The following paper was used to explore these subjects -- ><https://academic.oup.com/bioinformatics/article/36/1/1/5514039>

A significant improvement would be to make and manage an automatic job distribution system. This option has been previously explored but is not necessary for a basic Galaxy to operate. However, if the user amount increases and more resource-intensive tools are run, the instance RAM would not be enough for it to operate. Different types of job distribution systems could be implemented. Slurm was already used to schedule and cluster jobs. Slurm is a open-source cluster management and job scheduling system for large and small Linux clusters. The cluster workload manager has three key functions. First, it allocates (non-)exclusive access to compute nodes to users for some duration of time so they can run jobs. Second, it provides a framework for starting, executing, and monitoring different jobs on the allocated nodes. Finally, it arbitrates contention for resources by managing a queue of pending work. To improve job clustering and scheduling, Pulsar could be installed as a remote job system. Pulsar is a python server application of Galaxy that allows for the transfer of files to a remote job system. The output of the job would be returned by Pulsar afterward, eliminating the need for a shared file system.

Running Pulsar would be most advised in a multi-user environment like SURF. In DRE however, this would not be necessary yet. DRE uses Microsoft Azure desktop VMs to perform analysis. Those VMs do not allow for more than a few users. This would still make Slurm useful for clustering and job scheduling, but pulsar would not be necessary. The addition of the imaging analysis system CellProfiler would be important to install in DRE. As CellProfiler can be installed through Linux and Windows and does require a similar amount of resources as Galaxy, it would be possible to analyze imaged cardiac tissue in DRE. CellProfiler is commonly used to construct data pipelines for stained tissues. The program is simple to use and uses algorithms to quantitatively measure phenotypes from thousands of images automatically. Images made using fluorescence microscopy are measured by using either pre-made pipelines or step-by-step pipelines. The program pipelines can also be uploaded in Galaxy to use the resources efficiently. Having both CellProfiler and Galaxy installed would allow for Imaging data analysis in DRE for both new and more experienced researchers.

DRE is still in development and large improvements are expected to be made in the future by the anDREa consortium. SURF already has services available that would allow for restricted research-only access. This service is called Research Access Management (SRAM). It allows for sharing and accessing a specified environment, not necessarily present in SURF itself. This management system would be useful when collaborating with multiple institutes, as access can be given depending on the role of each researcher. Projects can be assessed privately, but data could be shared when applied for. This application would be redirected to the project’s owner and could contain a data request or pipelines used for data analysis. SRAM could also be attached to the SURF research drive or any other secure cloud storage. By using this management system for managing collaborations research could be done more efficiently through Galaxy. However, this depends on the GDPR-certification given to the Galaxy platform in the SURF environment. OpenID-connect could solve the issue of restricted access to Galaxy and is thus advised to be implemented. OIDC and the OAuth2.0 protocol allow for clients to be verified and identified by an authentication server. SURF also has a service that could provide this function for Galaxy called SURF-conext. This service is already used by various institutes including UMCU and Avans. Data encryption could be the final solution to securing Galaxy for the analysis of controlled data. Some methods allow for data to be analyzed whilst still being encrypted. Homomorphic encryption is a method that uses computation to do exactly that. see figure 4.7. The output of the analysis would be sent encrypted to the user or collaborator. Metadata and raw data would not have to be decrypted thanks to this process, only the job output would. Combining homomorphic encryption with SRAM would be considered securely enough for controlled data analysis whilst still implementing the FAIR principles. It should be possible to do this in DRE too, but this option has not been explored yet. The current version of Galaxy would not allow for data to be analyzed while encrypted, but this could be changed in the future.