.freesurfer as fs    # freesurfer
import nipype.interfaces.utility as util     # utility
import nipype.pipeline.engine as pe          # pypeline engine

#Specification of the folder where the dicom-files are located at
experiment_dir = '~SOMEPATH/experiment'

#Specification of a list containing the identifier of each subject
subjects_list = ['subject1','subject2','subject3','subject4']

#Specification of the output folder - where the T1 file can be found
data_dir_name = 'data'

#Node: SubjectData - we use IdentityInterface to creat our own node, to specify
#      the list of subjects the pipeline should be executed on
infosource = pe.Node(interface=util.IdentityInterface(fields=['subject_id']),
                                                      name="infosource")
```

```
infosource.iterables = ('subject_id', subjects_list)

#Node: Recon-All - to generate surfaces and parcellations of structural
#                  data from anatomical images of a subject.
reconall = pe.Node(interface=fs.ReconAll(), name="reconall")
reconall.inputs.directive = 'all'

#Because the freesurfer_data folder doesn't exist yet
os.system('mkdir %s'%experiment_dir+'/freesurfer_data')

reconall.inputs.subjects_dir = experiment_dir + '/freesurfer_data'
T1_identifier = 'struct.nii' #This is the name we manually gave the T1-file
```

Important: If we don't create the `freesurfer_data` folder before we specify the `subjects_dir` variable we would get following error:
TraitError: `The 'subjects_dir' trait of a ReconAllInputSpec instance must be an existing directory name, but a value of '~SOMEPATH/experiment/freesurfer_data' <type 'str'> was specified.`

Now we are ready to implement the pipeline and connection the `infosource` with the `reconall` node. As in the example about the conversion of the dicoms into niftis, we will use again a function called `pathfinder` to specify the exact location of the T1-file of each subject. Note that the function takes three arguments in this case.

```
#implementation of the workflow
reconflow = pe.Workflow(name="reconflow")
reconflow.base_dir = experiment_dir + '/workingdir_reconflow'

#defenition of the pathfinder function
def pathfinder(subject, foldername, filename):
    import os
    experiment_dir = '~SOMEPATH/experiment/experiment'
    return os.path.join(experiment_dir, foldername, subject, filename)

#connection of the nodes
reconflow.connect([(infosource, reconall,[('subject_id', 'subject_id')]),
                   (infosource, reconall,[(('subject_id', pathfinder, data_dir_name,
                                            T1_identifier),'T1_files')]),
                   ])

#run the recon-all pipeline (as recommended in serial mode)
reconflow.run(plugin='Linear')

#to delete the workingdir of the reconflow we use again the shell-command "rm".
#The important recon-all files are already stored in the "freesurfer_data" folder
os.system('rm -rf %s'%reconflow.base_dir)
```

CHAPTER

FOUR

EXAMPLE OF A FIRST LEVEL ANALYSIS PIPELINE

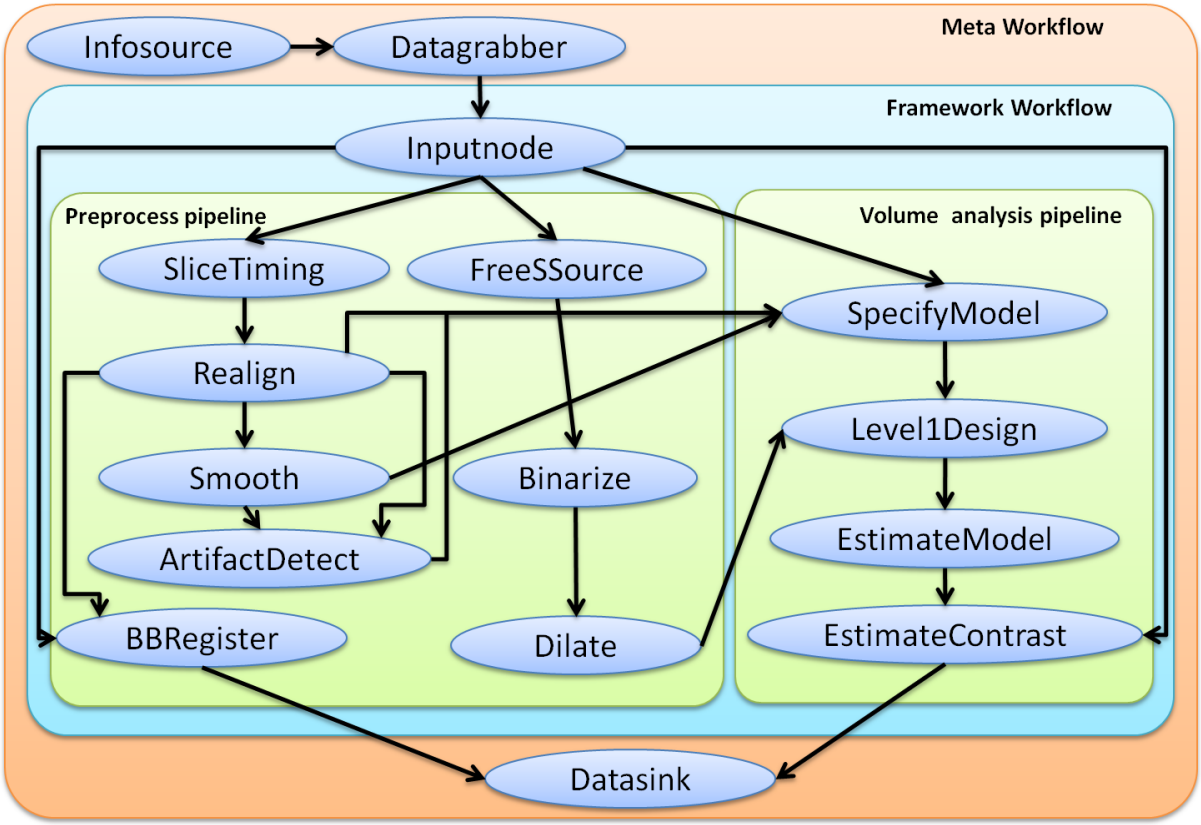
This is an example of a simple first level analysis pipeline. This pipeline takes the raw nifti data, does some preprocessing (e.g. realignment and smoothing) which is followed by the estimation of the concrete model.

Note: The normalization of the results to a common template will not be coverd by this chapter. A brief instruction on how to normalize your date with ANTS will be given in a latter chapter.

4.1 Define structure of your pipeline

The best way to build your own pipeline is first to think about the workflow the data should go through. In this example we want to first run some preprocessing like *SliceTiming*, *Realignment*, *ArtifactDetection*, *BBRegister* and *Smoothing*. Additionally we also want to take the subject specific aseg file from the freesurfer folder, binarize it and dialte it to create a mask for the level1desing. All this will be accommodate in a **preprocess pipeline**. After that we want to estimate a concrete model by running first *SpecifyModel*, create a *Level1Design*, *EstimateModel* and finally *EstimateContrast*. This will be done in a **volume analysis pipeline**.

The preprocess and the volume analysis pipeline together with the inputnode, will create the basic **framework workflow**. And this framework workflow will be connected to the infosource, the datagrabber and the datasink by a higher workflow, I'd like to call **meta workflow**. Visualized this pipeline would look like the following:



The connections between the nodes will make more sense once we look at the inputs a given node needs. Some may also ask why we haven't included the infosource, datagrabber and datasink nodes into the frameflow? The reason is that those nodes dependent highly on the paradigma specific parameters and will change for every model/experiment. That's why we want to separat them from the framework workflow which will stay more or less the same for similar experiments. This does also give us the opportunity to just import this framework workflow into a new pipeline script.

4.2 Write your pipeline script

Now that we have defined how the structure of our pipeline and the connections between them should be we can start with writing the pipeline script.

4.2.1 Import modules

First we have to import all necessary modules.

```
import os                                    # system functions
import nipype.algorithms.modelgen as model   # model generation
import nipype.algorithms.rapidart as ra      # artifact detection
import nipype.interfaces.freesurfer as fs    # freesurfer
import nipype.interfaces.io as nio           # i/o routines
import nipype.interfaces.spm as spm          # spm
import nipype.interfaces.utility as util     # utility
import nipype.pipeline.engine as pe          # pypeline engine
import nipype.interfaces.base as base        # base routines
import nipype.interfaces.fsl.maths as math   #for dilating of the mask
```

4.2.2 Define experiment specific parameters

I suggest to keep things that change often between versions, models, experiments like subject names, output folders and name of functional runs at one place so that they can be accessed more easily in case you want to change them.

```
#To better access the parent folder of the experiment
experiment_dir = os.getcwd()

#name of the subjects, functional fiels and output folders
subjects = ['subject1','subject2','subject3']
sessions = ['func1','func2']
nameOflevel1Out = 'results/level1_output'
nameOfWorkingdir = '/results/workingdir'

# Tell freesurfer what subjects directory to use
subjects_dir = experiment_dir + '/freesurfer_data'
fs.FSCommand.set_default_subjects_dir(subjects_dir)
```

Those specification mean that the name of the subjectfolders are "subject1", "subject2" and "subject3" and that each of those folders contain a "func1.nii" and "func2.nii" file which represents the nifti files for the first and the second functional run. But the exact structure will be defined with the datagrabber node.

4.2.3 Define a pipeline for the preprocess

We first define the preprocess pipeline with a node for slicetiming, realingment, artifaction detection, bbregister and smoothing.

```
#Initiation of the preprocess workflow
preproc = pe.Workflow(name='preproc')

#Node: Slicetiming
sliceTiming = pe.Node(interface=spm.SliceTiming(), name="sliceTiming")
sliceTiming.inputs.num_slices = 28
sliceTiming.inputs.time_repetition = 2.0
sliceTiming.inputs.time_acquisition = 2. - 2./28
sliceTiming.inputs.slice_order = range(1,28+1)      #for bottom up slicing
#sliceTiming.inputs.slice_order = range(28,0,-1)    #for top down slicing
sliceTiming.inputs.ref_slice = 1

#Node: Realign - for motion correction and to register all images to the mean image
realign = pe.Node(interface=spm.Realign(), name="realign")
realign.inputs.register_to_mean = True

#Node: Artifact Detection - to determine which of the images in the functional
#      series are outliers based on deviations in intensity or movement.
art = pe.Node(interface=ra.ArtifactDetect(), name="art")
art.inputs.norm_threshold       = 0.5
art.inputs.zintensity_threshold = 3
art.inputs.mask_type            = 'file'
art.inputs.parameter_source     = 'SPM'

#Node: BBRegister - to coregister the mean functional image generated by realign
#      to the subjects' surfaces.
bbregister = pe.Node(interface=fs.BBRegister(),name='bbregister')
bbregister.inputs.init = 'fsl'
bbregister.inputs.contrast_type = 't2'

#Node: Smooth - The volume smoothing option performs a standard SPM smoothing
volsmooth = pe.Node(interface=spm.Smooth(), name = "volsmooth")
volsmooth.inputs.fwhm = 6
```

```
#Node: FreeSurferSource - The get specifc files from the freesurfer folder
fssource = pe.Node(interface=nio.FreeSurferSource(),name='fssource')
fssource.inputs.subjects_dir = subjects_dir

#Node: Binarize -  to binarize the aseg file for the dilation
binarize = pe.Node(interface=fs.Binarize(),name='binarize')
binarize.inputs.min = 0.5
binarize.inputs.out_type = 'nii'

#Node: DilateImage - to dilate the binarized aseg file and use it as a mask
dilate = pe.Node(interface=math.DilateImage(),name='dilate')
dilate.inputs.operation = 'max'
dilate.inputs.output_type = 'NIFTI'

#Connect up the preprocessing components
preproc.connect([(sliceTiming, realign,[('timecorrected_files', 'in_files')]),
                 (realign, bbregister,[('mean_image', 'source_file')]),
                 (realign, volsmooth, [('realigned_files', 'in_files')]),
                 (realign, art,[('realignment_parameters','realignment_parameters'),
                                ('mean_image', 'mask_file'),
                                ]),
                 (volsmooth,art,[('smoothed_files','realigned_files'),
                                 ]),
                 (fssource, binarize, [('aseg','in_file')]),
                 (binarize, dilate,[('binary_file','in_file')]),
                 (realign, art,[('realignment_parameters','realignment_parameters'),
                                ('mean_image','mask_file'),
                                ]),
             ])
```

Note: If you are wondering how we knew which parameters to specify and which connections to establish. It is simple: Define or connect all mandatory inputs for each node. All the other optional inputs can be defined as you please and more importantly as your model demands. For more informations go to Interfaces and Algorithms.

4.2.4 Define a pipeline for the volume analysis

We than define the pipeline for the volume analysis with a node for model specification, first level design, parameter estimation and contrast estimation.

```
#Initiation of the volume analysis workflow
volanalysis = pe.Workflow(name='volanalysis')

#Node: SpecifyModel - Generate SPM-specific design information
modelspec = pe.Node(interface=model.SpecifySparseModel(), name= "modelspec")
modelspec.inputs.input_units             = 'secs'
modelspec.inputs.time_repetition         = 8.
modelspec.inputs.high_pass_filter_cutoff = 128
modelspec.inputs.model_hrf = True
modelspec.inputs.scale_regressors = True
modelspec.inputs.scan_onset = 4.
modelspec.inputs.stimuli_as_impulses = True
modelspec.inputs.time_acquisition = 2.
modelspec.inputs.use_temporal_deriv = False
modelspec.inputs.volumes_in_cluster = 1

#Node: Level1Design - Generate a first level SPM.mat file for analysis
level1design = pe.Node(interface=spm.Level1Design(), name= "level1design")
level1design.inputs.bases              = {'hrf':{'derivs': [0,0]}}
#level1design.inputs.bases              = {'fir':{'length':3, 'order' : 1}}
level1design.inputs.timing_units       = 'secs'
level1design.inputs.interscan_interval = modelspec.inputs.time_repetition
```

```
#Node: EstimateModel - to determine the parameters of the model
level1estimate = pe.Node(interface=spm.EstimateModel(), name="level1estimate")
level1estimate.inputs.estimation_method = {'Classical' : 1}

#Node: EstimateContrast - to estimate the first level contrasts we define later
contrastestimate = pe.Node(interface = spm.EstimateContrast(), name="contrastestimate")

#Connect up the volume analysis components
volanalysis.connect([(modelspec,level1design,[('session_info','session_info')]),
                     (level1design,level1estimate,[('spm_mat_file','spm_mat_file')]),
                     (level1estimate,contrastestimate,[('spm_mat_file','spm_mat_file'),
                                                       ('beta_images','beta_images'),
                                                       ('residual_image',
                                                        'residual_image')]),
                     ])
```

4.2.5 Define a framework workflow that contains the preprocess and the volume analysis

As we planed at the beginning we now want to integrate those two pipelines into a bigger framework workflow and want to add an inputnode that feeds these pipelines with parameters.

```
#Initiation of the framework workflow
frameflow = pe.Workflow(name='frameflow')

#Node: Inputnode - For this workflow the only necessary inputs are the functional
#      images, a freesurfer subject id corresponding to recon-all processed data,
#      the session information for the functional runs and the contrasts to be evaluated.
inputnode = pe.Node(interface=util.IdentityInterface(fields=['func','subject_id',
                                                             'session_info','contrasts']),
                    name='inputnode')

#Connect up the components into an integrated workflow.
frameflow.connect([(inputnode,preproc,[('func','sliceTiming.in_files'),
                                       ('subject_id','bbregister.subject_id'),
                                       ('subject_id','fssource.subject_id'),
                                       ]),
                   (inputnode, volanalysis,[('session_info','modelspec.subject_info'),
                                            ('contrasts','contrastestimate.contrasts'),
                                            ]),
                   (preproc, volanalysis, [('realign.realignment_parameters',
                                            'modelspec.realignment_parameters'),
                                           ('volsmooth.smoothed_files',
                                            'modelspec.functional_runs'),
                                           ('art.outlier_files',
                                            'modelspec.outlier_files'),
                                           ('dilate.out_file','level1design.mask_image'),
                                           ])
                   ])
```

4.2.6 Define infosource, datagrabber and datasink

We now have to create the infosource that defines the subjectlist, the datagrabber that grabbs the input and the datasink which defines where we want to store the important output at.

```
#Node: Infosource - we use IdentityInterface to creat our own node, to specify
#      the list of subjects the pipeline should be executed on
infosource = pe.Node(interface=util.IdentityInterface(fields=['subject_id']),
                     name="infosource")
infosource.iterables = ('subject_id', subjects)
```

```
#Node: DataGrabber - To grab the input data
datasource = pe.Node(interface=nio.DataGrabber(infields=['subject_id'],
                                               outfields=['func', 'struct']),
                     name = 'datasource')

#Define the main folder where the data is stored at and define the structure of it
datasource.inputs.base_directory = experiment_dir
datasource.inputs.template = 'data/%s/%s.nii'

info = dict(func=[['subject_id', sessions]],
            struct=[['subject_id','struct']])

datasource.inputs.template_args = info

#Node: Datasink - Create a datasink node to store important outputs
datasink = pe.Node(interface=nio.DataSink(), name="datasink")
datasink.inputs.base_directory = experiment_dir

#Define where the datasink input should be stored at
datasink.inputs.container = nameOflevel1Out
```

4.2.7 Define contrasts and model specification

We now set up the model specific components like contrasts, names of conditions, parametric modulators onset, duration of a trial etc.

```
#Names of the conditions
namesOfConditions = ['basic','condition1','condition2','condition3']

#Define different contrasts
cont1 = ('basic vs. conditions','T', namesOfConditions, [3,-1,-1,-1])
cont2 = ('all vs. condition1',  'T', namesOfConditions, [0,1,0,0])
cont3 = ('all vs. condition2',  'T', namesOfConditions, [0,0,1,0])
cont4 = ('all vs. condition3',  'T', namesOfConditions, [0,0,0,1])
cont5 = ('session1 vs session2','T', namesOfConditions, [1,1,1,1],[1,-1])

#store all contrasts into a list...
contrasts = [cont1, cont2, cont3, cont4, cont5]

#...and feed those contrasts to the inputnode filed 'contrasts'
frameflow.inputs.inputnode.contrasts = contrasts

#Function: Subjectinfo - This function returns subject-specific information about
#          the experimental paradigm. This is used by the SpecifyModel function
#          to create the information necessary to generate an SPM design matrix.
def subjectinfo(subject_id):

    from nipype.interfaces.base import Bunch

    namesOfConditions = ['basic','condition1','condition2','condition3']

    #Onset Times in seconds
    onsetTimes = [[1,10,42,49.6,66.1,74.1,97.6,113.6,122.2,130.2,137.2,153.7,169.2,
                   185.7,201.8,290.4,313.4,321.4,377.5,401.5,410,418.6,442.1,473.6],
                  [17.5,82.1,89.6,145.2,225.3,242.3,281.4,426.6],
                  [26,162.2,209.3,249.3,265.9,205.4,450.1,386],
                  [34,273.4,329.5,338.5,354,362,370,466.4]
                  ]

    #Define the parametric modulators
    para_modu = [None,
```

```
                Bunch(name=['target2','target3'], poly=[[1],[1]],
                      param = [[0,0,1,0,0,0,0,0],[0,0,0,0,1,0,0,1]]),
                Bunch(name=['target2','target3'], poly=[[1],[1]],
                      param = [[0,0,0,1,1,1,0,0],[1,0,0,0,0,0,1,1]]),
                Bunch(name=['target2','target3'], poly=[[1],[1]],
                      param = [[0,1,0,0,0,1,0,1],[0,0,0,0,1,0,1,0]]),
                ]

    output = []

    #We add the model specific parameters twice to the output list because we
    #have 2 functional runs which were performed identical.
    for r in range(2):
        output.append(Bunch(conditions=namesOfConditions,
                            onsets=onsetTimes,
                            durations=[[2] for s in namesOfConditions],
                            amplitudes=None,
                            tmod=None,
                            pmod=para_modu,
                            regressor_names=None,
                            regressors=None))
    return output  #this output will later be returned to inputnode.session_info
```

4.2.8 Define the meta pipeline

After setting up all the nodes, parameters and subpipelines, we now want to creat the pipeline that contains everything. The one that gets executed at the end.

```
#Initiation of the metaflow
metaflow = pe.Workflow(name="metaflow")

#Define where the workingdir of the metaflow should be stored at
metaflow.base_dir = experiment_dir + nameOfWorkingdir

#Connect up all components
metaflow.connect([(infosource, datasource,[('subject_id', 'subject_id')]),
                  (datasource,frameflow,[('func','inputnode.func')]),
                  (infosource,frameflow,[('subject_id','inputnode.subject_id'),
                                         (('subject_id', subjectinfo),
                                           'inputnode.session_info'),
                                         ]),
                  (frameflow,datasink,[('preproc.bbregister.out_reg_file',
                                        'bbregister'),
                                       ('volanalysis.contrastestimate.spm_mat_file',
                                        'vol_contrasts.@spm_mat'),
                                       ('volanalysis.contrastestimate.spmT_images',
                                        'vol_contrasts.@T'),
                                       ('volanalysis.contrastestimate.con_images',
                                        'vol_contrasts.@con'),
                                       ])
                  ])
```

Note: Some may wonder what the `@spm_mat`, `@T` and `@con` cause in the connection between the **frameflow** and the **datasink**. This specification means that all the `spm_mat_file`, `spmT_images` and `con_images` files all get saved into the same **datasink** folder with the name `vol_contrasts`. The `.@id` just specifies an identifier for the saving process.

4.2.9 Run the pipeline and generate the graph

Finally, after everything is set up correctly we can run the pipeline and let it draw the two graphs.

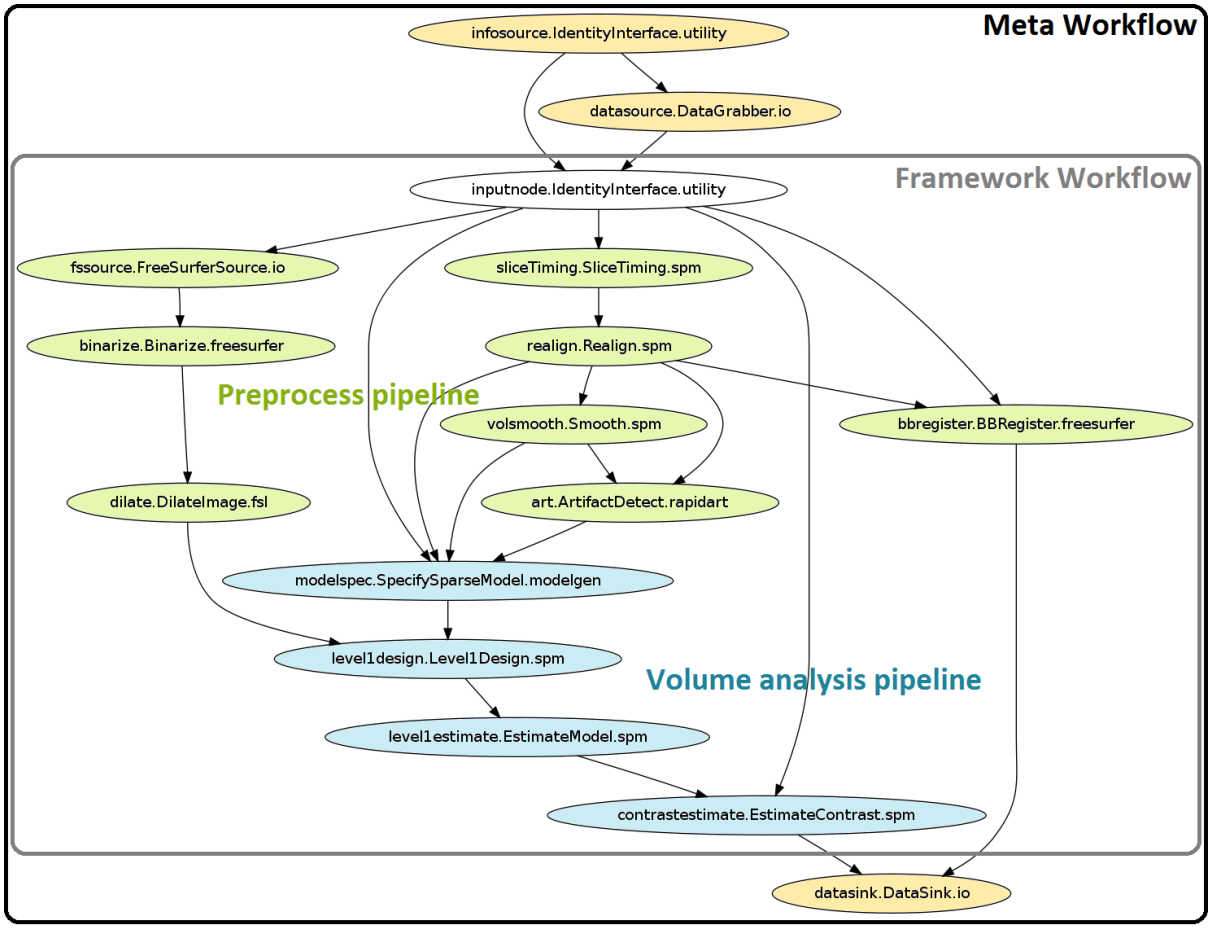
```
#Run the analysis pipeline and create the two graphs that visually represents the workflow.
metaflow.write_graph(graph2use='flat')
metaflow.run(plugin='MultiProc', plugin_args={'n_procs' : 2})
```

4.3 Visualization of the metaflow

Here are now the basic and the detailed graph of this example of our meta pipeline. Both graphs were created with the graph2use parameter set to 'flat'. The coloring was done additionally to underline the structure of the whole workflow.

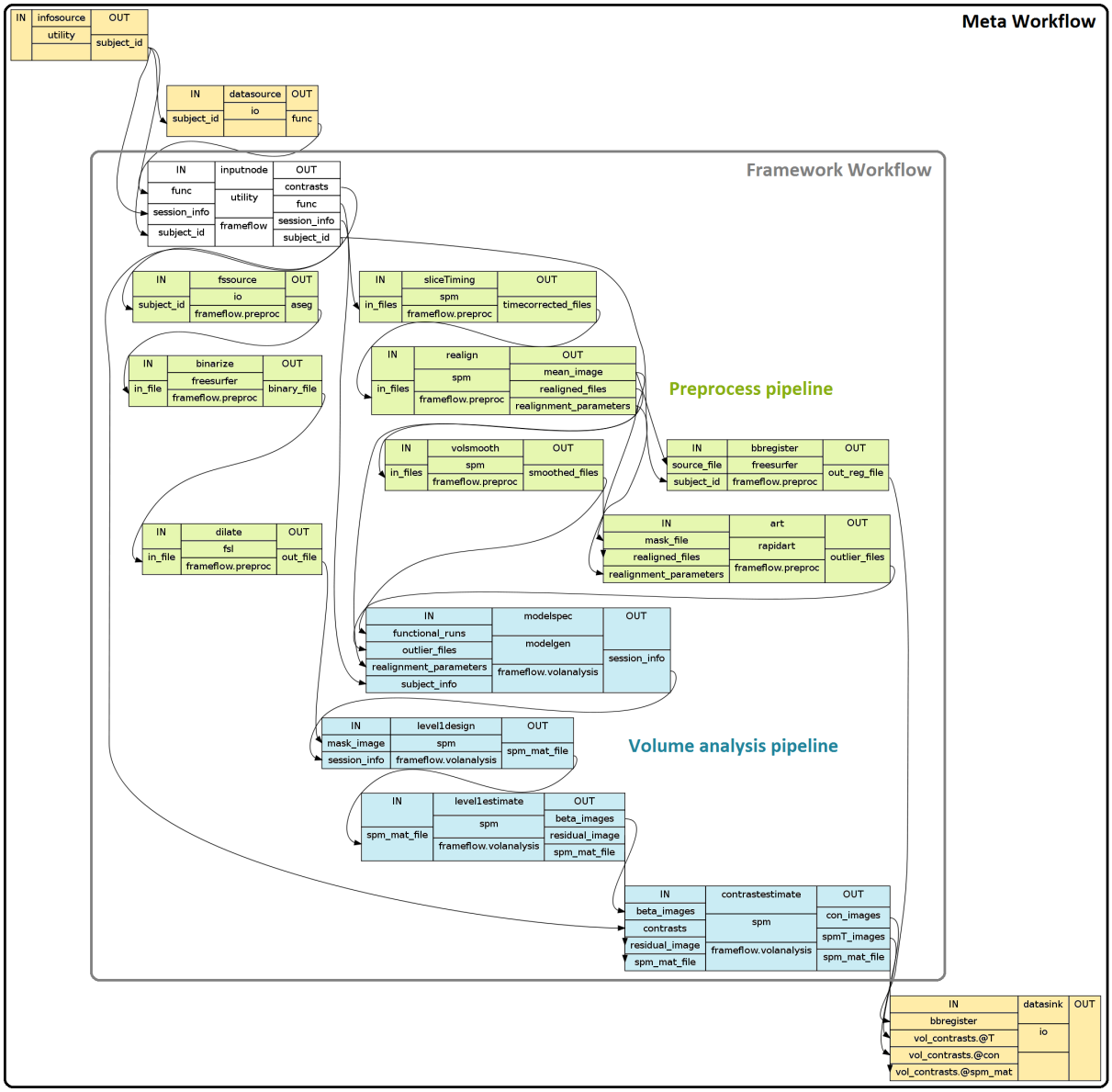
4.3.1 Basic graph

The basic graph showes how the nodes are connected to each other.



4.3.2 Detailed graph

The detailed graph shows how the different inputs and outputs of each individual nodes are connected to each other.


Meta Workflow
IN
infosource
utility
OUT
subject_id
IN
datasource
io
OUT
subject_id
func
Framework Workflow
IN
inputnode
utility
frameflow
OUT
contrasts
func
session_info
subject_id
func
session_info
subject_id
IN
fssource
io
frameflow.preproc
OUT
aseg
subject_id
IN
sliceTiming
spm
frameflow.preproc
OUT
timecorrected_files
in_files
IN
binarize
freesurfer
frameflow.preproc
OUT
binary_file
in_file
IN
realign
spm
frameflow.preproc
OUT
mean_image
realigned_files
realignment_parameters
in_files
Preprocess pipeline
IN
volsmooth
spm
frameflow.preproc
OUT
smoothed_files
in_files
IN
bbregister
freesurfer
frameflow.preproc
OUT
out_reg_file
source_file
subject_id
IN
dilate
fsl
frameflow.preproc
OUT
out_file
in_file
IN
art
rapidart
frameflow.preproc
OUT
outlier_files
mask_file
realigned_files
realignment_parameters
IN
modelspec
modelgen
frameflow.volanalysis
OUT
session_info
functional_runs
outlier_files
realignment_parameters
subject_info
IN
level1design
spm
frameflow.volanalysis
OUT
spm_mat_file
mask_image
session_info
Volume analysis pipeline
IN
level1estimate
spm
frameflow.volanalysis
OUT
beta_images
residual_image
spm_mat_file
spm_mat_file
IN
contrastestimate
spm
frameflow.volanalysis
OUT
con_images
spmT_images
spm_mat_file
beta_images
contrasts
residual_image
spm_mat_file
IN
datasink
io
OUT
bbregister
vol_contrasts.@T
vol_contrasts.@con
vol_contrasts.@spm_mat

CHAPTER

FIVE

UNDER CONSTRUCTION

Guide about a second level analysis, an example how to prepare your data, how to create some basic functional slice pictures and perhaps even how to use ANTS for normalization is still under cunstruction.

It all will follow as soon as possible.