

Scientific Workflows with Knowledge Graphs

Authors
Yigit Ilk (yigit.ilk@tum.de)

Affiliations
Technical University of Munich, TUM school of Computation, Information and Technology
Bachelor Practical course in Data Engineering

Advisors
Prof. Dr. Maribel Acosta

This project aims to simplify the data flow between multiple projects and working locations in one large Research focus of Material Science: CRC:1625

Introduction

Communication between 16+ Projects and countless Researchers collaborating is no simple feat. That is why we want to create Workflows and better implementations and visuals of data flow from one source to another.

Approach

- Provide Workflows based on general user patterns
- Provide Dataflows based on general content
- Create code that can run in project management server to create graphs automatically using Python for backend and using Neo4J Graph Database for visuals

As CRC has a lot of researchers working under it, thus Our goal is to make a real-time status update platform, where all parties involved can see up-to-date information

Transferring Generalized Workflows that provide flow of information...

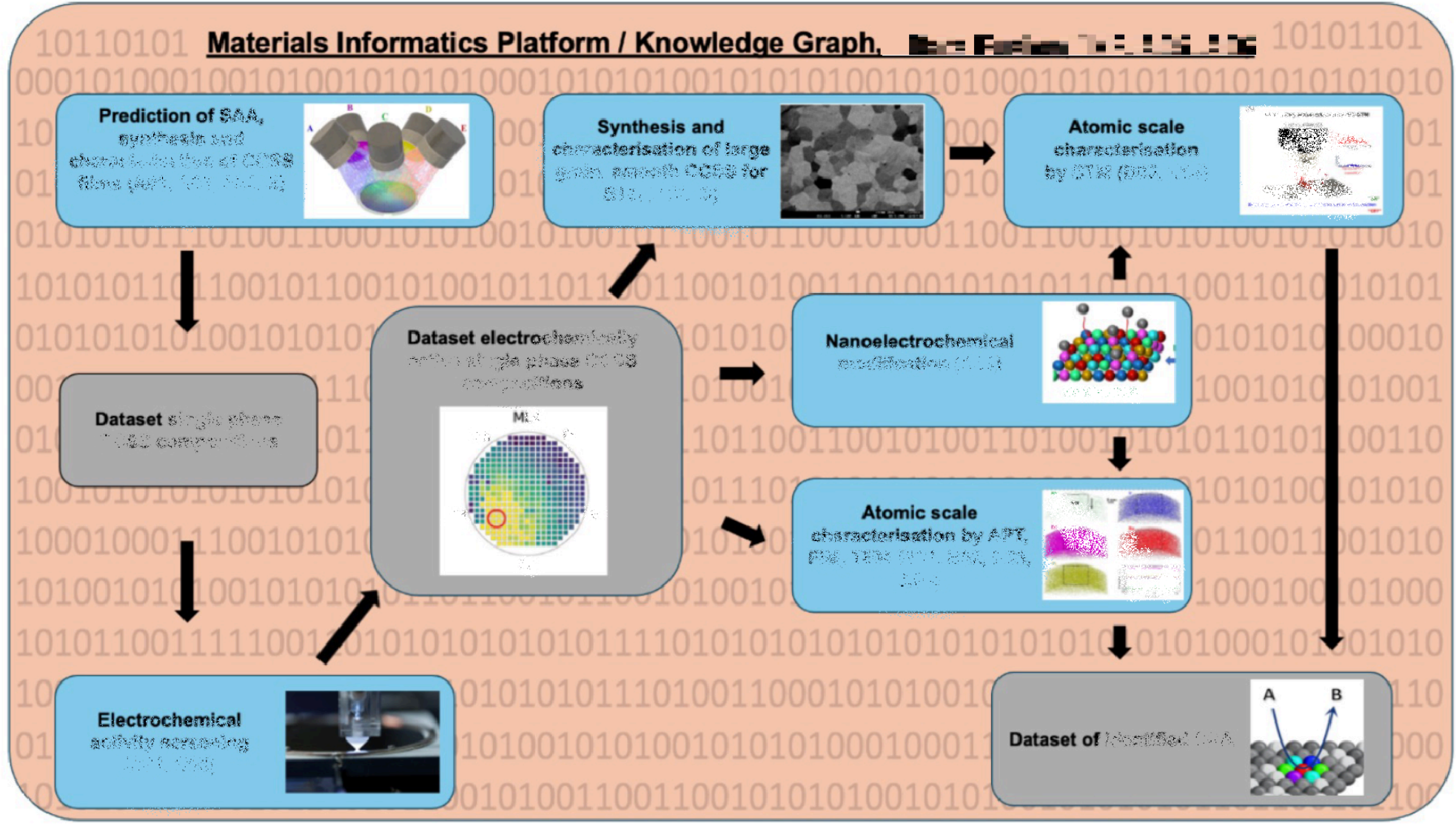
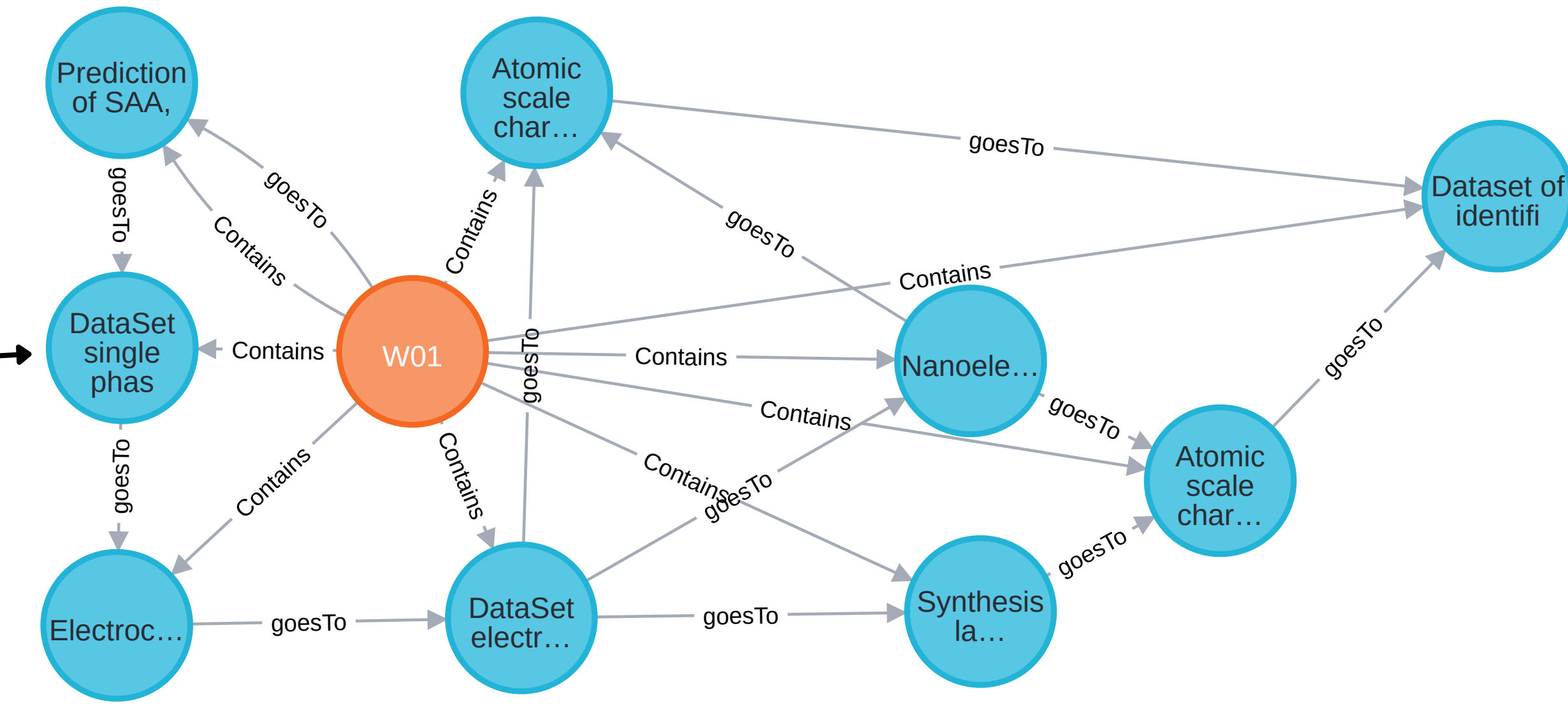


Figure 27: Exemplary workflow to identify SAA: Data-guided prediction-synthesis-characterisation (methods in blue, datasets in grey).

Into queryable and digitally easily modifiable graph databases that allow for system integrations



Usual Databases

SQL Query
SELECT h.HandoverId, u1.UserName as source, u1.Id as sourceId, u2.UserName as target, u2.Id as DestinationId, h.DestinationComments as comments, h.SampleObjectid as SampleId
FROM v.Handover h, v.Handover u1, v.Handover u2
WHERE h.HandoverId = u1.Objectid AND
h.DestinationUserid = u2.Id AND
o_createdBy = u1.Id;

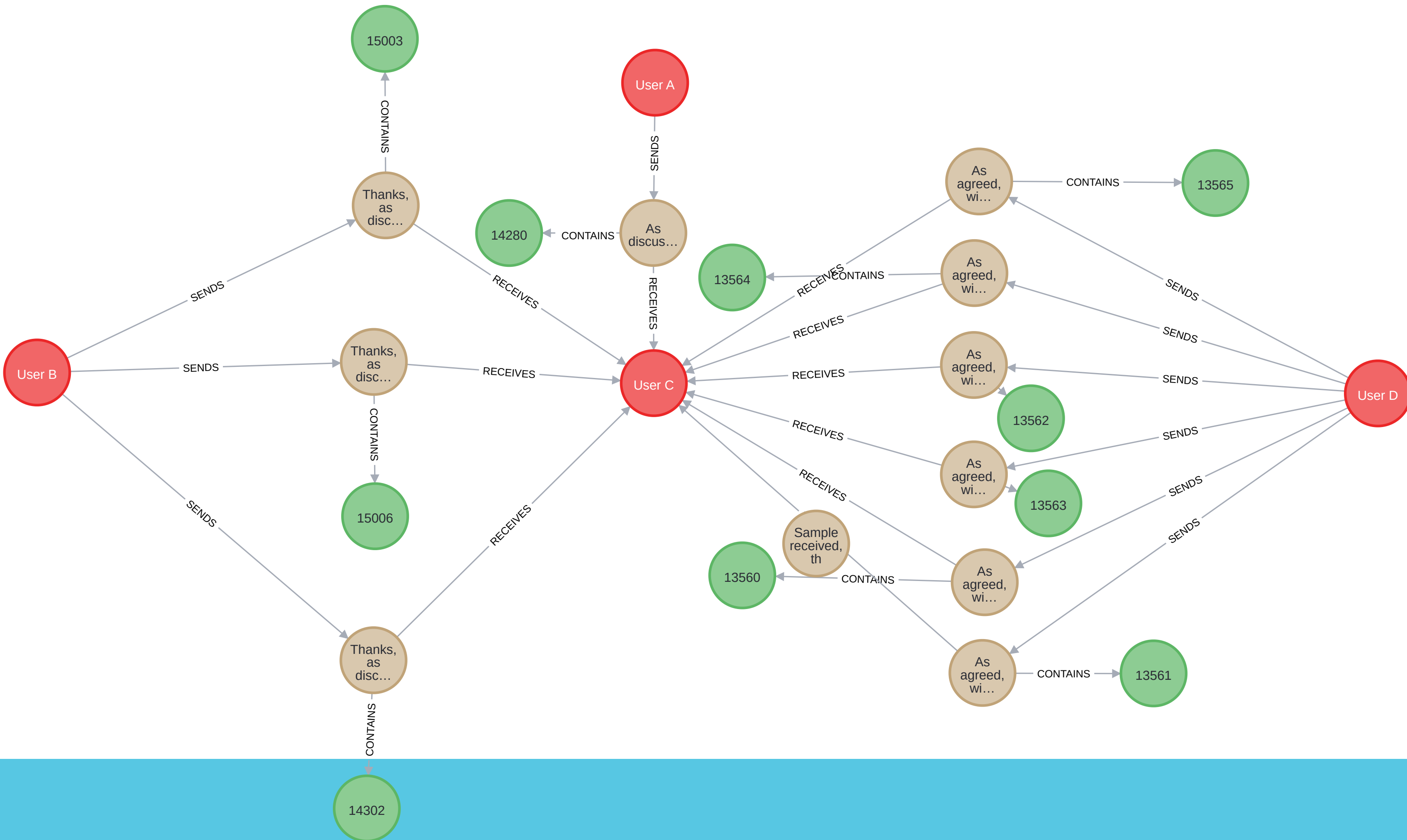
HandoverId	source	sourceId	target	DestinationId	comments	SampleId
1	User A	1	user C	3	As discus...	14280
2	User B	2	User C	3	Thanks, as disc...	15006
3	User D	4	User C	3	As agreed, wi...	13562

Instead of a fixed SQL Schema which requires background knowledge to query information and connect multiple dots together,

we can visualize even the querying of data flow for a better user experience and ease of data flow visualization connecting all the dots automatically.

Graph Databases

MATCH p=(u:User)-[]->(n:Handover)-[]->>() RETURN p



Results & Conclusion

We created easily viewable and queryable database nodes that are able to show the flow of data

We wrote backend Python Code in order to get data from the already existing user portal to create easily viewable data flow history

Showing that the data visualisation and understanding in science doesn't always have to be overcomplicated ;)

Thus enabling more efficient understanding of information flow between researchers in this project and helping them find necessary data faster.

Future Work

As we have access to the already existing user portal, with more researchers and more projects using it, we will be able to get more data:

-Using that, we want to create more extensive workflows from the data traffic users have and provide further suggestions for :

- *Which person the data they provide should be transmitted to
- *Which department is most likely required to process the data now before further use
- *let researchers know beforehand the status of a request before it is their time to work on the data

As this research is still in development stage, most of the data and information is omitted from this Poster.