x5gon-dev-db

X5GON Univ-Nantes | May 20, 2019

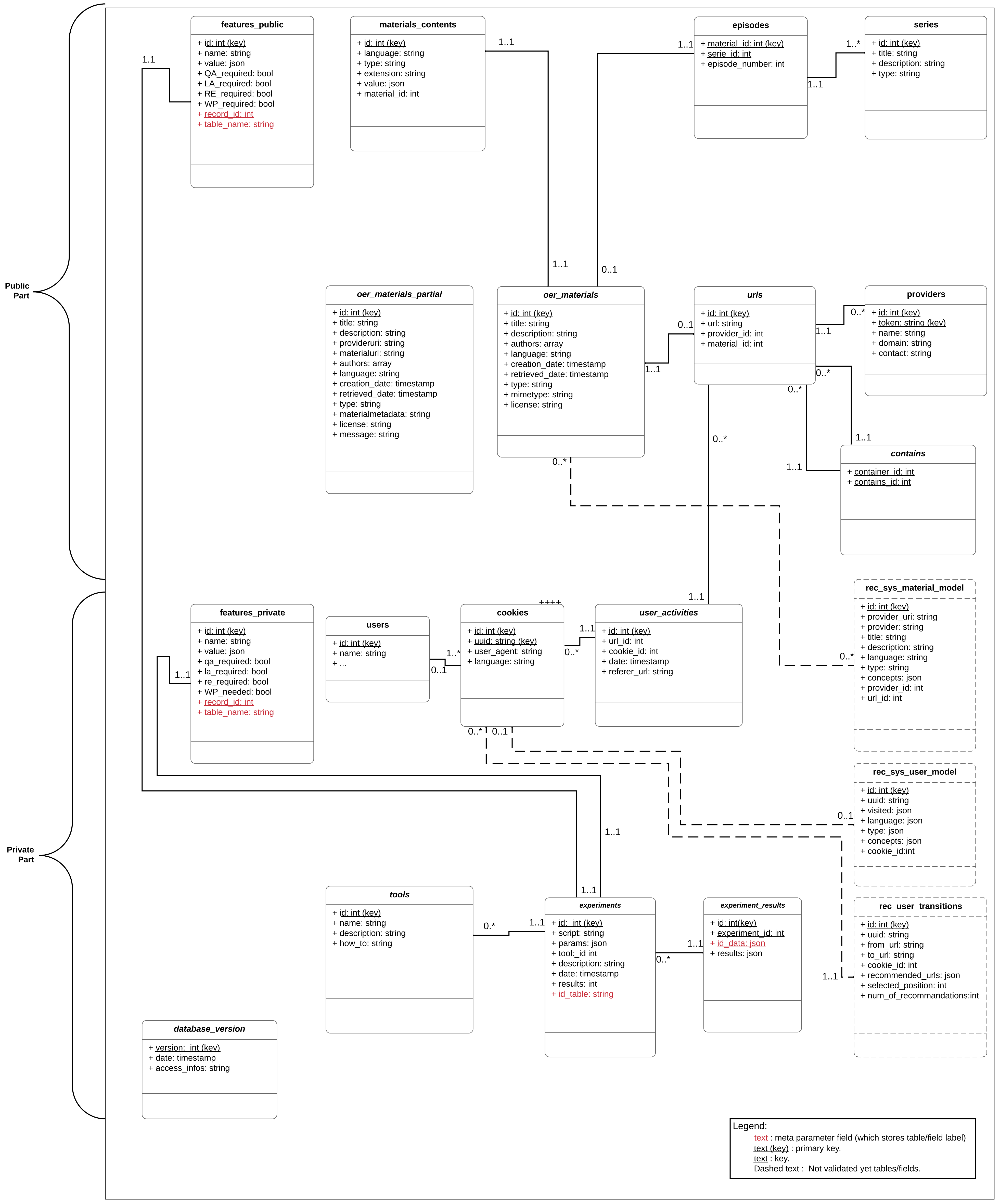

Table	Description	Attributes	Unanswered questions	Comments Any further questions or elements needed to be added ?
oer_materials *|oers|*	* Stores all OERs collected from OER repositories.	* Type: resource type (media, pdf,pptx ...)	*** What is exactly a resource** ? * A page containing many resources: which are considered as resources ? * Or only final urls are the resources ? * Current Model is flexible, any url can be an OER or not.	
oer_materials_partial *|oers|*	* Stores **not treated** collected OERs from OER repositories.			
providers *|providers|*	* Stores all providers : all registered OER repositories.	* Token: like an ID, is unique for the repositories , generated when provider enters consortium. * Domain: Url of repository * Contact: email of repository contact person.		
materials_contents *|oers|*	* Stores all resource content informations: transcriptions and available translations.	* language: content language. * value: content value. * type: content type: transcription, translation.... * extension: content extension.	* Content infos is in a standalone table for the size of transcription/translation infos.	* Note that the table can be merged with public_oers.
features_public	* Stores all WP final results meant to be public and shared with exeternal/internal parties through the 'External partners API'.	* value: metric value for a specific oer/user. * <WP/qa/la/re>_required: boolean representing whether the metric is used by WP or not.	* We think that the best solution to share WP metrics is to store it in DB then will be fetched by "DEV API": because will be a different ways to generate those metrics => better to homogenize.	* Note that the table can be merged with oer_materials.
features_private	* Stores all WP final results meant to be private and not shared with exeternal parties.	* value: metric value for a specific oer/user. * <WP/qa/la/re>_required: boolean representing whether the metric is used by WP or not.		
episodes <|oers|	* Stores the link between the OERs and the "Series".	* episode: the number of episode in the series.		* An OER appears in at most one series. * In some series, episodes are not numbered. In thses cases, "episode_number" are set to none.
series *|series|*	* Stores all series and their related informations.			
urls *|activitties|*	* Stores the pages visited in OER repository.	* material_id:link could be deduced either from "container_id".	* Table key is not clear: because records here are representing "accessed pages" ? * We have "visited_url" to make the link to "public_oers": we can remove "id_resource " ?	* For non-OER urls, no content information is kept.
contains *|urls|*	* Stores pairs (container_id--contains_id) where the webpages (url_container) links to webpages (url_contains).	* container_id: url contains 'contains_id' * contains_id		
cookies *|cookies|*	* Stores the different user (cookies)	* uuid: cookie id. * user_agent: agent infos. * language: user languages (settable preference).	* "browser_agent" and "language": are user preferences and could be changed , is it good to store it ? * **Geolocalisation / IP** infos are not stored ? * **User consent** is not stored ?	
users *|users|*				
user_activities *|activities|*	* Stores all users activities in OER repositories pages traced by the snippet.	* cookie_id: user cookie unique id.	* **what is exactly a resource** ? * How precise is the snippet ? * The visited_url can be an OER or not.	
tools *|tools|*	* Stores all development (intermediate) tools used and developed during our experiments.	* name: tool name. * description: how it works and what results provides. * how_to: how to execute it + some of execution examples.		
experiments *|experiments|*	* Stores all configurations "experiments/tests" .	* tool_id: tool id. * script: path to the executable of tool. * params: the specific parameters passed to the script. * date: execution timestamp. * id_table: "public_oers/ user_activity" on which data the script is executed.		
experiment_results *|experiments|*	* Stores all results for given "experiments/tests".	* id_data: "oer_id / user_activity /user": the result of the "script trial " on a specific "id_data". * results: possibly a json.		* **"id_data" :** a tool result can applied on a "user / oer" depending on tool aims. In both cases, this field references the concerned field of the table provided by "id_table" (experiment).
database_versions *|versions|*	* Stores all "dataset versions". This will be updated directly by the "versionning system". * The last record is the the actual version of the dataset.	* date: dataset timestamp. * access_infos: dataset storage and access infos ("Storage DB" and "Dev DB" dumps , files...)	* Which is better for the "versionning type": **Whithout versioning**: -Always we have only one dataset. -We can't detect anomalies in the dataset. +This is optimized in storage matter. **Periodic Versioning**: +We have many versions of working datasets. +We can detect and isolate anomalies. +- More or less optimized in storage matter. **RealTime Versioning**: -so many versions of working datasets. + Detect and isolate anomalies(more precisely) - NOT optimized in storage matter.	
oer_materials_partial	* Stores oer_materials partially crawled/treated by the pipeline.			
rec_sys_material_model	Coming soon ...			
rec_sys_user_model	Coming soon ...			
rec_user_transitions	* Stores users transition logs when passing within the recommanded list of resources.			