

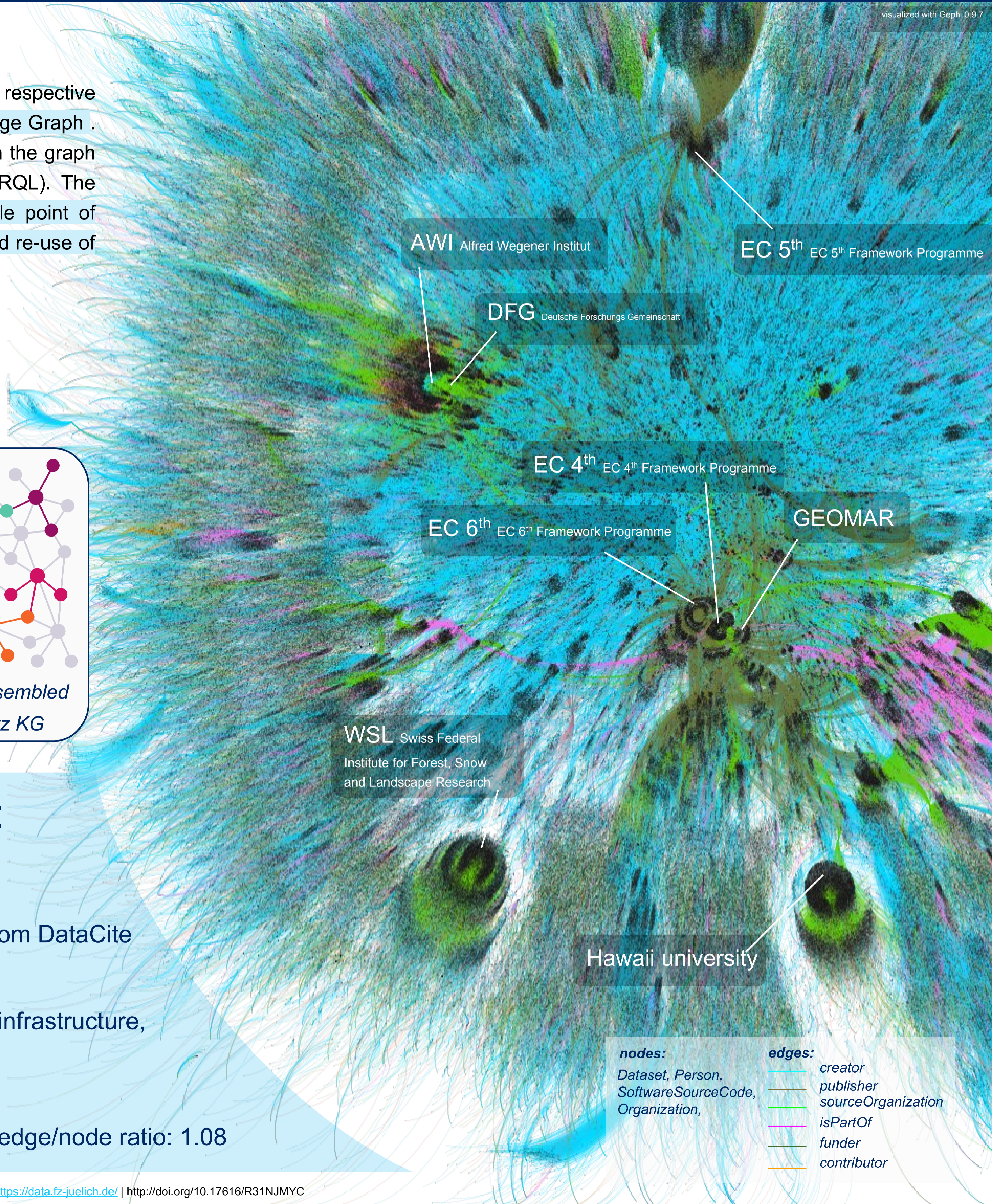
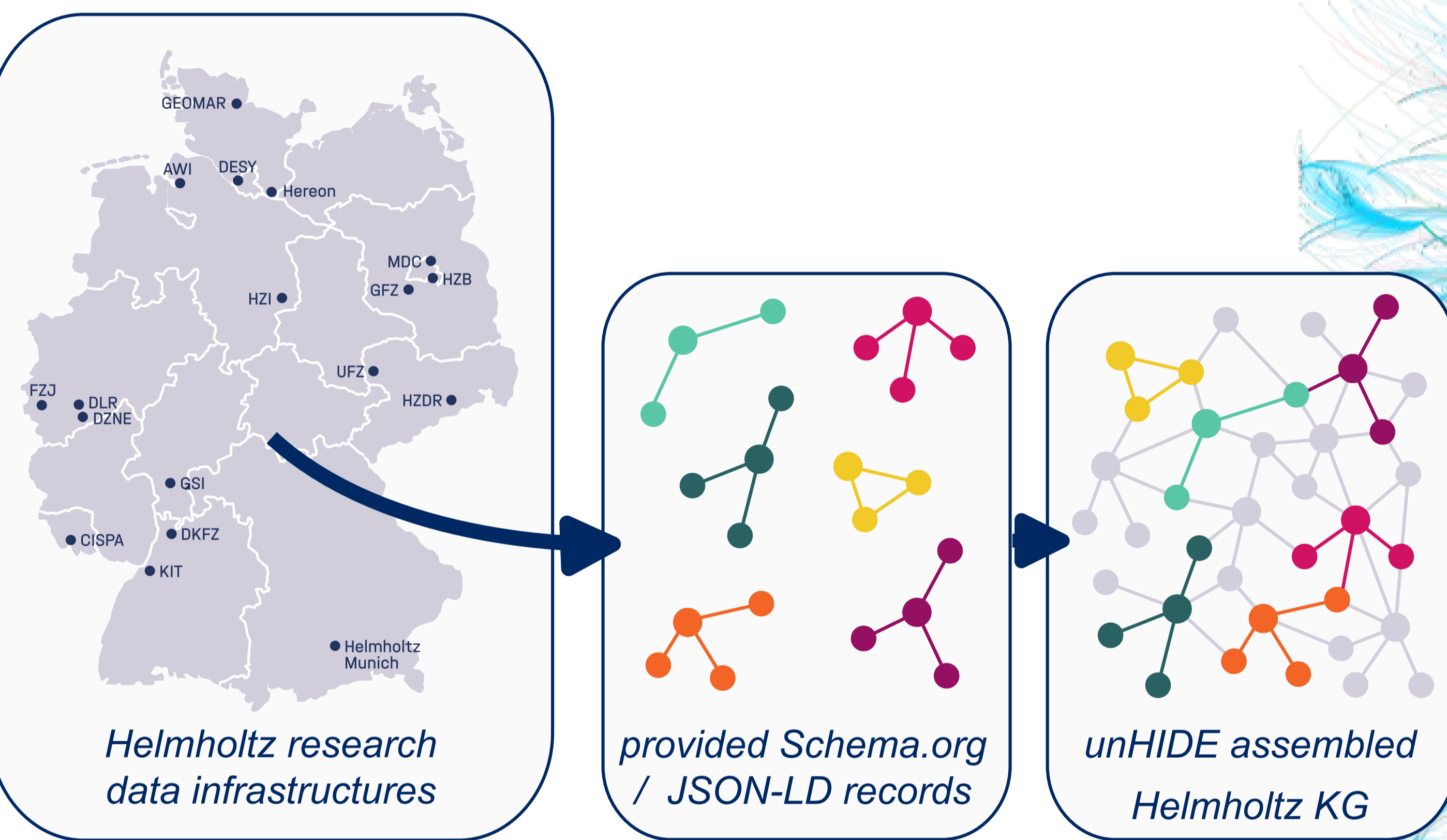
The Helmholtz Knowledge Graph & the unified Helmholtz information and data exchange (unHIDE)

Jens Bröder¹ // Volker Hofmann¹ // Pier Luigi Buttigieg² // Oonagh Mannix³ // Gabriel Preuss³ // Vivien Serve³ // Said Fathalla¹ // Fiona DMello¹ // Stefan Sandfeld¹

¹ Forschungszentrum Jülich; Institute for Advanced Simulation – Materials Data Science and Informatics (IAS-9)
² GEOMAR Helmholtz Centre for Ocean Research Kiel
³ Helmholtz Zentrum Berlin for Materials and Energy,

Connecting Helmholtz (meta)data

Helmholtz (meta)data is siloed in data infrastructures located in the respective Helmholtz centres. unHIDE connects this data to the Helmholtz Knowledge Graph. The used technology is co-developed with external stakeholders. Data in the graph is provided for re-use for humans (UI, API) and machines (API, SPARQL). The graph improves visibility for researchers and centers, provides a single point of access to information in the Helmholtz association and allows queries and re-use of data across infrastructures, centers and research fields.



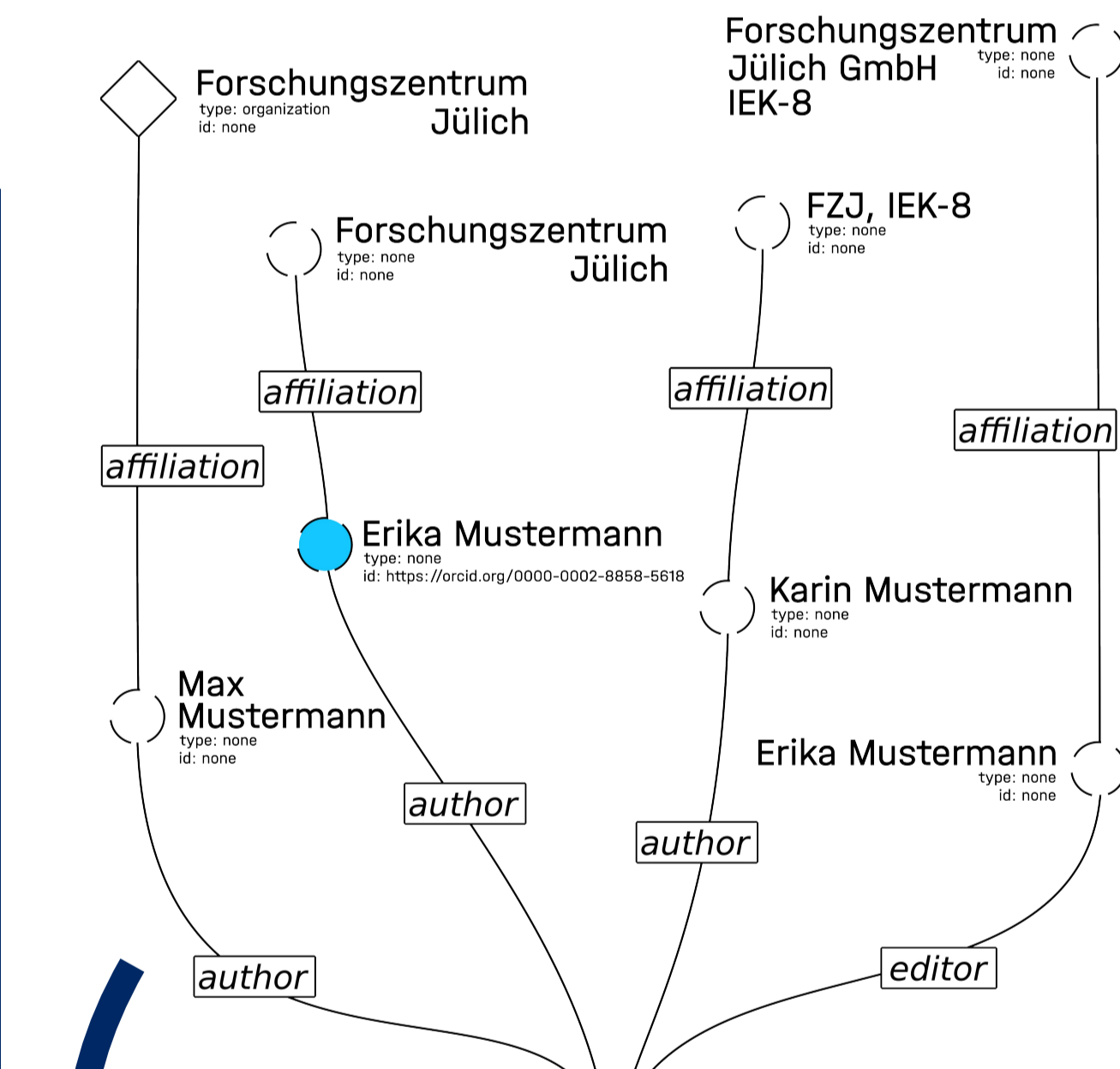
Improving records in Helmholtz data holdings

With the Helmholtz Knowledge graph unHIDE reveals the current state of (meta)data across Helmholtz data holdings. Data connectivity and connection quality is low. Structured areas of the graph do not reflect Helmholtz entities (centers, repositories or infrastructures). We improve Helmholtz metadata records by uplifting provided documents and feed uplifted data back to the data source. For well known entities (i.e. Helmholtz research centers) we harmonize records through SPARQL updates. We resolve entities by assigning types and ids. Where possible, we infer entity types and ids through the assembled data.

Original provided record

```

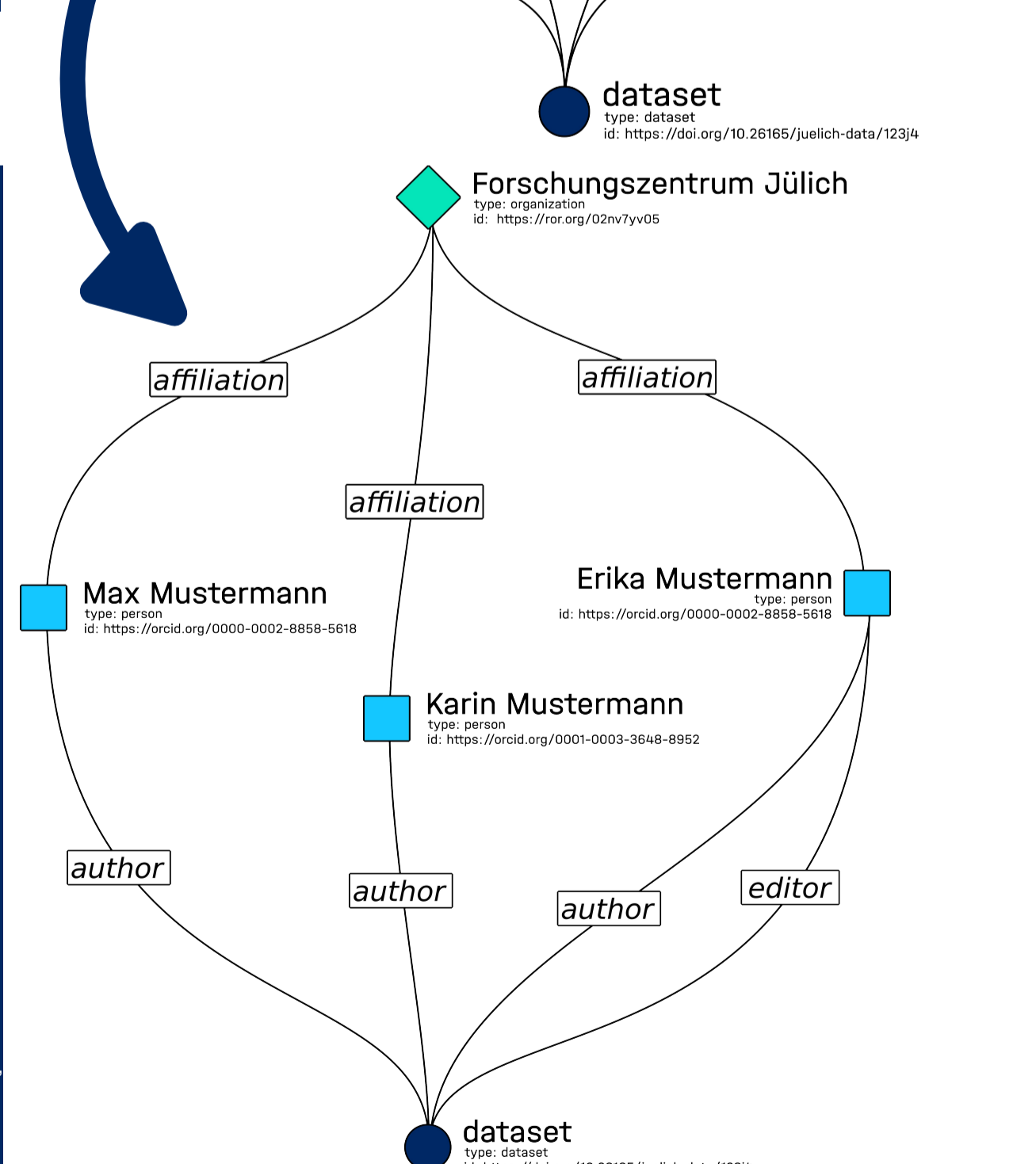
{
  "@context": "http://schema.org",
  "@id": "https://doi.org/10.28165/juelich-data/1234",
  "@type": "dataset",
  "author": [
    {
      "affiliation": {
        "@type": "Organization",
        "name": "Forschungszentrum Jülich GmbH, IEK-8"
      },
      "name": "Max Mustermann"
    },
    {
      "@id": "https://orcid.org/0000-0002-8858-5618",
      "affiliation": {
        "name": "Forschungszentrum Jülich GmbH"
      },
      "name": "Erika Mustermann"
    },
    {
      "affiliation": {
        "name": "FZJ, IEK-8"
      },
      "name": "Karin Mustermann"
    },
    {
      "affiliation": {
        "@type": "Organization",
        "name": "Forschungszentrum Jülich GmbH, IEK-8"
      },
      "name": "Erika Mustermann"
    }
  ],
  "editor": {
    "name": "Name of the dataset."
  }
}
  
```



unHIDE uplifted record

```

{
  "@context": "http://schema.org",
  "@id": "https://doi.org/10.28165/juelich-data/1234",
  "@type": "dataset",
  "author": [
    {
      "@type": "Person",
      "affiliation": {
        "@id": "https://ror.org/02nv7yv05",
        "@type": "Organization",
        "name": "Forschungszentrum Jülich GmbH"
      },
      "name": "Max Mustermann"
    },
    {
      "@id": "https://orcid.org/0000-0002-8858-5618",
      "@type": "Person",
      "affiliation": {
        "@id": "https://ror.org/02nv7yv05",
        "@type": "Organization",
        "name": "Forschungszentrum Jülich GmbH"
      },
      "name": "Erika Mustermann"
    },
    {
      "@id": "https://orcid.org/0001-0003-3648-8952",
      "@type": "Person",
      "affiliation": {
        "@id": "https://ror.org/02nv7yv05",
        "@type": "Organization",
        "name": "Forschungszentrum Jülich GmbH"
      },
      "name": "Karin Mustermann"
    },
    {
      "@id": "https://ror.org/02nv7yv05",
      "@type": "Organization",
      "name": "Forschungszentrum Jülich GmbH"
    }
  ],
  "editor": {
    "name": "Name of the dataset."
  }
}
  
```



Helmholtz Knowledge Graph facts:

harvested infrastructures:

Rodare¹ | Pangaea² | JülichData³ | 16/18 Helmholtz libraries | 12 (all) Helmholtz GitLab instances | Helmholtz Sub-Graph from DataCite

- 2.15 mio records harvested
- representing: documents, datasets, software, institutions, infrastructure, other digital resources, researchers & experts, projects

visualized sub-graph:

- # records 552 k | # nodes: 2.26 mio | # edges: 2.43 mio | edge/node ratio: 1.08

¹ <https://rodare.hzdr.de/> | <http://doi.org/10.17616/R3BR40> ² <https://www.pangaea.de/> | <http://doi.org/10.17616/R3XS37> ³ <https://data.fz-juelich.de/> | <http://doi.org/10.17616/R31NJMYC>

follow our progress

