

Data model for Social Media Crawling

Webinar organised by the DW2020 synergy group

Presenter: Stelios Andreadis, ITI-CERTH


Wednesday 3 March, 2021

11.00-12.30 CET












Aim & Steps followed



Creation of a data model to organize elements of social media data and standardize their attributes and relations

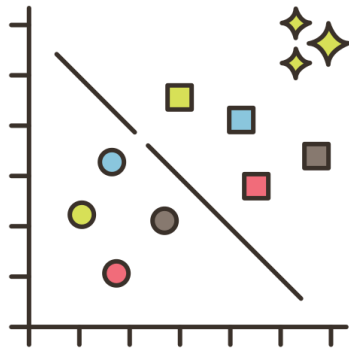
- *As generic as possible  Selection of most popular social media platforms that also offer an API to collect data (crawling)*
- *Collect most fundamental post & user attributes of each platform*
- *Comparative table – merge same attributes, cluster similar attributes*
- *Include potential analysis outcomes*
- *Design & Implementation of data model*

Comparative table

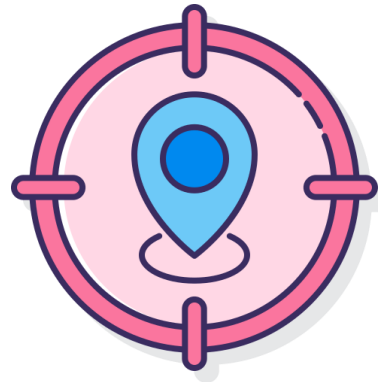
					
	user.id user.id_str user.name user.screen_name	from	username	snippet.channelId snippet.channelTitle	owner owner_name
	text	message	caption	snippet.title snippet.description	title
	created_at timestamp_ms	created_time updated_time	timestamp	snippet.publishedAt	date_upload date_taken last_update
	coordinates place	coordinates place	X	X	latitude longitude
	quote_count reply_count retweet_count favorite_count	shares reactions	X	statistics.viewCount statistics.likeCount statistics.dislikeCount statistics.favoriteCount statistics.commentCount	views
	entities.media.media_url_ https extended_entites.media.m edia_url_https	full_picture picture	media_url permalink	X	<i>can be synthesized</i>

etc.

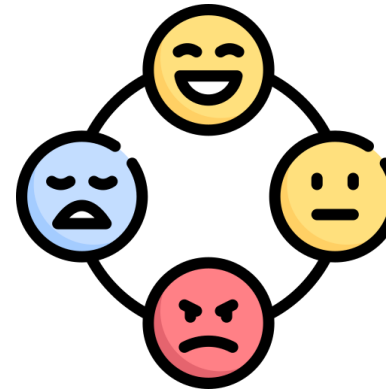
Potential analysis



Clustering



Geotagging

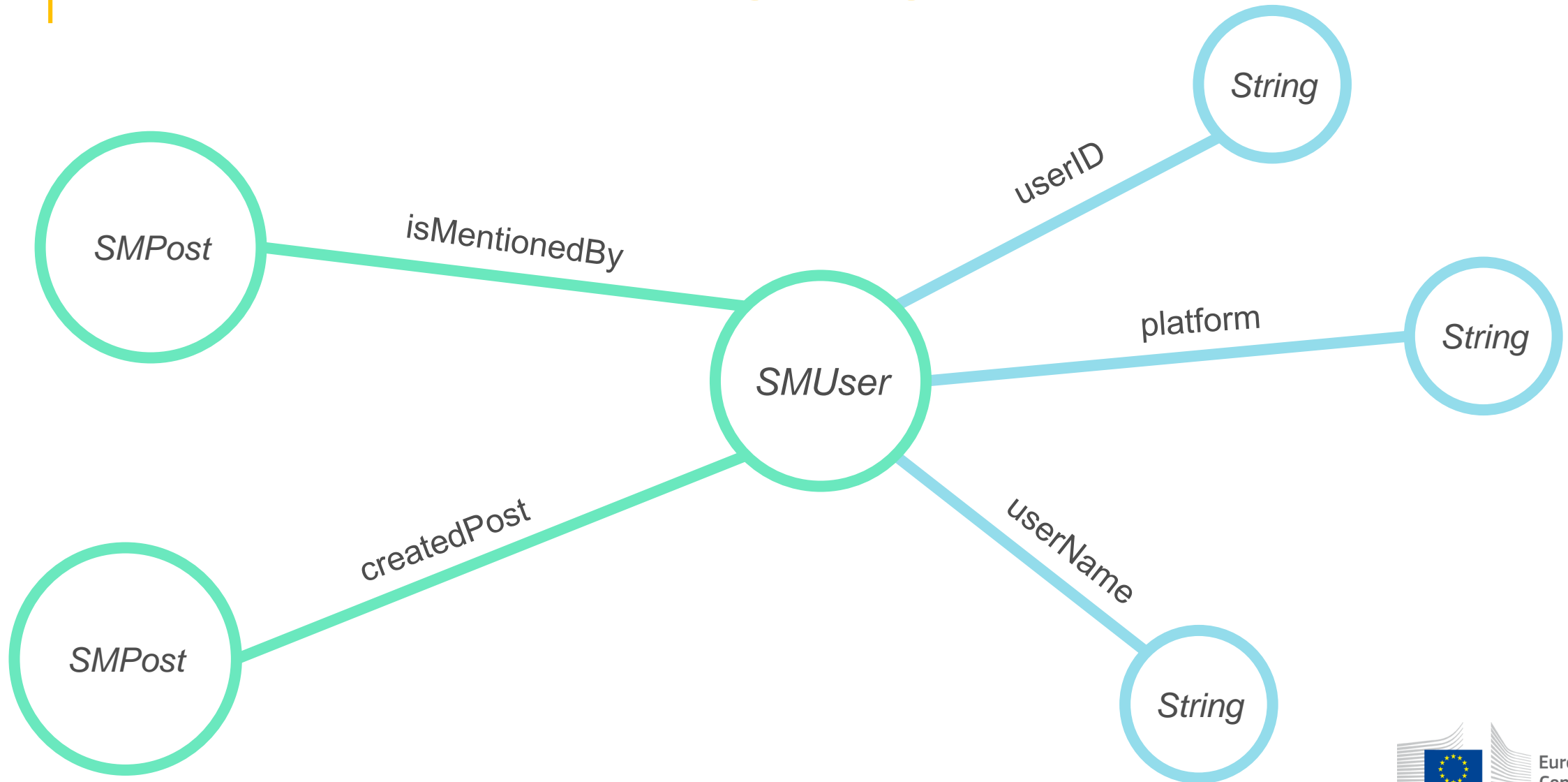


*Sentiment
analysis*

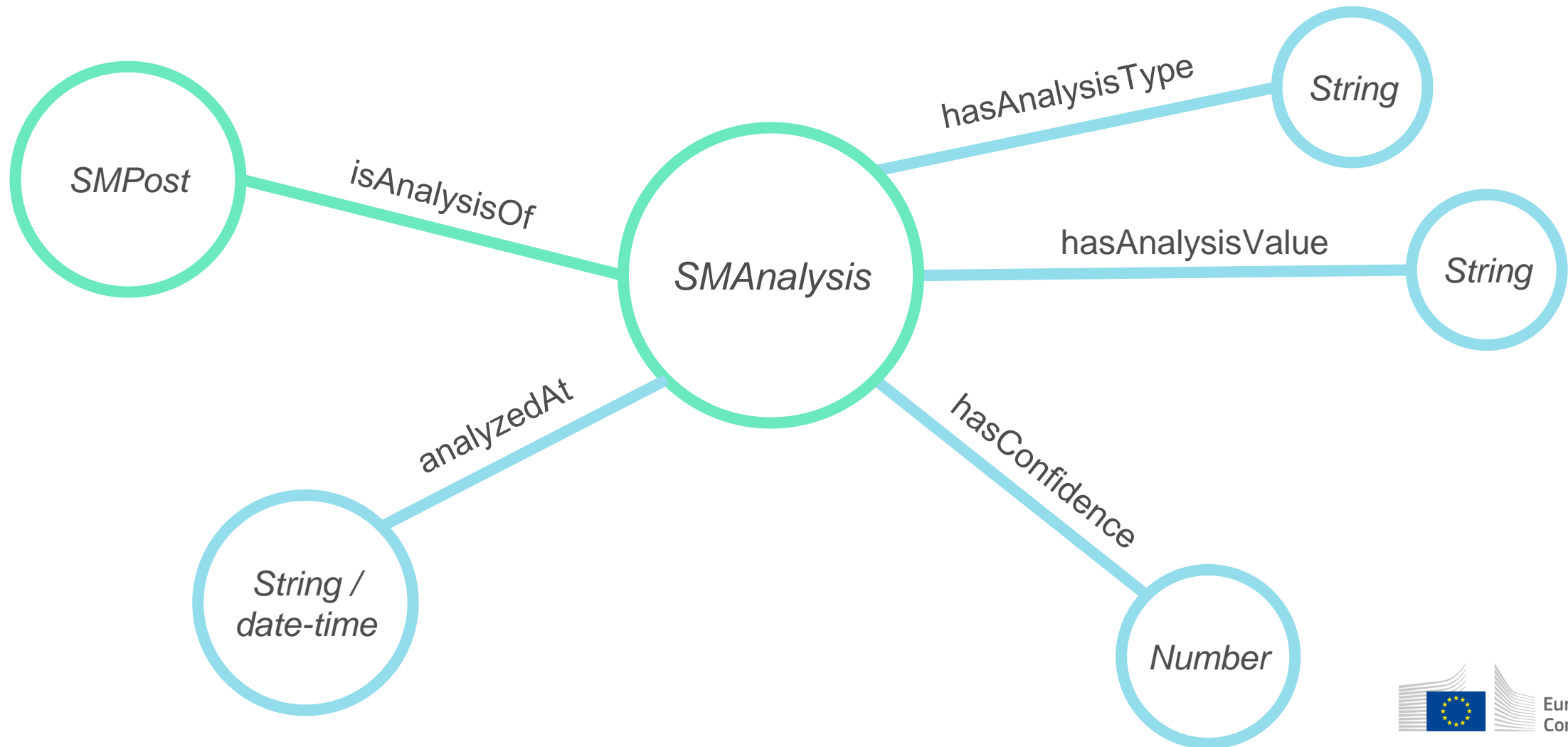


*Fake news
detection*

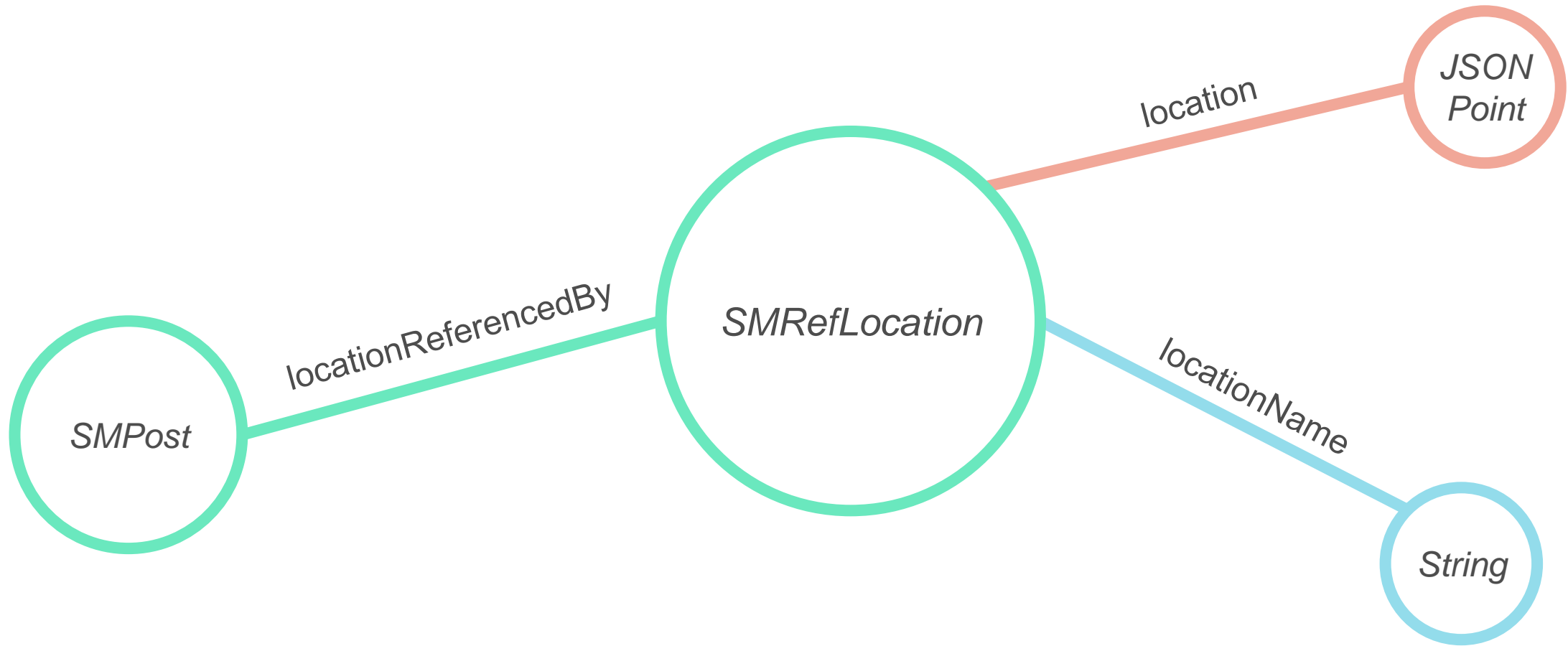
Data Model - SMUser



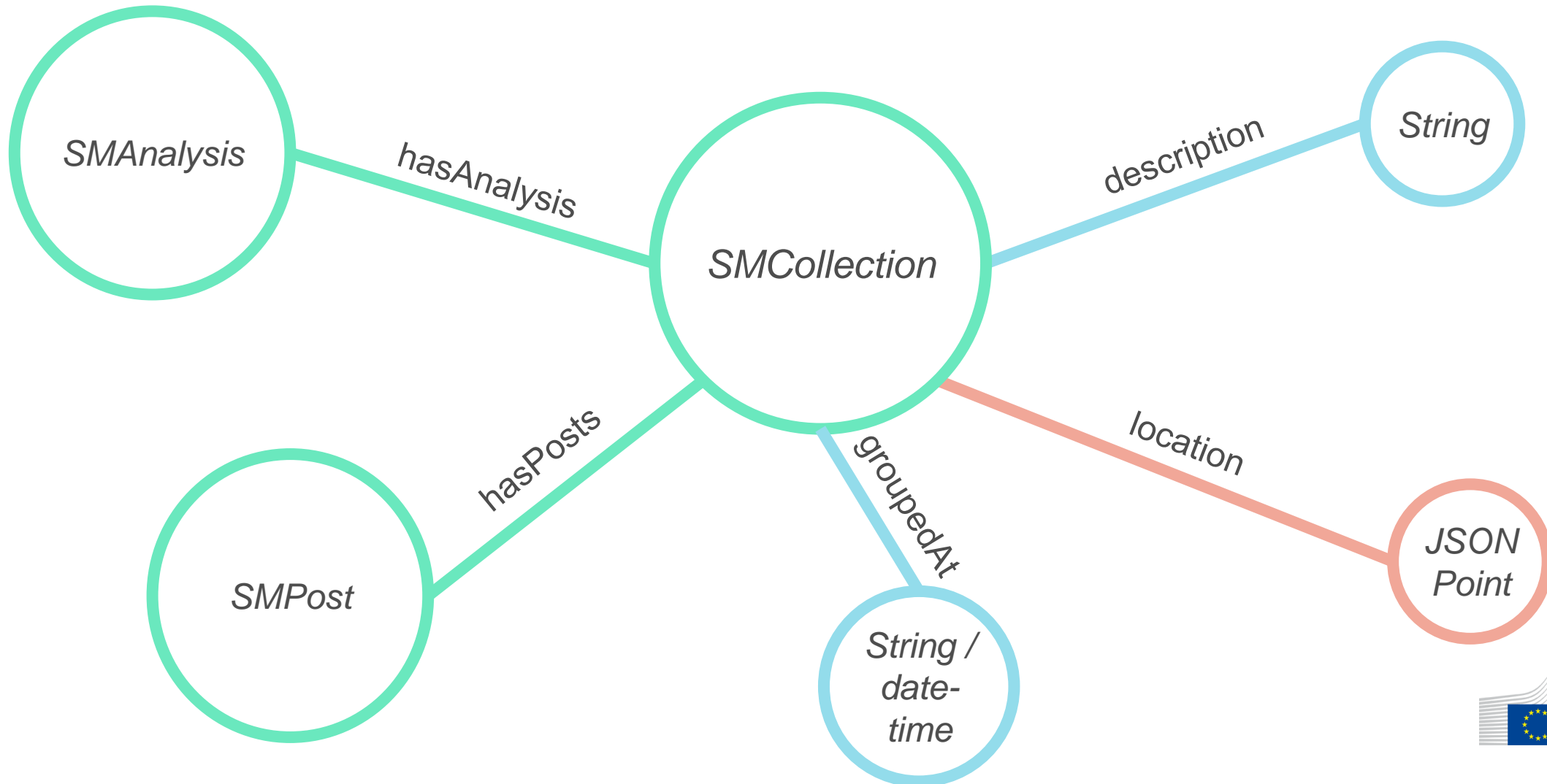
Data Model - SMAnalysis



Data Model - SMRefLocation



Data Model - SMCollection



Thank you for your attention

Stelios Andreadis, ITI-CERTH

 *andreadisst@iti.gr*