# SOFTWARE

## Data Factory (DF)

### Feature Engineering

**Component:** Brain morphological features  
**Contributing task(s):** T8.3.11 Brain morphological features UCL  
**Description:** A privacy preserving approach for the generalised principal component analysis of large image datasets.

**Dependencies:**

|  |  |  |
| --- | --- | --- |
| DATA | Hospital | Lille Hospital, Tel Aviv Hospital, Milano Hospital, Freiburg Hospital, CHUV Hospital |
|  | Reference | TBI (Traumatic Brain Injury), PPMI (Parkinson's Progression Markers Initiative), ADNI |
| SOFTWARE | Data Factory (DF) | Data Anonymisation, Workflow Engine, Data Pipeline processes, Data Quality Processes, Data Storage |

**Releases:**

|  |  |
| --- | --- |
| Planned functionalities at M18 | Further testing and code optimisation of image factorisation method (handle missing data, make more probabilistic, make computationally stable, etc). |
| Planned functionality at M12 | Initial implementation of image factorisation method without distributed computing (MS126). |
| Planned functionalities at M24 | A functioning distributed implementation of the image factorisation method, with features obtained by the approach available for data mining. |

**User and Use cases:**

|  |  |
| --- | --- |
| User | Use cases |
| Neuroscientist Clinical Researcher Developer - Software Developer - Methods SP8 Platform Developer | Compute a PCA-like analysis over thousands of 3D patient scans distributed across several hospitals, in situations where some parts of the data may be missing. Principal components serve as features for data mining. |