Assumption: Rubric to be used to assist a researcher in determining what data or software should be deposited in a FAIR aligned repository to communicate knowledge.

Simulation / Experiment Descriptors			**Simulation / Experiment Descriptor Classes**			
			SAVE MORE OUTPUT		SAVE LESS OUTPUT	
Descriptor	**Descriptor definition**		Class 1	Class 2	Class 3	**Theme**
Model/Code Availability/Ease of use	How accessible is this particular version of the model/code? How often does the code version change? Ease of software installation, setup, etc. IP barriers, embargo periods for new model development?		Difficult to acquire & manage	Model is shareable, but specific changes were implemented that make it unique.	Validated version of a highly accessible model was used/. Easy to install and run on many environments	Accessibility
Platform/System Availability	How specialized the platform needed is (particular hardware, compilers, source code needed)		Requires resources that are more difficult to get access to. Could be scale of resources or type. E.g. general desktop computing vs specific HPC.		Does not require special hardware resources to run	Accessibility
where/how was this run?	cloud vs. server (computational efficiency)		Cloud storage might be cheap, so can save more output with less cost issues		If cloud egress costs are high and cloud storage costs are high	Accessibility
Model Re-usability (setup etc)	Ease of software installation, setup, etc.		Greater difficulty means more to save, continual evolution of the underlying system but containerization may change this		Easy means little data to save	Accessibility
Human Effort	Person-hours required to reproduce dataset		Significant time & expertise required to replicate simulation. Likely will require contact with & guidance from original data producer(s).		Trivial effort required to replicate simulation for most end users.	Accessibility
Simulation Inputs	How much effort is it to get and manage all the inputs used by the simulation?		If inputs are difficult to acquire & manage, retaining output lowers burden for others who might want to re-run model or use outputs.		Easy to acquire & manage	Accessibility

Assumption: Rubric to be used to assist a researcher in determining what data or software should be deposited in a FAIR aligned repository to communicate knowledge.

Simulation / Experiment Descriptors			**Simulation / Experiment Descriptor Classes**			
			SAVE MORE OUTPUT		SAVE LESS OUTPUT	
Descriptor	**Descriptor definition**		Class 1	Class 2	Class 3	**Theme**
Output Usability	How easy is it to use the outputs outside the original context? Does it adhere to standards? What community standard? Are the metadata sufficient for someone else to understand the output.		Simulation outputs structured and aligned with community conventions		Simulation outputs provided in proprietary format. Obscure or undefined standards make usablility difficult.	Accessibility
Conformance to open or established standards	Ability of common software to read the data in future; ease with which data curators will be about to perform long-term preservation.		Community accepted standards compliance as a good base state minimum, assuming long-term stability in the standard; better adherence makes more data useful		Lack of conformance makes data far less useful and less reason to save	Accessibility
Archive Accessibility Provided by Data Curator	How easy is it to access the data? Can you bring analysis compute to the data?		Easily accessible compute co-located near the data. Data volume reduction capabilities provided to support targeted data transfer.		Data are only available for full file/granule download	Accessibility
Longevity of the technology	How long will the technology be usable, e.g. data formats, programming languages					Accessibility
Used in a "Highly Influential Scientific Assessment"	As defined, for example, by OMB "Revised Information Quality Bulletin for Peer Review" (2004 Apr 15): a scientific assessment whose "dissemination could have a clear and substantial impact on important public policies (including regulatory actions) or private sector decisions with a potential effect of more than $500 million in any one year or that the dissemination involves precedent setting, novel and complex approaches, or significant interagency interest."		Need to keep data for future fact checking.	Subset of data may enable fact checking, e.g. all data not needed	No, not used in any HISA.	Community Commitment
Part of Set? -Continuum of coordinated experiments to solo/smaller events	Is this model output part of a larger set, that is of value as a whole? (e. g., intercomparisons)		Yes, output is part of a larger set of related experiments.	Subsets more appropriate for some kinds of ensembles.	Full output may not need to be preserved.	Community Commitment

Assumption: Rubric to be used to assist a researcher in determining what data or software should be deposited in a FAIR aligned repository to communicate knowledge.						
Simulation / Experiment Descriptors			**Simulation / Experiment Descriptor Classes**			
			SAVE MORE OUTPUT		SAVE LESS OUTPUT	
Descriptor	**Descriptor definition**		Class 1	Class 2	Class 3	**Theme**
Benchmark	Is this potentially a benchmark for comparision?		Yes, output is a community reference dataset		Full output may not need to be preserved.	Community Commitment
Computational Cost	The economic cost (combination of run time and computer access costs) of completing the simulations		High computational cost and can only be produced with specialized platforms	Moderate computational cost, but access to needed platforms straightforward	Small computational cost with no special platform needs	Cost
Storage needs/costs	The volume of output that is actually generated by the model experiment or simulation.		Expensive storage can put a cap on how much data are saved			Cost
Data transfer cost	Limitations on transferring data		If you can use subsetting tools to reduce transfer cost			Cost
Archiving/Curation Cost	The economic cost of archiving the simulations - who will pay for it now and in the future? And for how long? Is there the availability of a budget, storage space, repo, etc. Willingness and means to curate, maintain, and migrate as needed, now and into the future. This includes the availability of a suitable repository within budget		If willingness and means exist, keeping more output is appropriate. Good organzation and control reduces human resource cost.	If willingness, but fewer means. (Potentailly keeping a documented workflow, notebooks and code, and subsets of data)	If no willingness and means, there is less value in keeping data.	Cost
Feature Reproducibility	The ability to reproduce specific (atmospheric) features (of given scale) within an acceptable statistical range of error.		Would be difficult to reproduce due to nonlinearity of phenomena being studied	Would be difficult to reproduce some feature details, but general findings are robust	No issues with reproducibility (could be due to study subject or to model packaging, e.g. containerization)	Reproducibility