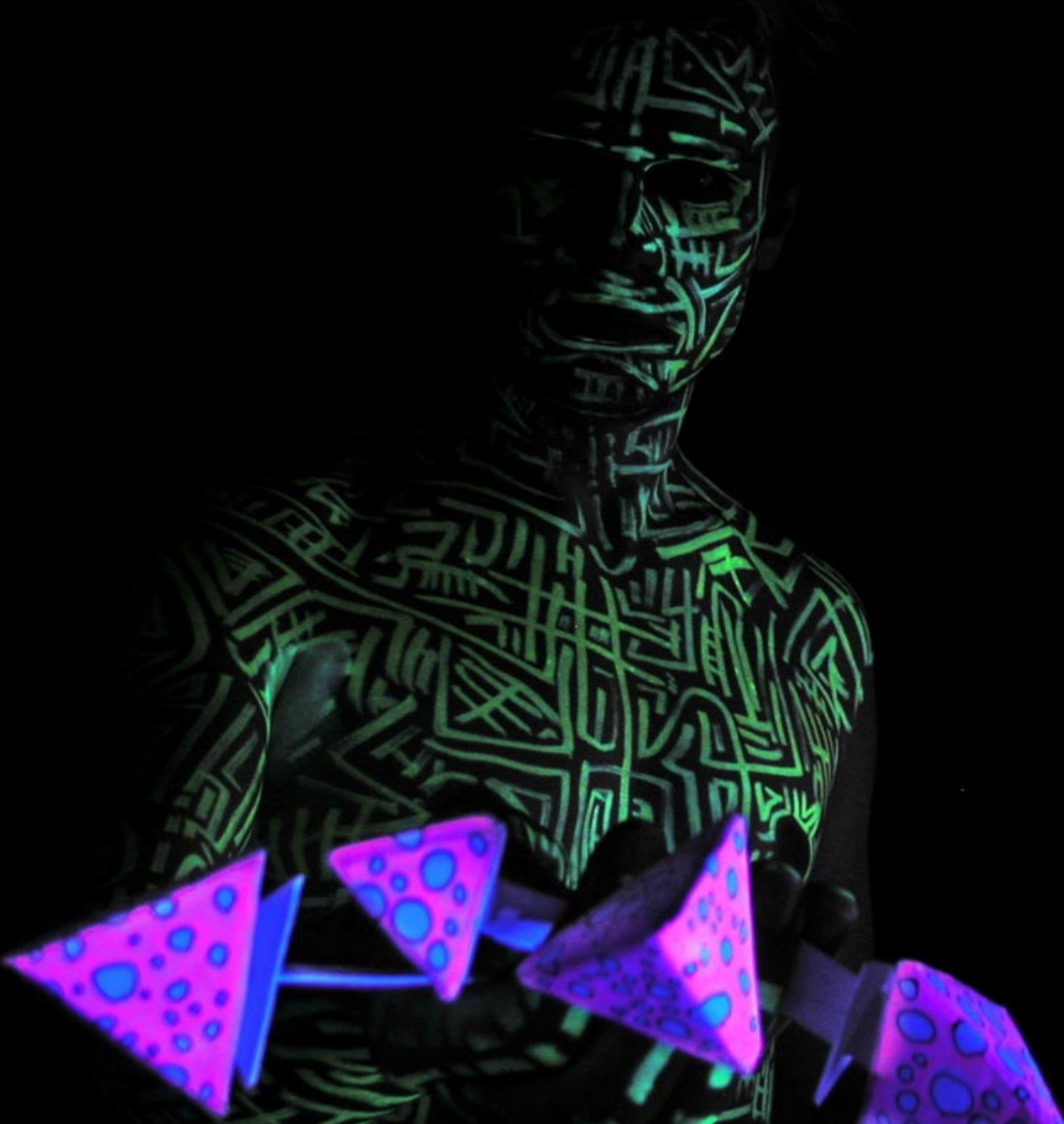


Provenance Information in the Web of Data

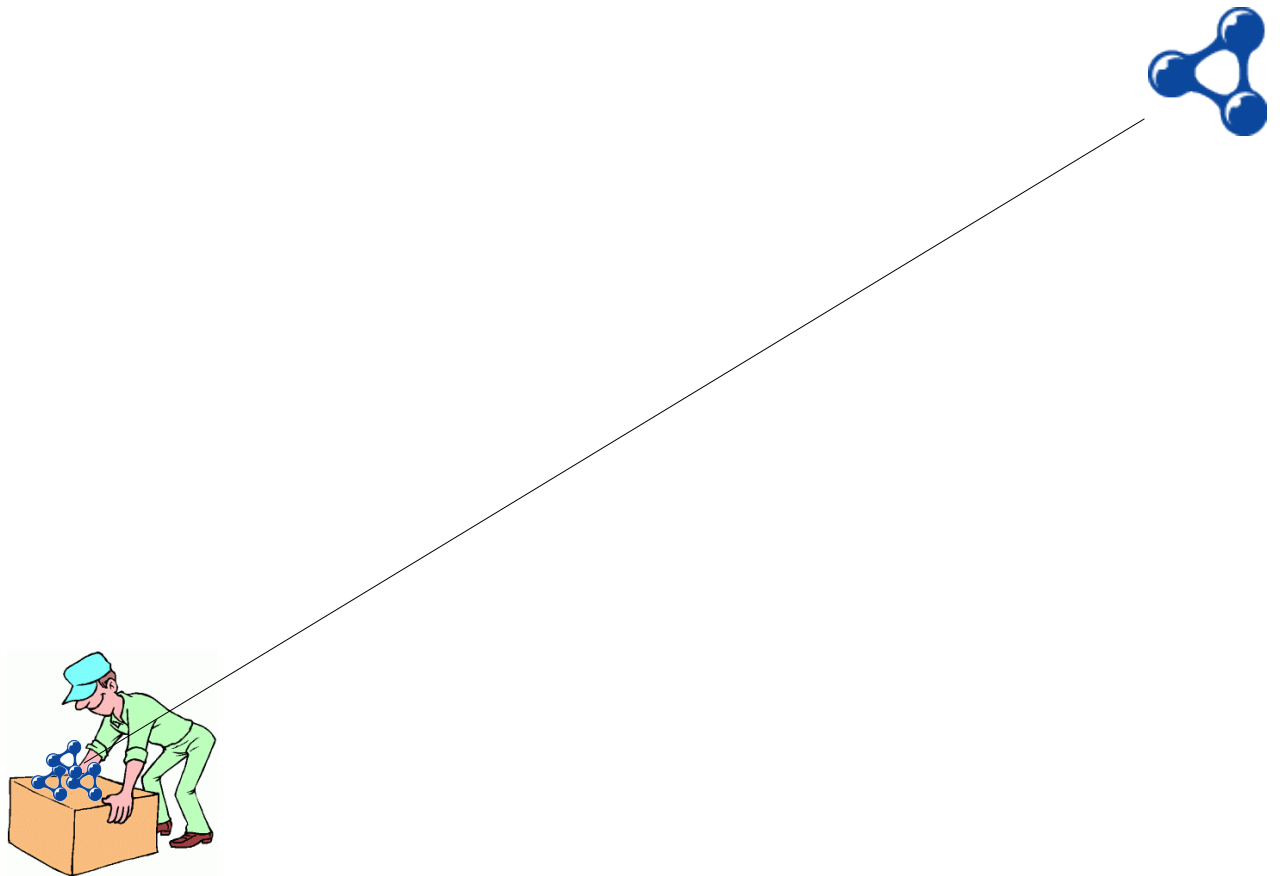


Olaf Hartig

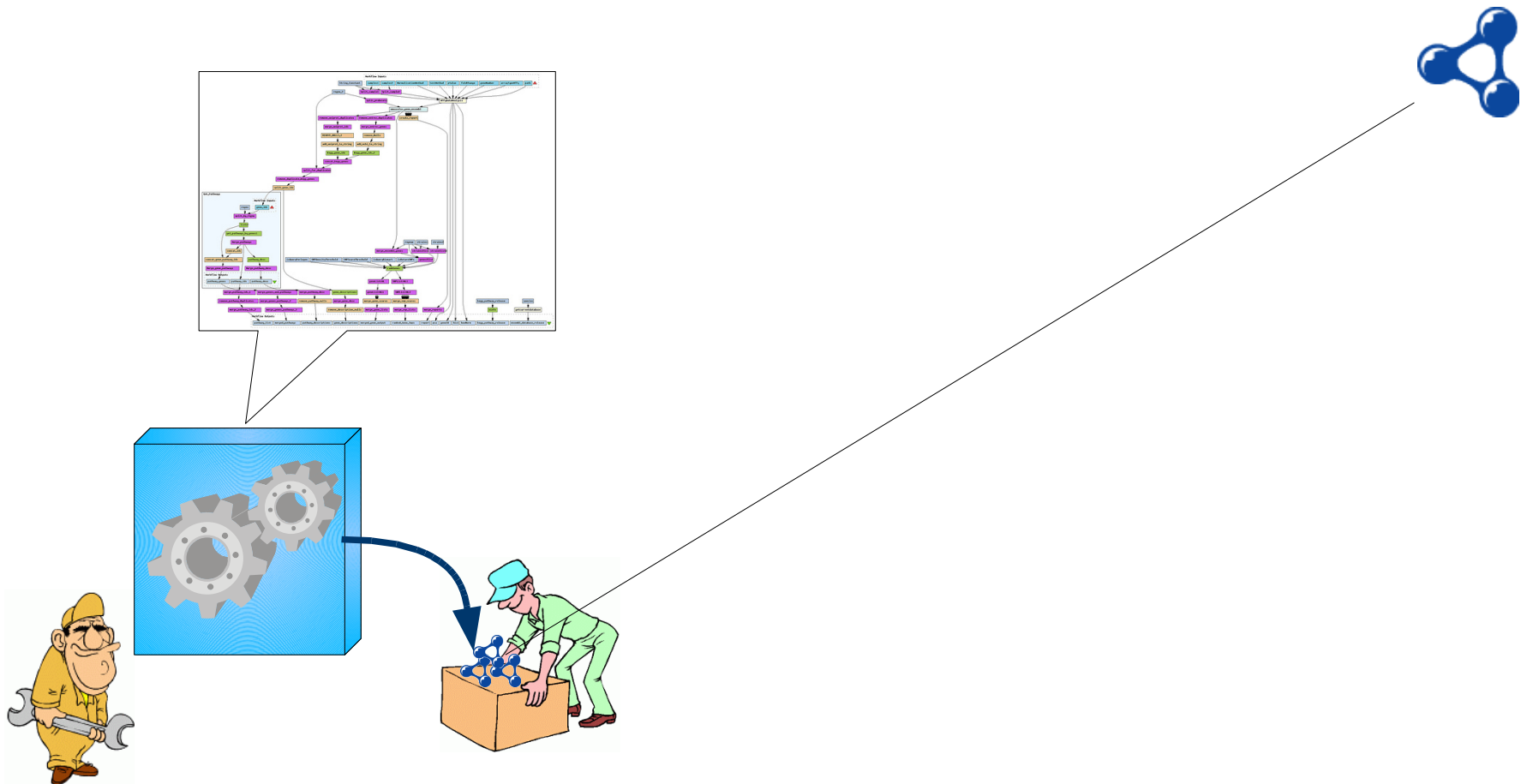
Humboldt-Universität zu Berlin

<http://olafhartig.de/foaf.rdf#olaf>

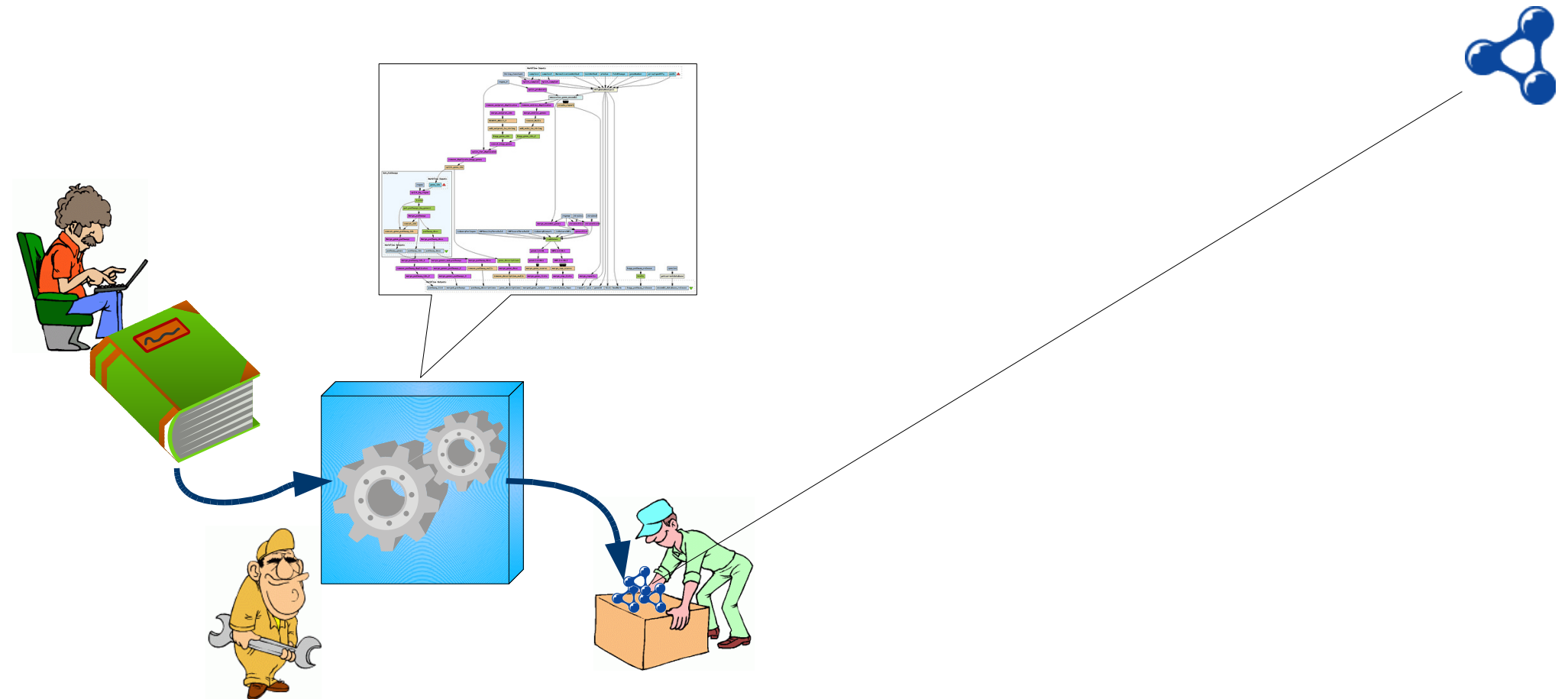
- **Provenance of a data item: information about the history**



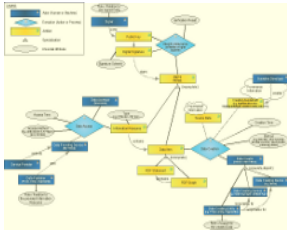
- **Provenance of a data item: information about the history**



- **Provenance of a data item: information about the history**



Outline



**Towards a model of
Web data **provenance****



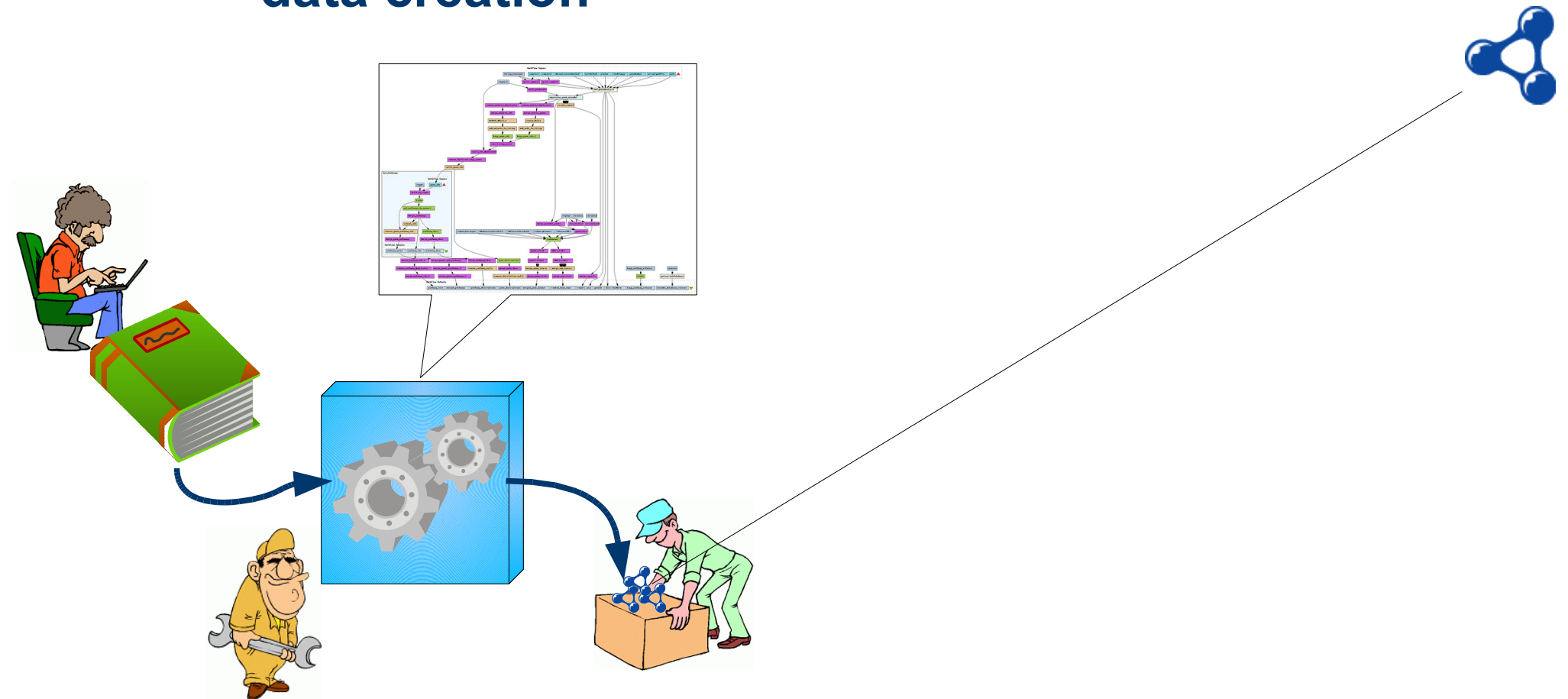
**Provenance information
in the Web of data **today****

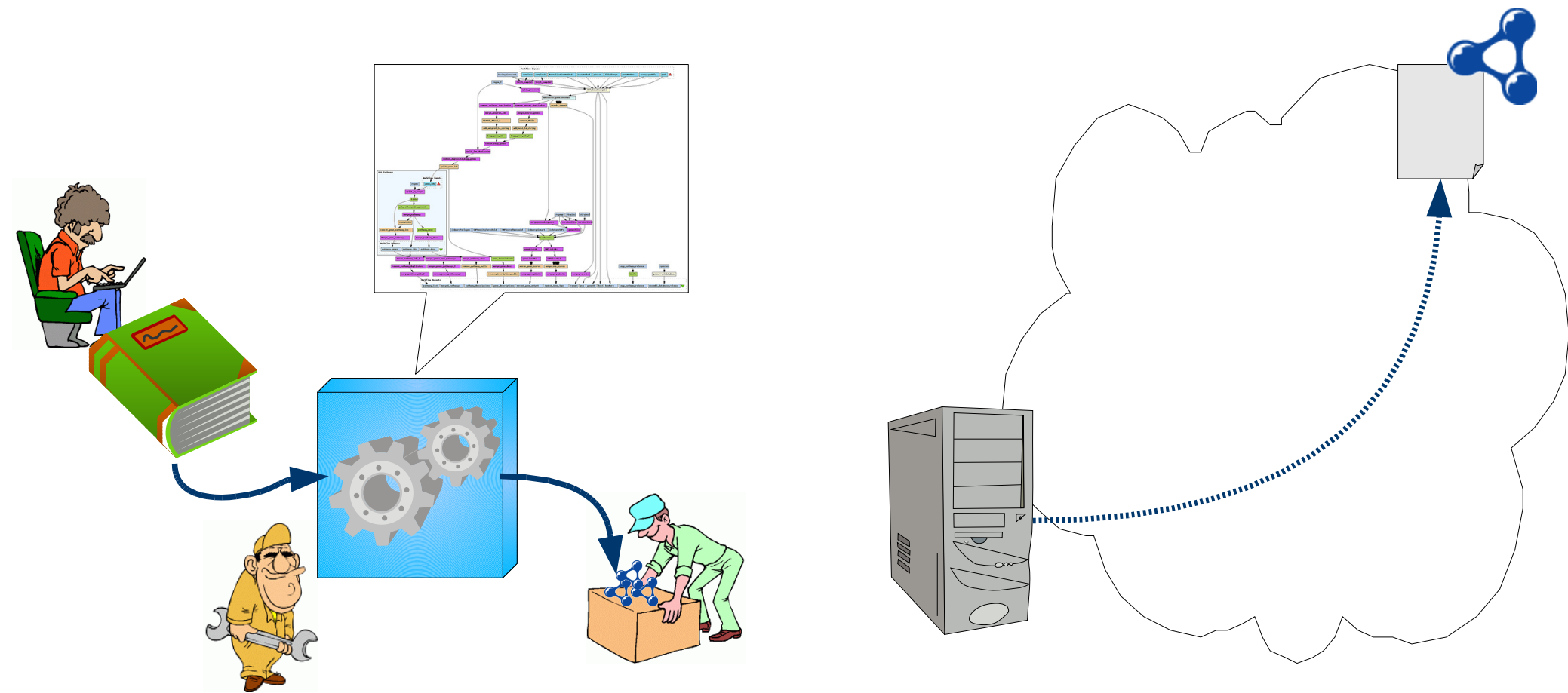


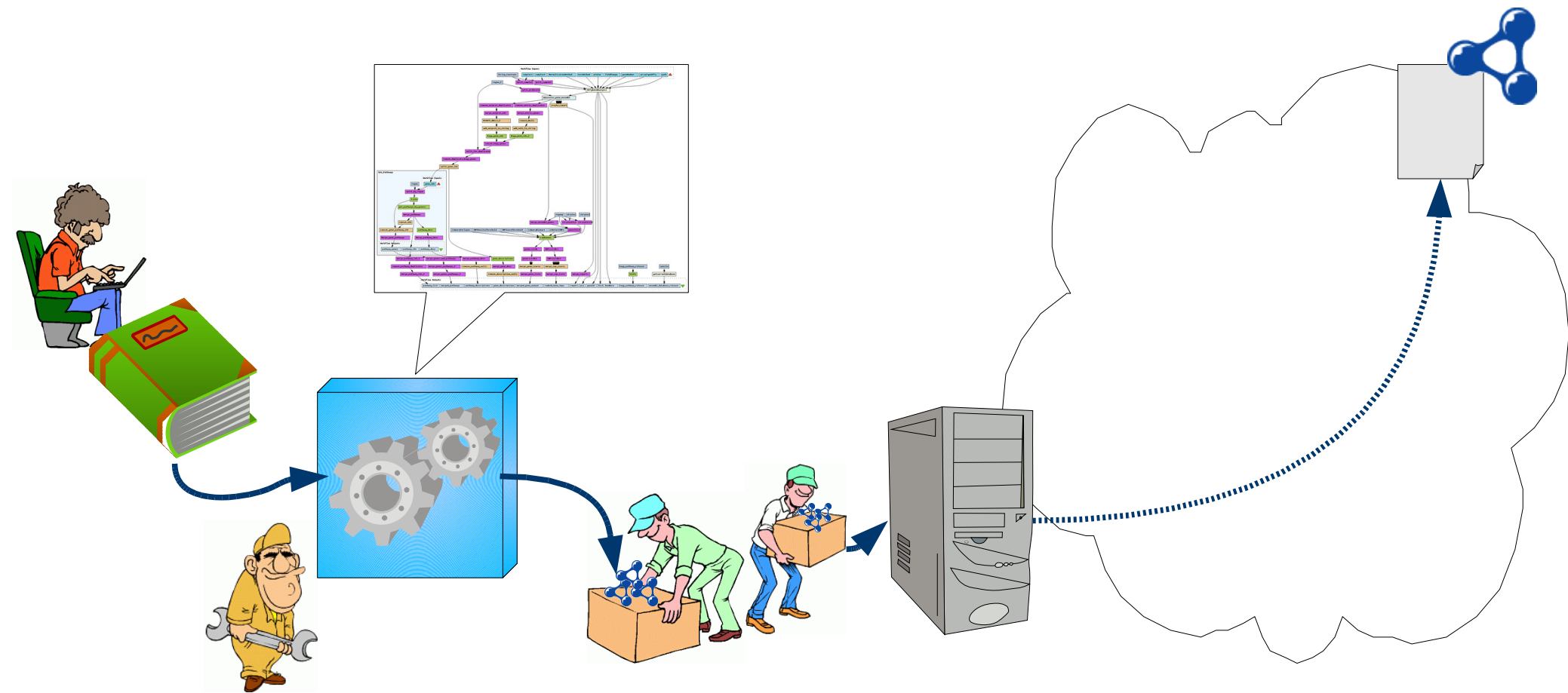
**Upcoming
tasks**

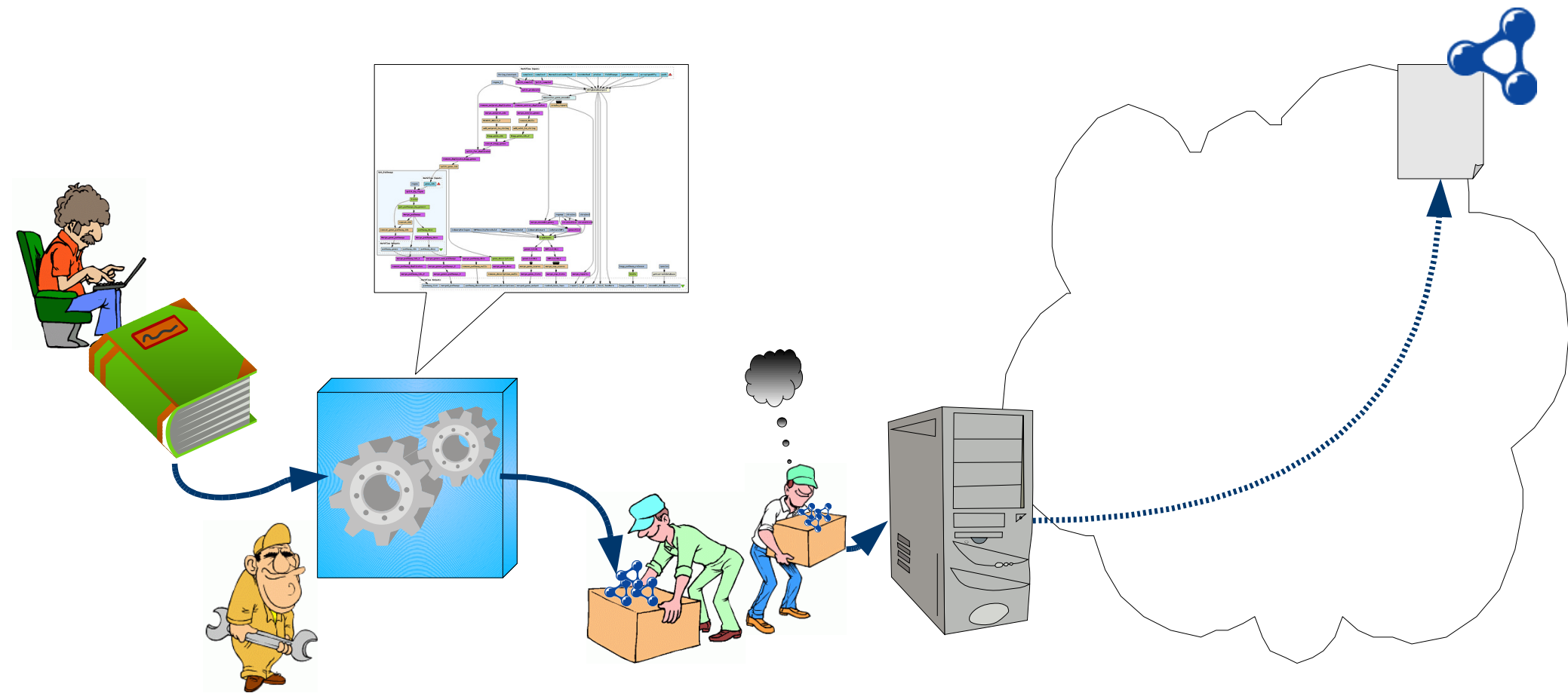
Existing Provenance Research

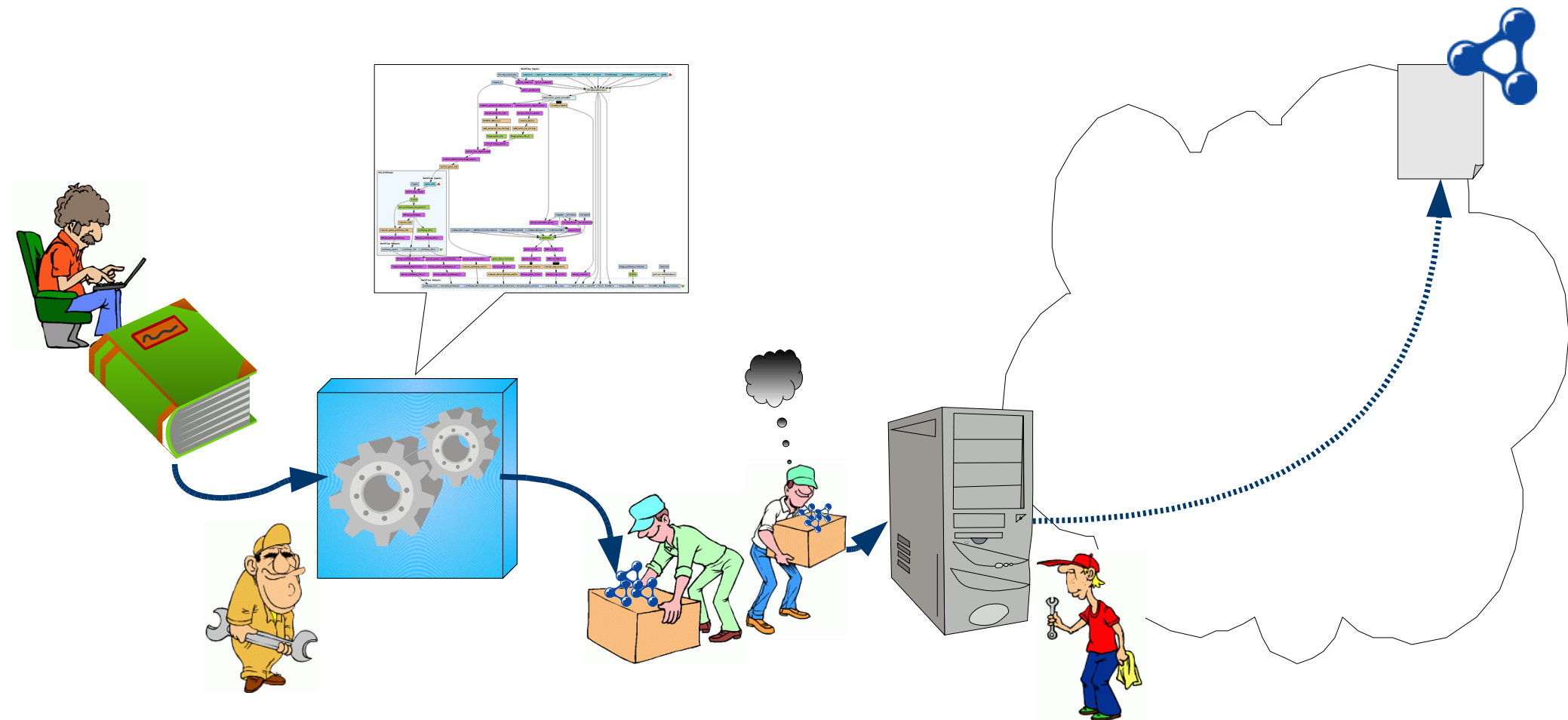
- Main research areas: (scientific) workflows, DBMSs
- General focus:
data creation











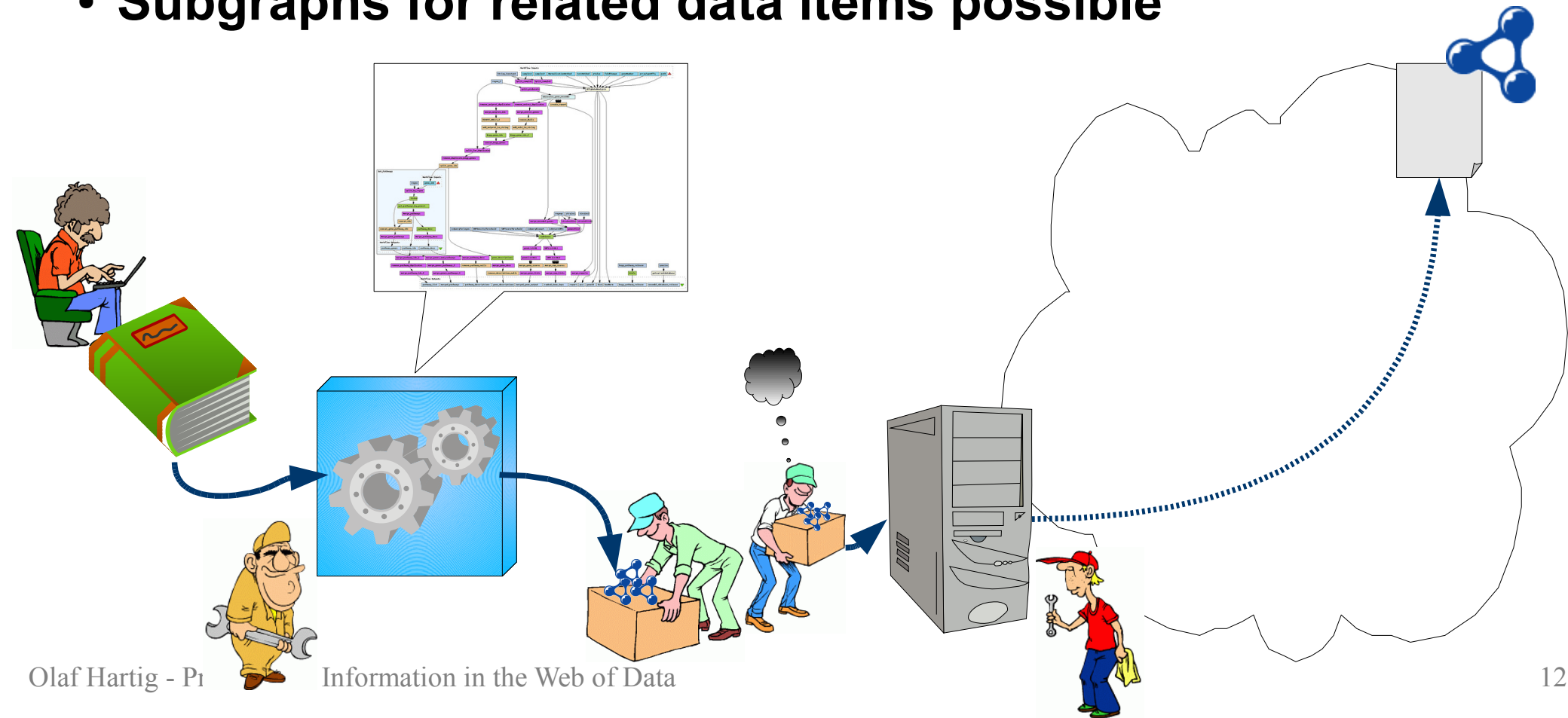


**Web data provenance
comprises
two dimensions:**

Data Creation • Data Access

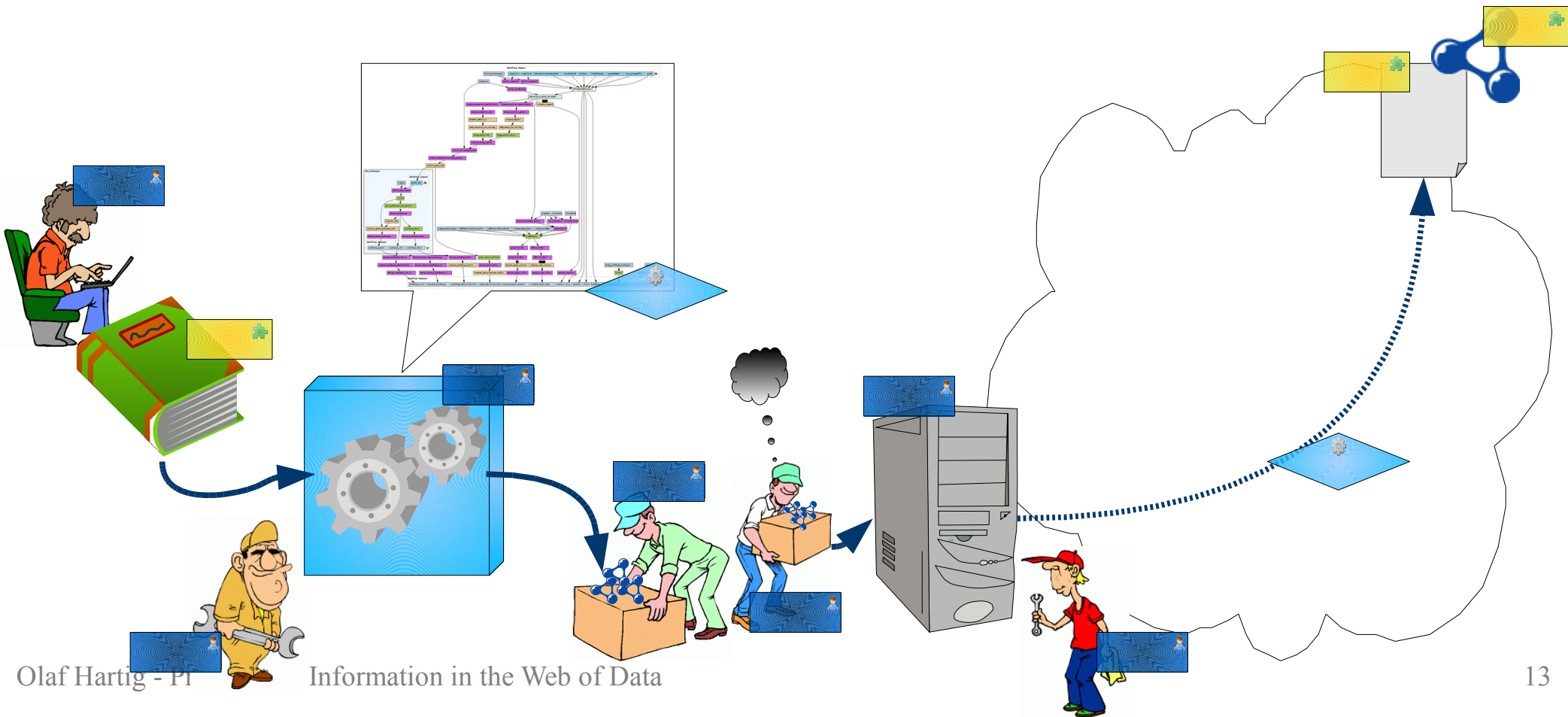
Basics of the Provenance Model

- **Provenance graph** describes provenance of a data item
 - **Nodes: provenance elements** – pieces of provenance info
 - **Edges:** relate provenance elements to each other
 - **Subgraphs for related data items possible**



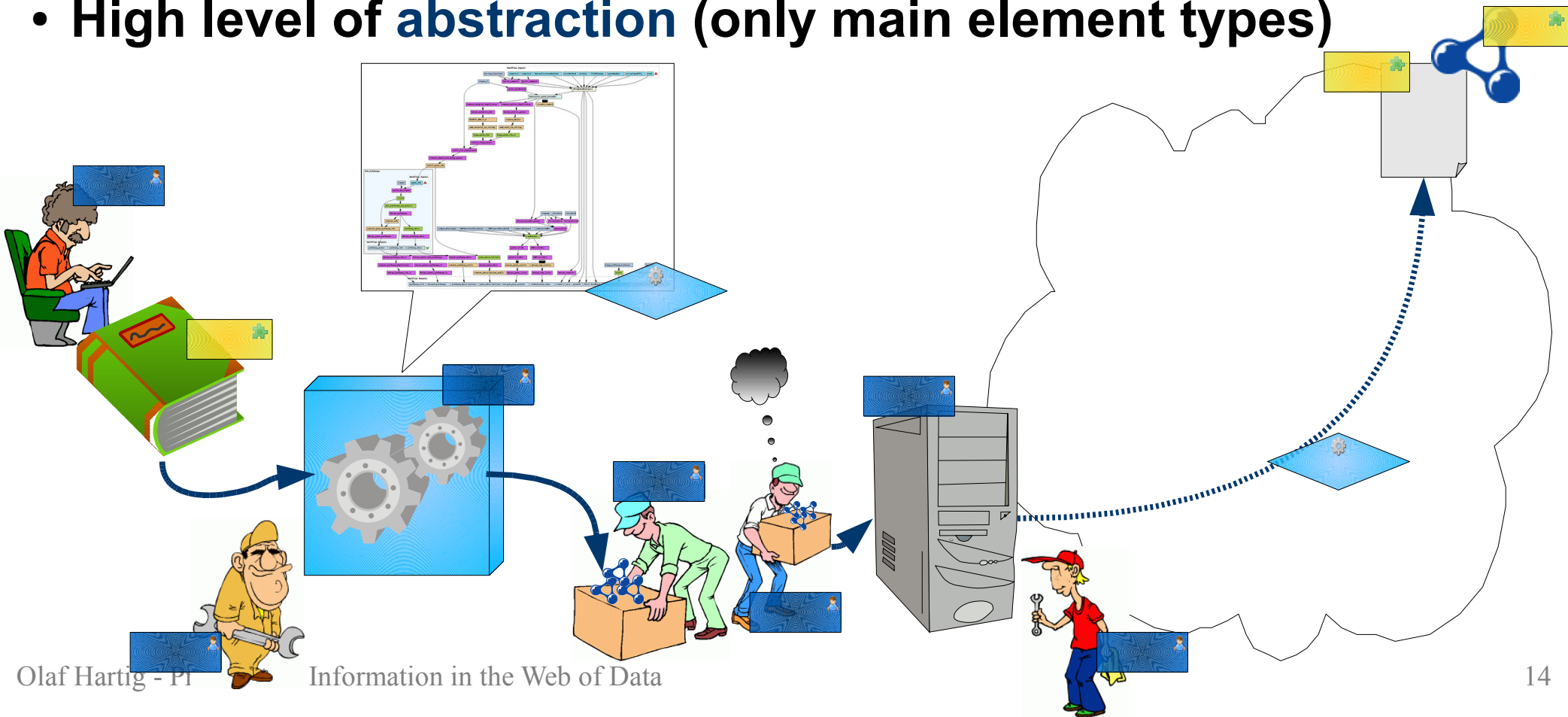
Basics of the Provenance Model

- **Provenance model** defines:
 - Types of provenance elements
 - Relationships



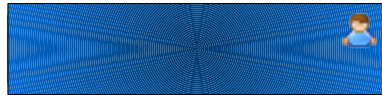
Basics of the Provenance Model

- **Provenance model** defines:
 - Types of provenance elements
 - Relationships
- High level of **abstraction** (only main element types)

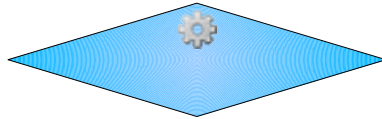


Basics of the Provenance Model

- **General differentiation:**



Actors

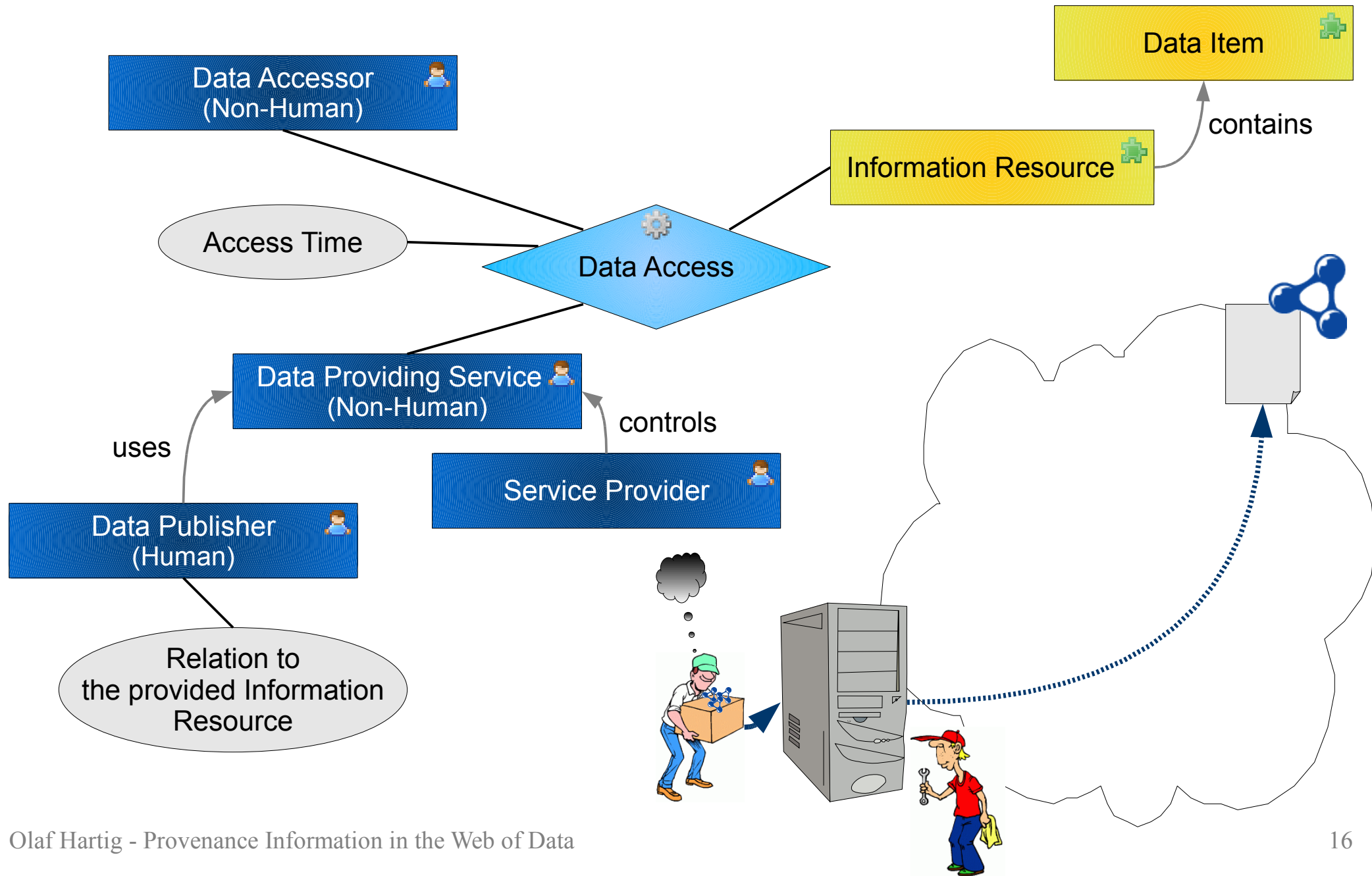


Executions

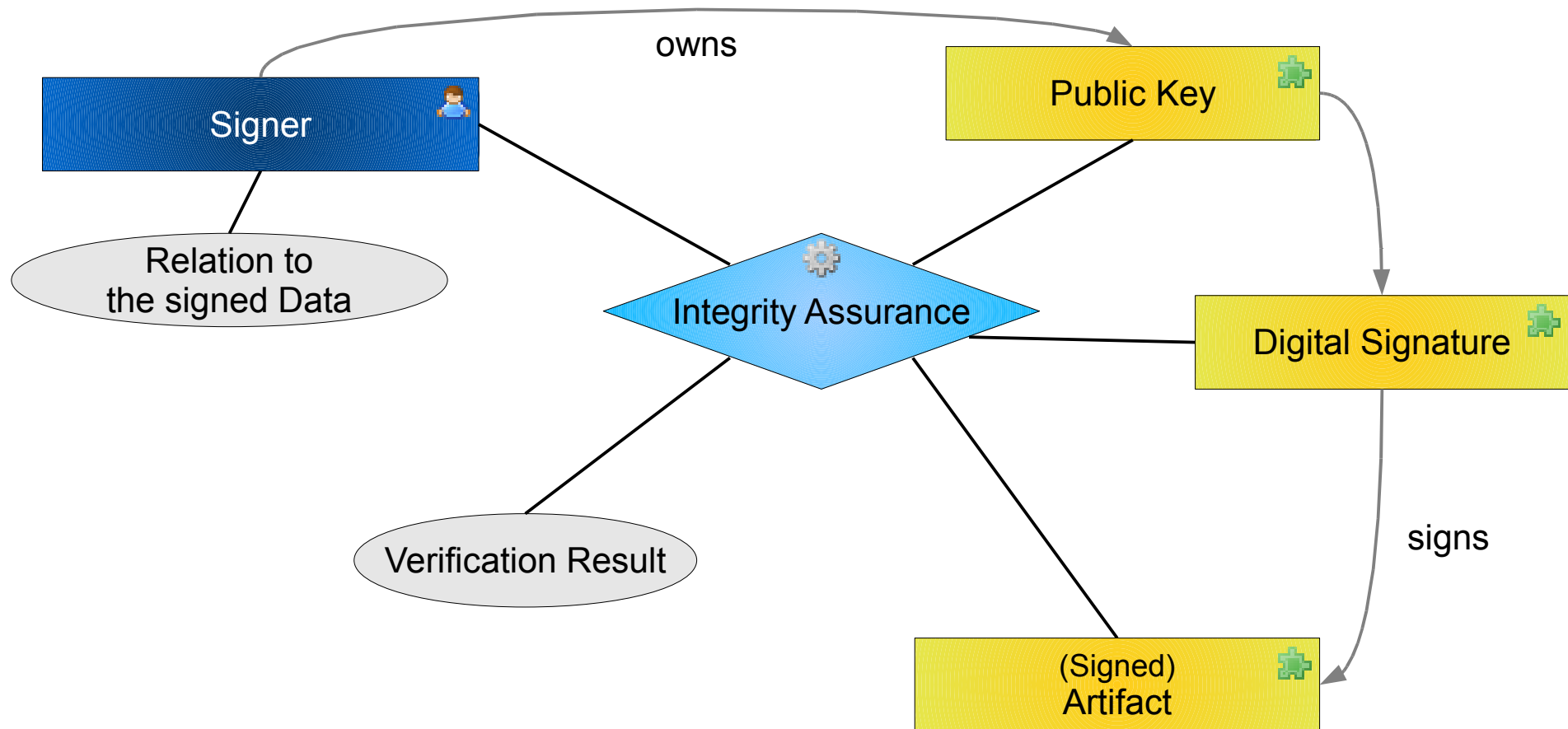


Artifacts

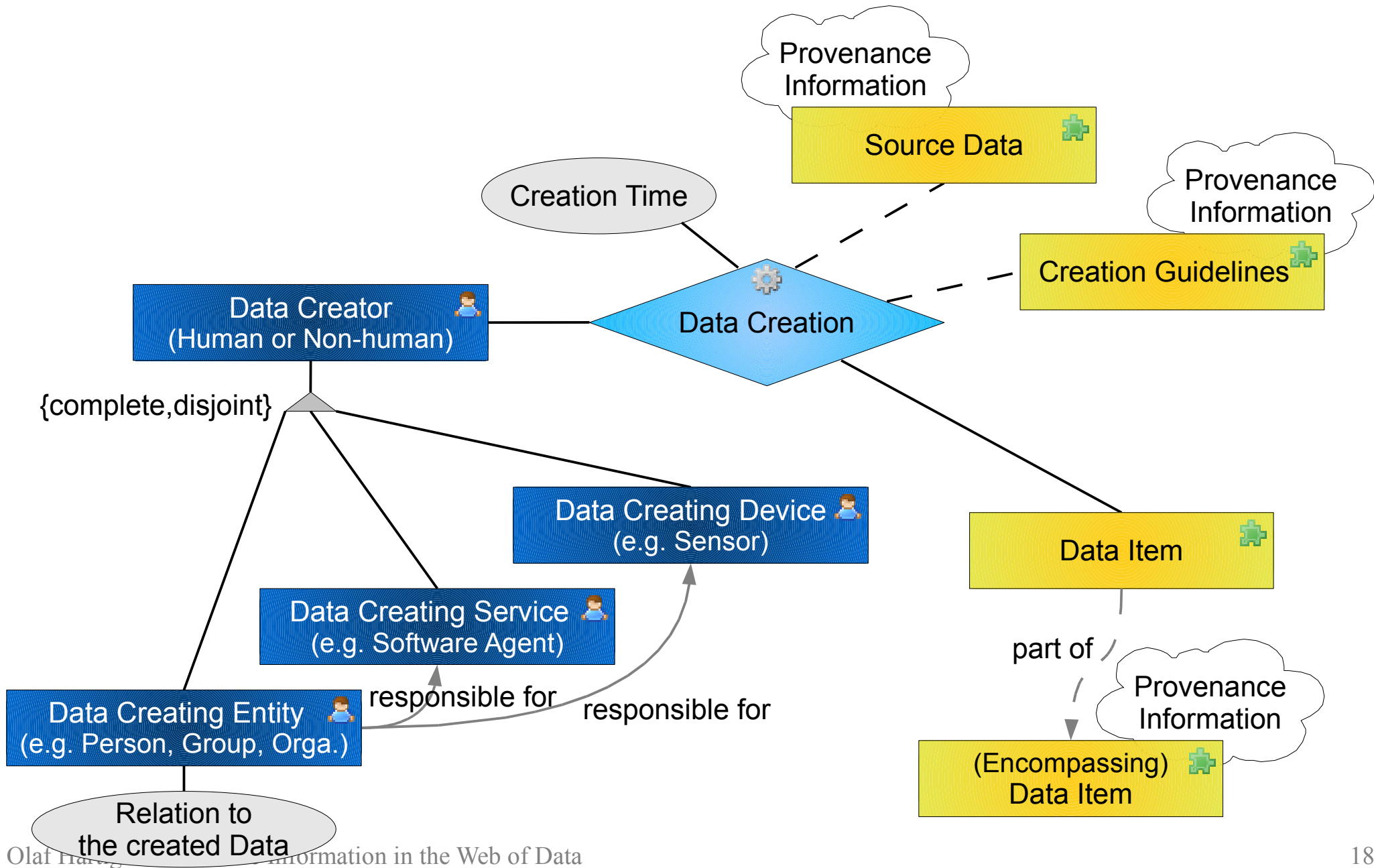
Data Access Dimension



Data Access Dimension cont.



Data Creation Dimension





**Provenance information
in the Web of data **today****

Provenance-related Vocabularies



DC – Dublin Core Metadata Terms



FOAF – Friend of a Friend



SIOC – Semantically-Interlinked Online Communities

- **SWP** – Semantic Web Publishing vocabulary
- **WOT** – Web of Trust schema
- **OMV** – Ontology Metadata Vocabulary
- **PML** – Proof Markup Language
- **Changeset** vocabulary
- **Ouzo Provenance Ontology**

Main Issues Today



- **Vocabularies:**
 - **Partly unsuitable**
 - **Lack of certain features**
 - **Coverage of provenance model impossible**

Provenance-related Vocabularies



DC – Dublin Core Metadata Terms

Property	Occurrences*
dc:creator	about 24,284
dc:contributor	476
dc:source	about 3,631
dc:created	about 82,720
dc:modified	about 12,020
dc:provenance	7

*Measured by querying Sindice; Feb. 7, 2009 (by that time Sindice indexed about **48,99 million** documents)

Main Issues Today



- Vocabularies:
 - Partly unsuitable
 - Lack of certain features
 - Coverage of provenance model impossible
- **General lack of provenance-related metadata on the Web of data**

Possible Reasons



- **Lack of suitable vocabularies**
- **Lack of usable tools**
- **Ignorance / lack of sensitization**



**Upcoming
tasks**

Address the Issues

- Let's **develop a vocabulary** for Web data provenance
 - **Proposal: refine the presented provenance model**
 - **Integrate existing vocabularies for specific types of provenance elements**

Address the Issues

- Let's **develop a vocabulary** for Web data provenance
 - **Proposal: refine the presented provenance model**
 - **Integrate existing vocabularies for specific types of provenance elements**
- Let's **develop usable tools** for data providers
 - **Edit and publish provenance-related metadata**
 - **Automatic generation if possible**

Address the Issues

- Let's **develop a vocabulary** for Web data provenance
 - Proposal: refine the presented provenance model
 - Integrate existing vocabularies for specific types of provenance elements
- Let's **develop usable tools** for data providers
 - Edit and publish provenance-related metadata
 - Automatic generation if possible
- Let's **raise awareness** of data providers
 - Probably the hardest task
 - Maybe voidID can help

Thank you!

Olaf Hartig

Humboldt-Universität zu Berlin

<http://olafhartig.de/foaf.rdf#olaf>





**These slides have been created by
Olaf Hartig**

<http://olafhartig.de>

**This work is licensed under a
Creative Commons Attribution-Share Alike 3.0 License**
(<http://creativecommons.org/licenses/by-sa/3.0/>)

Attribution:

- <http://www.flickr.com/photos/adrenalin/3032734/>
- <http://www.hasslefreeclipart.com>
- <http://www.flickr.com/photos/dullhunk/428079229/>
- <http://www.flickr.com/photos/darwinbell/1337963794/>
- <http://www.flickr.com/photos/alandd/2780700767/>
- http://www.flickr.com/photos/simeon_barkas/2872099696/
- <http://www.flickr.com/photos/robinh00d/122544491/>
- <http://www.flickr.com/photos/adrenalin/3032747/>