

Clinical data to knowledge graph:

The MeDaX pipeline

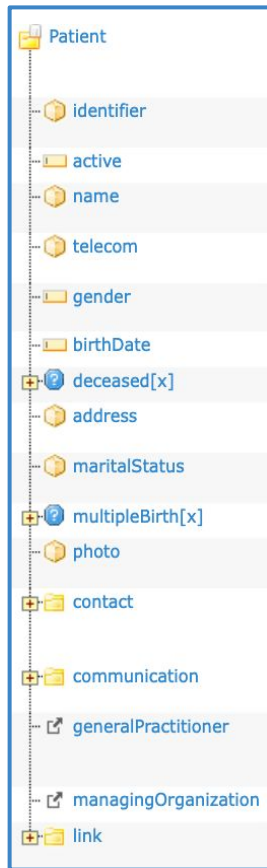
Ilya Mazein, Tom Gebhardt,
Lea Michaelis, Ron Henkel,
Benjamin Winter, Sebastian Berthe,
Dagmar Waltemath and Judith A.H. Wodke

Affiliations



FHIR data

- Format
- Structure
- File



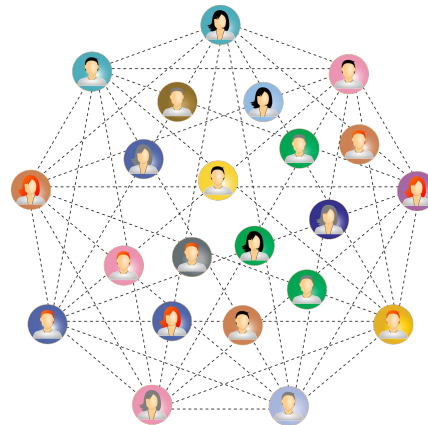
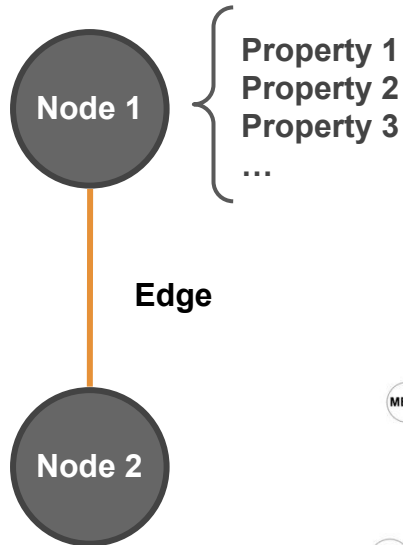
hl7.org/fhir/patient.html

```

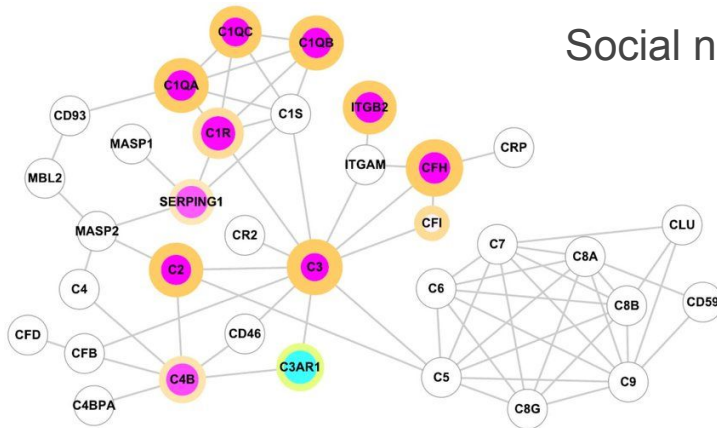
{
  "resourceType": "Patient",
  "id": "123456789",
  "meta": {
    "version": 1,
    "lastUpdated": "2014-08-01T12:00:00.000+01:00",
    "source": "http://hl7.org/fhir/StructureDefinition/Patient",
    "profile": "http://hl7.org/fhir/StructureDefinition/Patient"
  },
  "text": {
    "status": "complete",
    "div": "Patient information"
  },
  "identifier": [
    {
      "system": "http://hl7.org/fhir/StructureDefinition/Patient",
      "value": "123456789",
      "type": "id"
    }
  ],
  "name": [
    {
      "family": "Doe",
      "given": "John",
      "prefix": "Mr",
      "suffix": "Jr"
    }
  ],
  "telecom": [
    {
      "system": "phone",
      "value": "1234567890",
      "use": "home"
    }
  ],
  "gender": "male",
  "birthDate": "1970-01-01",
  "deceased": {
    "x": false
  },
  "address": [
    {
      "line": [
        "123 Main St",
        "Suite 100"
      ],
      "city": "Anytown",
      "state": "CA",
      "postalCode": "90210",
      "country": "US"
    }
  ],
  "maritalStatus": "M",
  "multipleBirth": {
    "x": false
  },
  "photo": [
    {
      "url": "http://example.com/patient-photo.jpg",
      "data": "data:image/jpeg;base64,..."
    }
  ],
  "contact": [
    {
      "name": "John Doe",
      "relationship": "self",
      "telecom": [
        {
          "system": "phone",
          "value": "1234567890",
          "use": "home"
        }
      ]
    }
  ],
  "communication": [
    {
      "language": "en",
      "text": "Patient information"
    }
  ],
  "generalPractitioner": [
    {
      "reference": "http://hl7.org/fhir/StructureDefinition/Practitioner",
      "type": "Practitioner"
    }
  ],
  "managingOrganization": [
    {
      "reference": "http://hl7.org/fhir/StructureDefinition/Organization",
      "type": "Organization"
    }
  ],
  "link": [
    {
      "target": "http://hl7.org/fhir/StructureDefinition/Patient",
      "relationship": "self"
    }
  ]
}
```



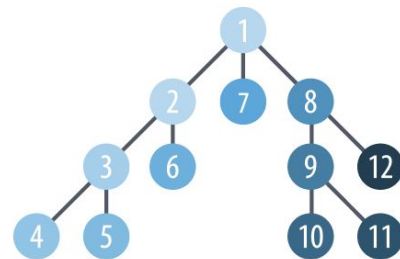
Graph databases



Social networks



Protein interactions

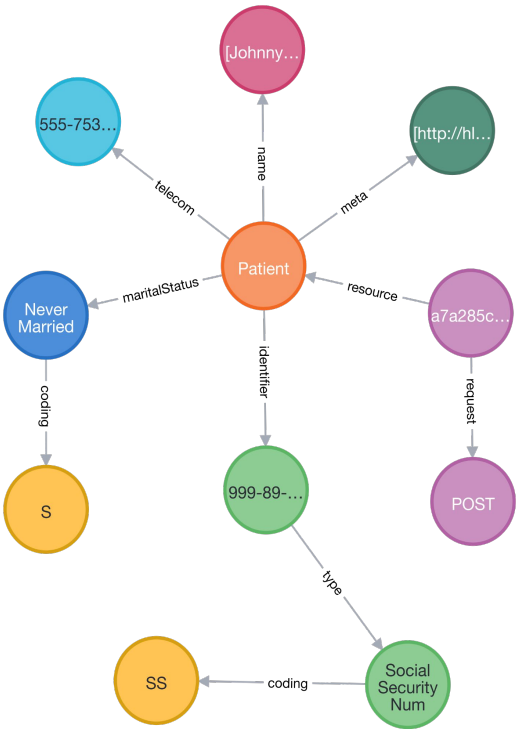


Traversal algorithms



Conversion

```
"entry": [ {
  "fullUrl": "urn:uuid:a7a285c0-4714-dd3c-4837-8719c9b67873",
  "resource": {
    "resourceType": "Patient",
    "id": "a7a285c0-4714-dd3c-4837-8719c9b67873",
    "meta": {
      "profile": [ "http://hl7.org/fhir/us/core/StructureDefinition/us-core-patient" ]
    },
    "identifier": [ {
      "type": {
        "coding": [ {
          "system": "http://terminology.hl7.org/CodeSystem/v2-0203",
          "code": "SS",
          "display": "Social Security Number"
        } ],
        "text": "Social Security Number"
      },
      "system": "http://hl7.org/fhir/sid/us-ssn",
      "value": "999-89-9526"
    } ],
    "name": [ {
      "use": "official",
      "family": "Schoen8",
      "given": [ "Johnny786", "Vince741" ]
    } ],
    "telecom": [ {
      "system": "phone",
      "value": "555-753-6560",
      "use": "home"
    } ],
    "gender": "male",
    "birthDate": "2021-05-22",
    "maritalStatus": {
      "coding": [ {
        "system": "http://terminology.hl7.org/CodeSystem/v3-MaritalStatus",
        "code": "S",
        "display": "Never Married"
      } ],
      "text": "Never Married"
    },
    "request": {
      "method": "POST",
      "url": "Patient"
    }
  }
}]
```

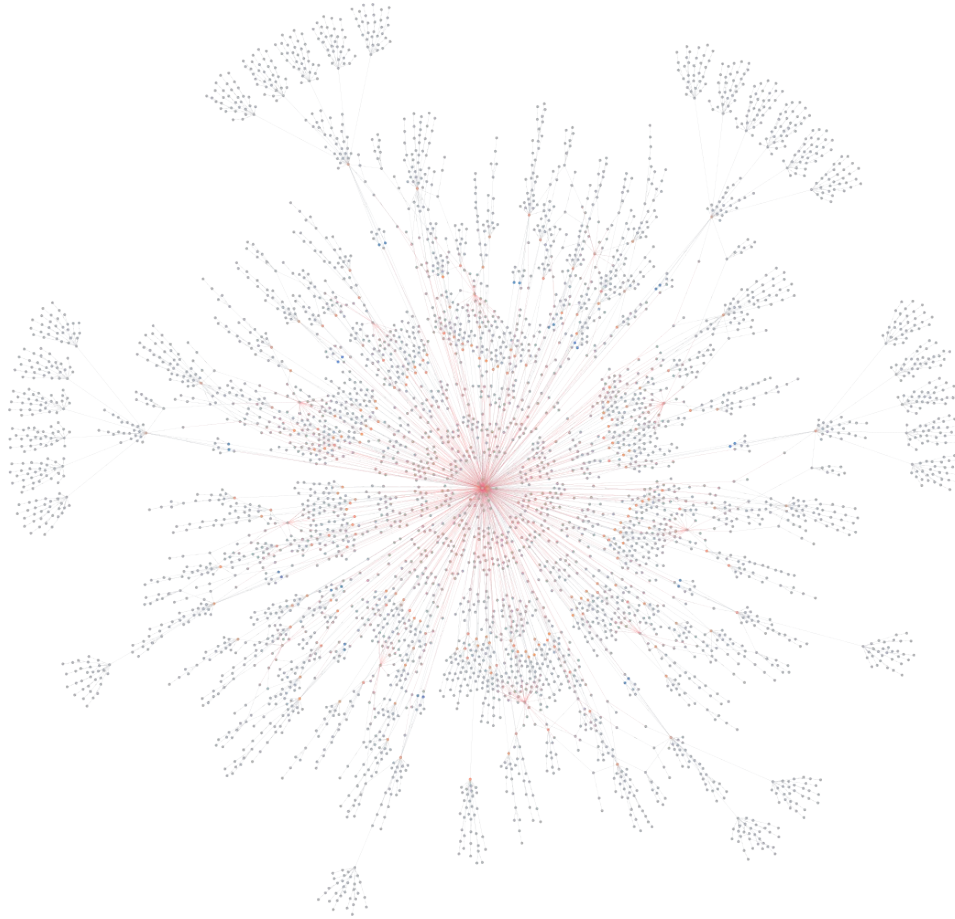


FHIR data

NetworkX graph



Conversion



157 Resource types



Generic conversion



Large graphs

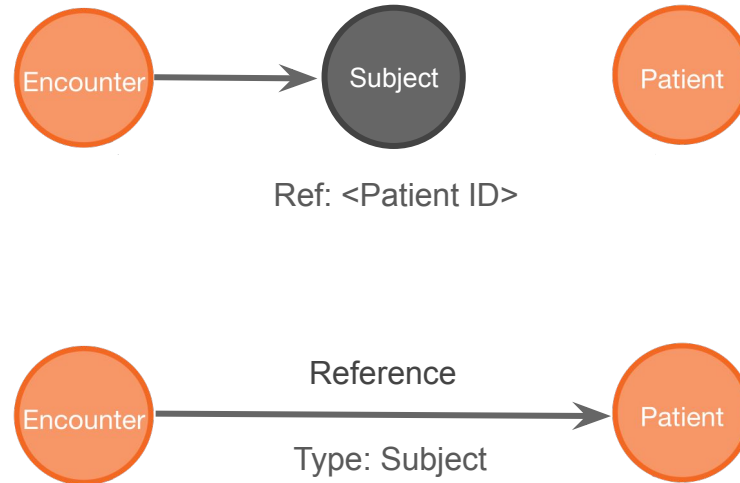


Optimisation

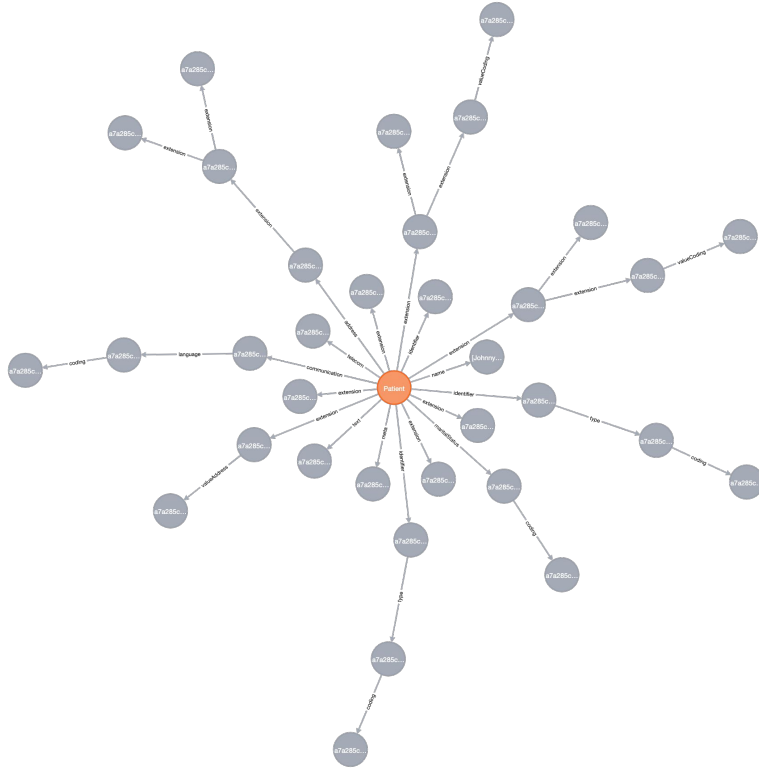


Optimisation

Reference path reduction strategy



Optimisation



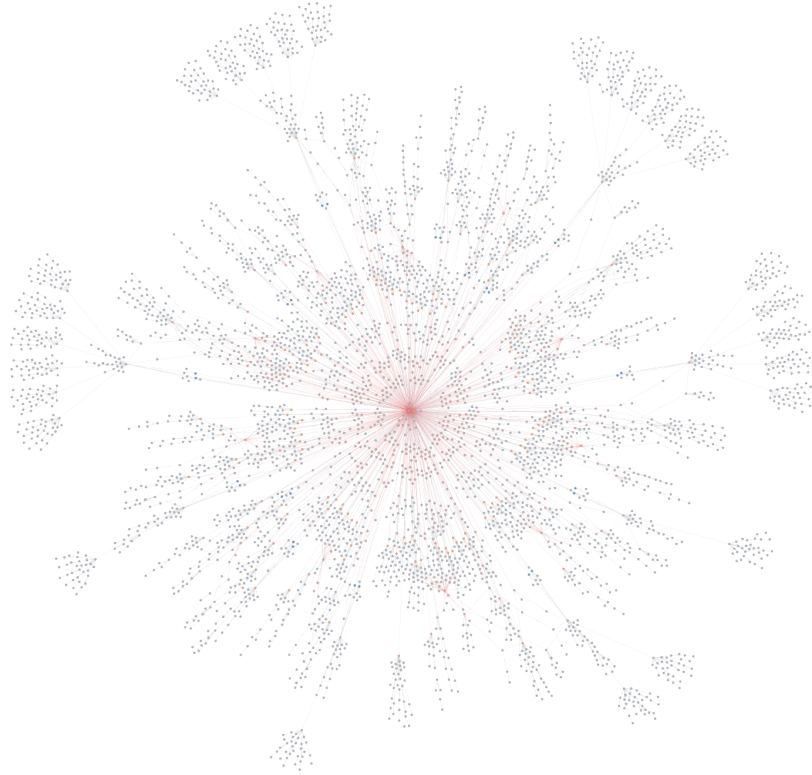
Feature paths before convolution



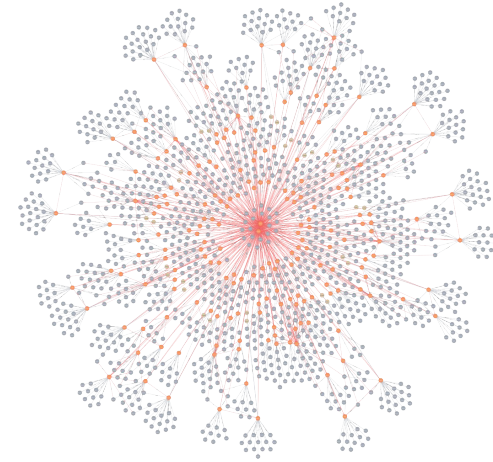
Feature paths after convolution



Optimisation



Generic bundle conversion



Optimised structure



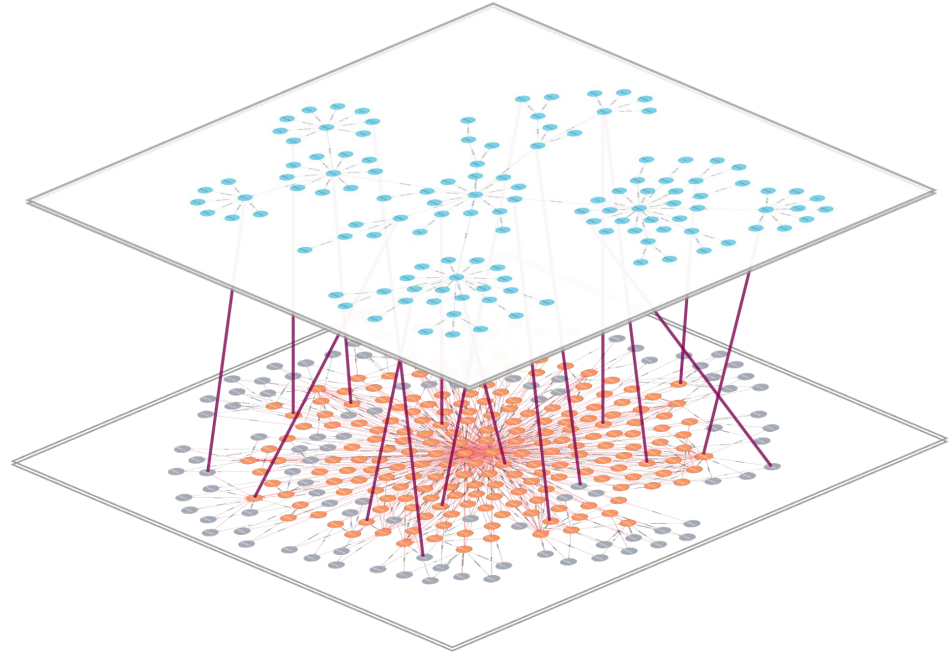
Enrichment

Ontology integration



BioCypher

biocypher.org/BioCypher



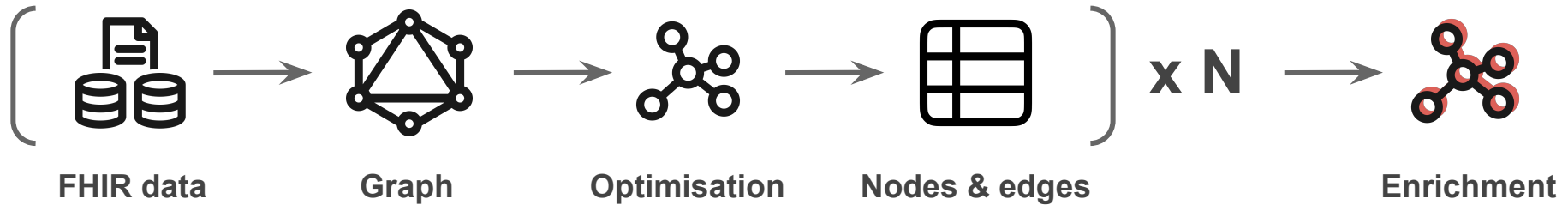
Challenges and solutions

Prototype set up at



Challenge: memory shortage

Solution: batching approach



Challenges and solutions

Challenge: duplicate Resources in different bundles

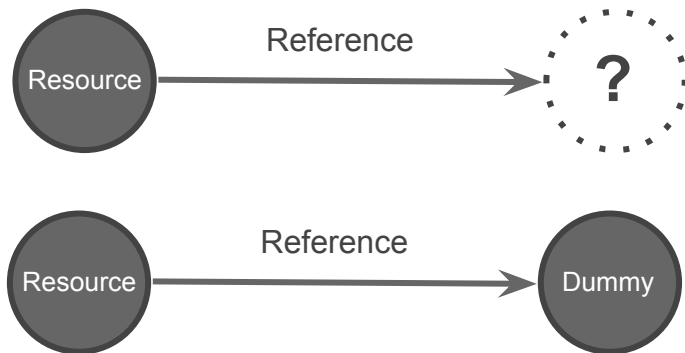
Solution: new unique ID for every Resource



Unique ID: <Resource type> + <Resource ID>

Challenge: missed References between Resources in different batches

Solution: dummy nodes for Reference targets

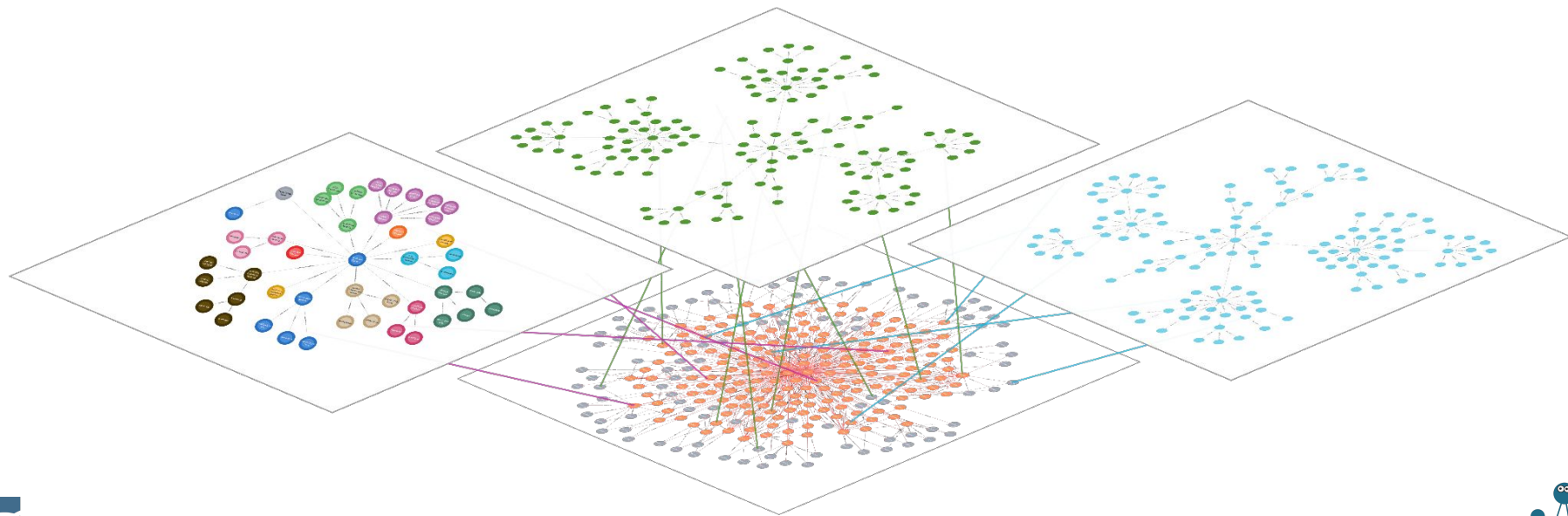


+ filtering in Neo4j import

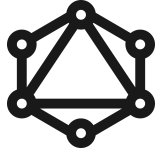


Perspectives

Integration of diverse data sources, i.e. Novel Analysis and Recommendation approach for study Features using graph technology (NARF), semantically enriched tabular data.



MeDaX pipeline



tinyurl.com/medax-pipeline



Acknowledgements



Sebastian Berthe

Kai Fitzer

Sherry Freiesleben

Robert Gött

Markus Mandalka

Seyed Reza Mazlooman

Sarah Braun

Tom Gebhardt

Lea Gütebier

Ron Henkel

Lea Michaelis

Benjamin Winter

Dagmar Waltemath

Sebastian Lobentanzer

Julio Saez-Rodriguez

Funded by



Federal Ministry
of Education
and Research



References

D. Bender and K. Sartipi, “HL7 FHIR: An Agile and RESTful approach to healthcare information exchange,” in Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, Porto, Portugal: IEEE, Jun. 2013, pp. 326–331. doi: 10.1109/CBMS.2013.6627810.

I. Mazein et al., “Graph databases in systems biology: a systematic review,” Brief. Bioinform., vol. 25, no. 6, p. bbae561, Sep. 2024, doi: 10.1093/bib/bbae561.

I. Mazein et al., “MeDaX: A Knowledge Graph on FHIR,” in Studies in Health Technology and Informatics, J. Mantas, A. Hasman, G. Demiris, K. Saranto, M. Marschollek, T. N. Arvanitis, I. Ognjanović, A. Benis, P. Gallos, E. Zoulias, and E. Andrikopoulou, Eds., IOS Press, 2024. doi: 10.3233/SHTI240423.

S. Lobentanzer et al., “Democratizing knowledge representation with BioCypher,” Nat. Biotechnol., vol. 41, no. 8, pp. 1056–1059, Aug. 2023, doi: 10.1038/s41587-023-01848-y.

T. Gebhardt, I. Mazein, L. Michaelis, R. Henkel, D. Waltemath, and J. Wodke, MeDaX pipeline. (Apr. 16, 2025). Zenodo. doi: 10.5281/ZENODO.15229077.

L. Gütebier et al., “Similarity scoring as a novel approach towards reusing and combining data from existing clinical and health studies,” German Medical Science, 2024, doi: 10.3205/24GMDS015.