Overview
The task at hand is the identification of one or more imaging artifacts that can affect the normal operation of microscopes at 3Scan. This is intentionally open-ended, as we are not looking for a "winning" algorithm, but instead we are interested to the approach and tools you use in tackling a real-world problem.

Submissions should include code, figures, and narrative available via git url from github or other similar providers. Enough information/documentation should be included to reproduce the results, this might include a list of dependant packages or instructions to run the associated code. We will begin reviewing submissions on the 27th of July, but the final date for consideration will be the 29th of July.

Task:
The goal of this exercise if to identify one or both of the following imaging artifacts:

- bubbles - These are caused by a pump/flow malfunction, and minimally interrupt imaging. They are non-destructive to data, however downstream processing can easily mistake them for other features. A couple examples of this defect can be see here, or here. In the data set associated with this, we see a some in slices numbered 15092 and 15172, but they are present in several others as well.
- rolling - Is a serious problem for imagery. When the slice curls back into the field of view it obstructs the imagery of the sample being sliced. This manifests at a large black blurry splotch coming in from the top of the image. A particularly bad bout of them can be seen here. In the dataset for this assignment, it can be seen in slice 15462 and two others that can be found by inspection.

Automatic identification of either/both of these artifacts should be a fairly easy undertaking with typical image-processing -or- machine learning techniques, and a few (2-4) hours of work at most.

Description of the Sample:
The data provided for this exercise come out of a sample containing ~16000 images of a mouse brain. A subsection of these data were selected 56 images total. These raw images were 4096 x 53203 pixels and have been cropped to 4096 x 8400 to minimize advantages gained from having a powerful computer or fast internet connection. This sampling represents 0.5% of our full data set of this sample.

The tissue is embedded in a polymer matrix, which surrounds the relevant anatomical features. The images contain a margin on both left and right sides which do not contain tissue or matrix being sectioned.

Description of the Data:

There are two forms of data provided for this exercise, the largest of which are the images. The images are a small subsection of our slice images. Their filename includes both the slice number in this sample and the slice unique id separated by an underscore. For example:

15022_14bcf5f8-d543-4c41-bb28-436e0f8c68ed.png

is the 15022'th slice we have taken on this sample, and 14bcf5f8-d543-4c41-bb28-436e0f8c68ed is a uuid4 that will connect this image with the _id field in the metadata.

The metadata, named metadata.json, looks suspiciously like a MongoDB dump, and contains 56 records which tie directly to the second raw images themselves via the _id field. This metadata contains many useful fields as well as many irrelevant to this analysis. In particular, each record has a "properties" field which represents a machine snapshot taken at the time of slicing. Some other relevant information can be found in the "stats" field, which contains some image statistics about the parent image that this image was cropped from.

Bonus:

Following on that there are a number of other interesting things that might be able to do with this dataset, including:

- Identification of any process parameters correlated with the artifacts.
- tissue analysis/segmentation
- vessel identification/tracking

None of these are in any way mandatory, just possibly interesting!