Wissenschaftsgeleitete Forschungsinfrastrukturen für die Geistes- und Kulturwissenschaften in Deutschland
3. Workshop: Politische Perspektive

STANDARDS UND SCHNITTSTELLEN

Henning Lobin

Regine Stein

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

Standards

Was ist ein Standard?

„Ein Standard ist ein **Dokument**, das **Anforderungen**, **Spezifikationen**, **Richtlinien** oder **Merkmale** bietet, die **konsequent genutzt** werden können, um sicherzustellen, dass die **Materialien**, **Produkte**, **Prozesse** und **Dienstleistungen** für ihren Zweck **geeignet** sind.“ (ISO)

Standards lassen sich in **zwei Kategorien** unterscheiden:

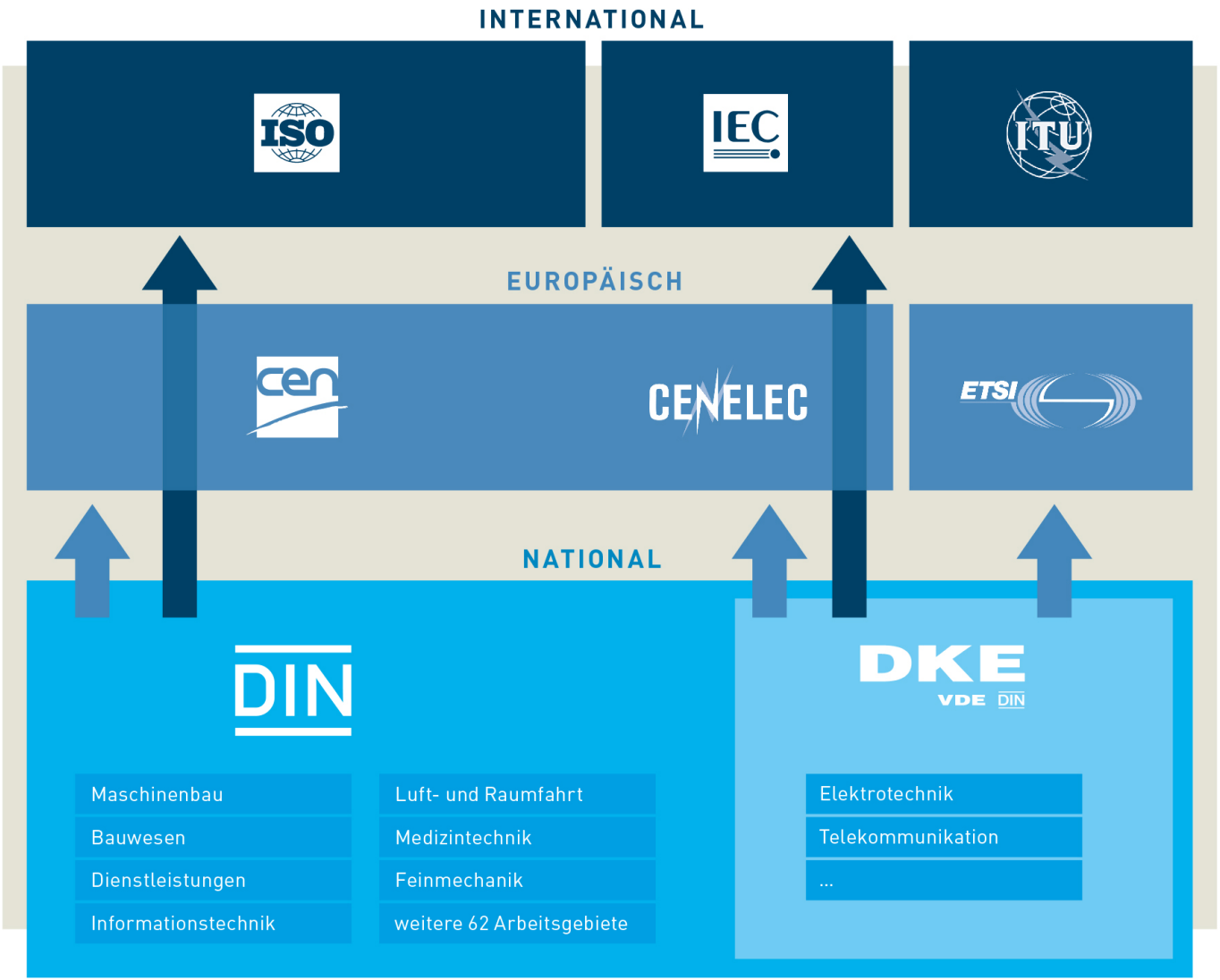
- *de jure*-Standards
- *de facto*-Standards

de jure-Standards

- Werden in der Regel durch nationale oder internationale **Normungsgremien** erarbeitet.
- Ihre **Verwendung kann** institutionell **gefordert** werden
- **Organisationen** z.B.:
  - *International Organization for Standardization* (ISO)
    - *Deutsches Institut für Normung* (DIN)
    - *American National Standards Institute* (ANSI)
  - *World Wide Web Consortium* (W3C)
  - *Text Encoding Initiative* (TEI)

C

-
-
-



ISO: Internationale Organisation für Normung

IEC: Internationale Elektro-technische Kommission

ITU: Internationale Fernmeldeunion

CEN: Europäisches Komitee für Normung

CENELEC: Europäisches Komitee für Elektrotechnische Normung

ETSI: Europäisches Institut für Telekommunikationsnormen

DIN: Deutsches Institut für Normung e.V.

DKE: Deutsche Kommission Elektrotechnik Elektronik Informationstechnik in DIN und VDE

DIN und DKE vertreten die nationalen Interessen in der europäischen und inter-nationalen Normung.


1987!
THE VILLARD ROOM
NO SMOKING

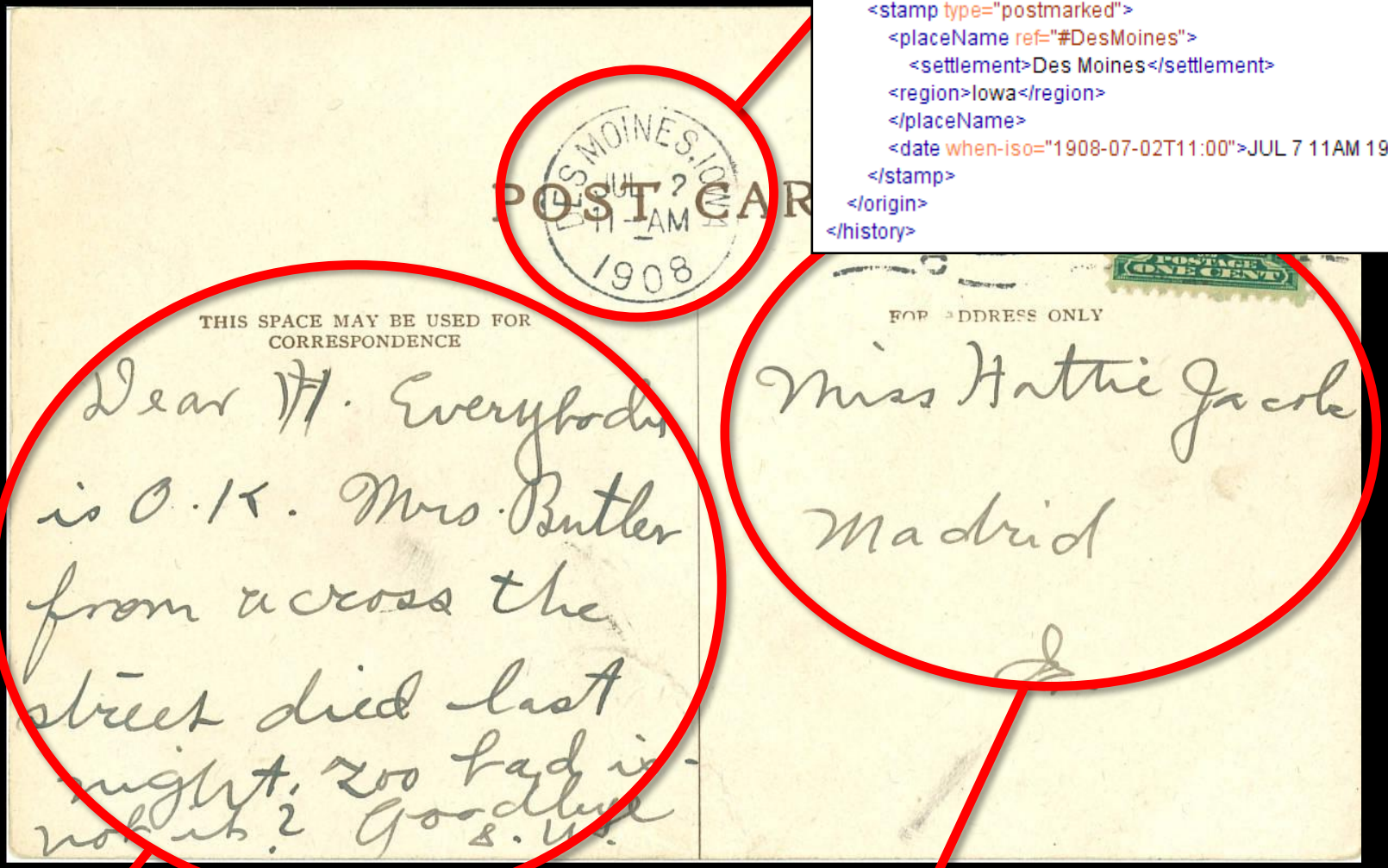
de facto-Standards

de facto-Standards werden …

- **ohne gesetzlichen Auftrag** entwickelt,
- durch eine **einzelne Firma/Organisation** oder mehrere **Unternehmen** erarbeitet,
- durch **weite Verbreitung und Annahme** zu formalen Normen aufgewertet.
- Beispiele:
  - **MP3** (Fraunhofer)
  - **docx** (Microsoft)
  - **PDF** (Adobe)
- In einigen Fällen werden *de facto*-Standards **nachträglich zu *de jure*-Standards** eingebracht (z.B. docx, PDF)

Digital Humanities (DH)

- DH gewinnt an **zunehmender Bedeutung** innerhalb der Geistes- und Kulturwissenschaften.
- In vielen DH-Projekten werden **digitale Ressourcen** erstellt.
- Diese Ressourcen müssen **wiederverwendbar und interoperabel** sein.
- In solchen Forschungsinfrastrukturen für die Geistes- und Sozial-wissenschaften wie CLARIN und DARIAH können die Wissen-schaftler/innen **gemeinsam** an solchen Ressourcen **mit den gleichen Werkzeugen arbeiten**.
- Die Verwendung von Standards **verbessert diese Arbeit signifikant**.
- *Special Interest Group „**TEI for Linguists**“*, in Zusammenarbeit mit **ISO** TC 37/SC 4




```
<history>
  <origin>
    <stamp type="postmarked">
      <placeName ref="#DesMoines">
        <settlement>Des Moines</settlement>
      <region>Iowa</region>
      </placeName>
      <date when-iso="1908-07-02T11:00">JUL 7 11AM 1908</date>
    </stamp>
  </origin>
</history>
```

```
<div type="back" facs="#noble0337b">
  <div type="left">
    <salute>Dear <persName ref="#HJ">H</persName>.</salute>
    <p>Everybody <lb/>is O.K. Mrs. Butler <lb/>from across the <lb/>street died
      last <lb/>night. Too bad is <lb/>not it?</p>
    <signed>Goodbye <lb/><persName>S. W.</persName></signed>
  </div>
```

```
<div type="right">
  <p>
    <address>
      <addrLine>Miss <persName ref="#HJ">Hattie Jacobs</persName></addrLine><lb/>
      <placeName ref="#Madrid"><settlement>Madrid</settlement><lb/>
      <region>Ia</region></placeName>
    </address>.
  </p>
</div>
```

Beispiel: Laurent Romary

Normierung von Sprachressourcen in der ISO

Der **ISO-Normenausschuss ISO TC 37/ SC4** „Management von Sprachressourcen“ ist …

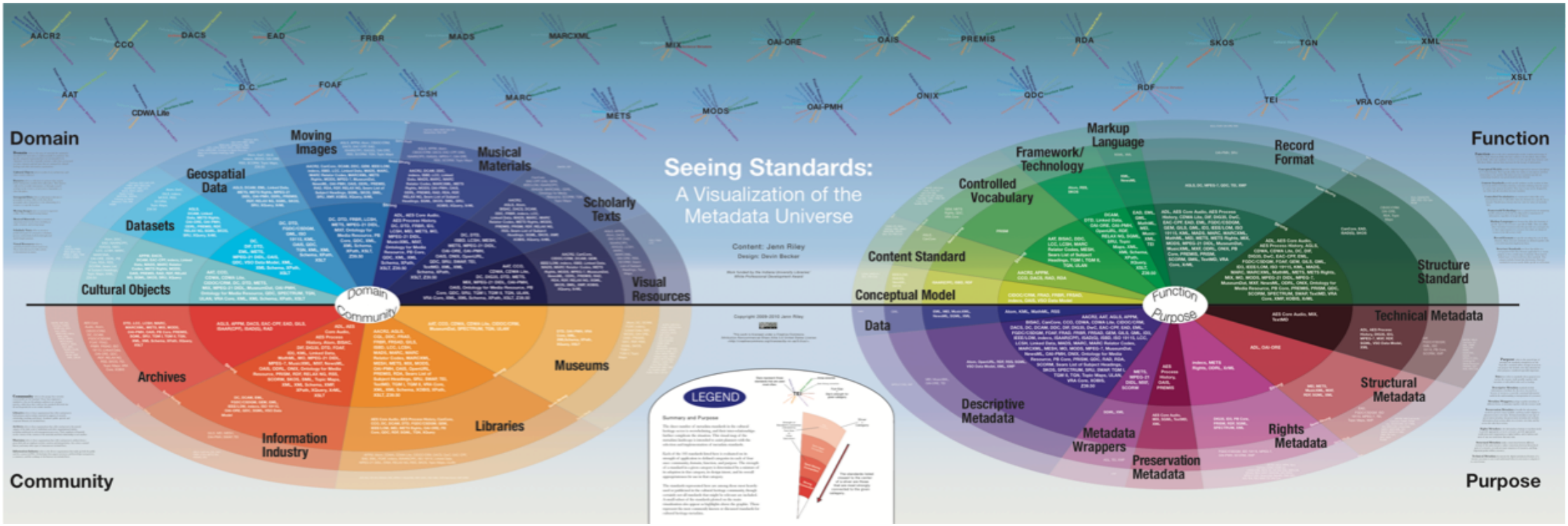
- für **alle Arten von Sprachressourcen** zuständig,
- in seinen Aktivitäten **in mehrere Arbeitsgruppen aufgeteilt**:
  - WG 01 „Grundlegende Beschreibungsmittel und Mechanismen für Sprachressourcen“
  - WG 02 „Semantische Annotation“
  - WG 03 „Multilinguale Informationsdarstellungen“
  - WG 04 „Lexikalische Ressourcen“
  - WG 05 „Workflow für das Management von Sprachressourcen“
  - WG 06 „Linguistische Annotation“

CLARIN und DARIAH in der Standardisierung

- **Offizielle Liaison** vom **CLARIN ERIC** und **ISO TC37 SC4**
- **Laurent Romary** (bis 09/2018 einer der Direktoren des **DARIAH ERIC**)
  - Chairman von ISO TC 37
  - TEI: Mitglied im *Board of Directors*
- **Andreas Witt** (CLARIN-D) leitet eine Arbeitsgruppe der ISO
- **Piotr Bański** (CLARIN-D):
  - Co-Leitung TEI *Special Interest Group „TEI for Linguists“*
  - Sprecher das europäischen *CLARIN Standards Committee*

Metadaten, Normdaten, Ontologien

Standards nach Kategorien



Seeing Standards: A Visualization of the Metadata Universe
Jenn Riley, Devin Becker
CC-BY-NC-SA
http://jennriley.com/metadatamap/

FAIR-Prinzipien als Basis

To be Findable:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
  F2. data are described with rich metadata.
  F3. (meta)data are registered or indexed in a searchable
  F4. metadata specify the data identifier

To be Accessible:

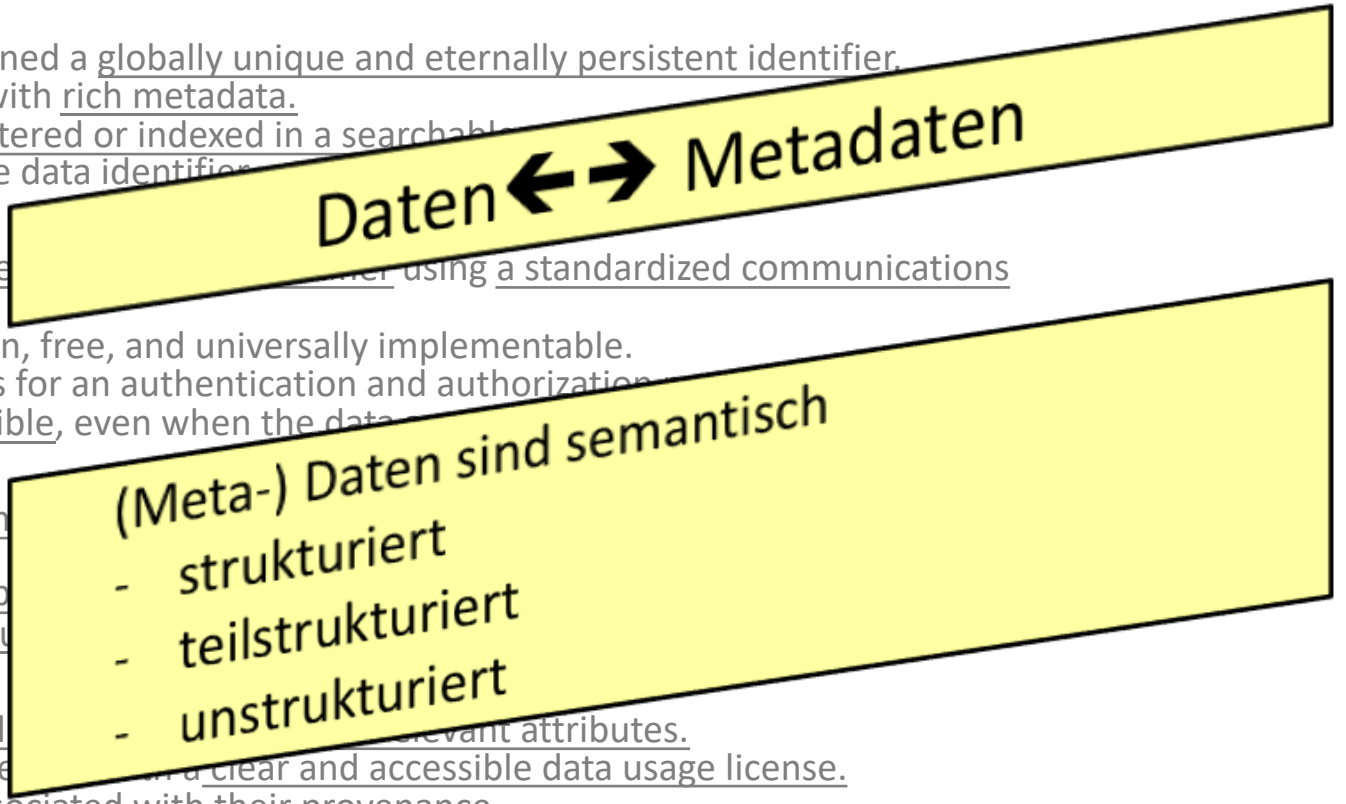
- A1 (meta)data are retrie … using a standardized communications protocol.
  A1.1 the protocol is open, free, and universally implementable.
  A1.2 the protocol allows for an authentication and authorization
  A2 metadata are accessible, even when the data

To be Interoperable:

- I1. (meta)data use a form … representation.
  I2. (meta)data use vocab …
  I3. (meta)data include qu …

To be Re-usable:

- R1. meta(data) have a pl … levant attributes.
  R1.1. (meta)data are rele … a clear and accessible data usage license.
  R1.2. (meta)data are associated with their provenance.
  R1.3. (meta)data meet domain-relevant community standards.



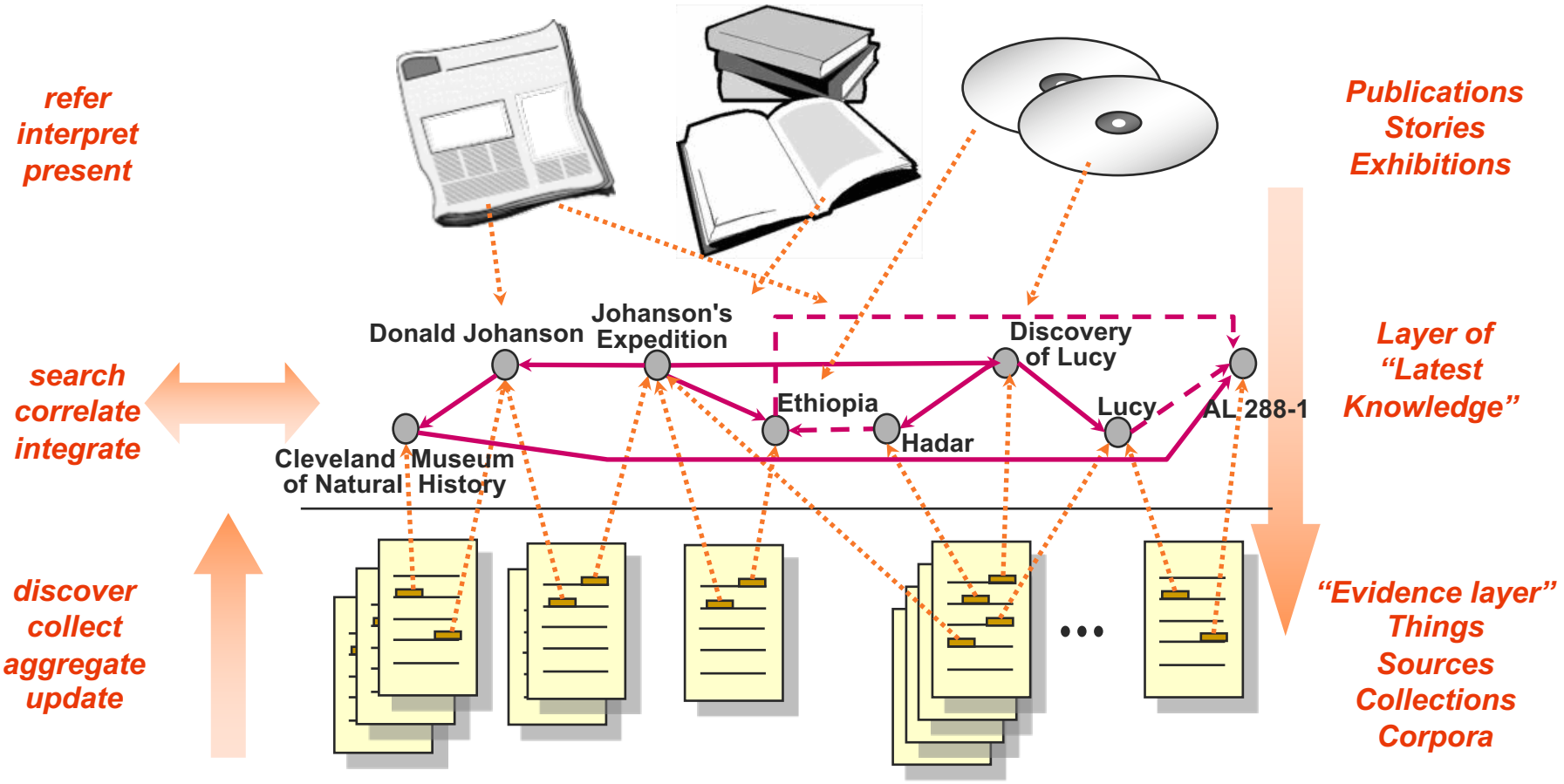
https://www.force11.org/group/fairgroup/fairprinciples

Kontextualisierung von Information

http://www.gettyimages.com/detail/112912339/McClatchy-Tribune.
Florence Thompson, the 'Migrant Mother' in Dorothea Lange's famous 1936 photo, holds up her likeness during an interview, October 10, 1978.

The CIDOC CRM

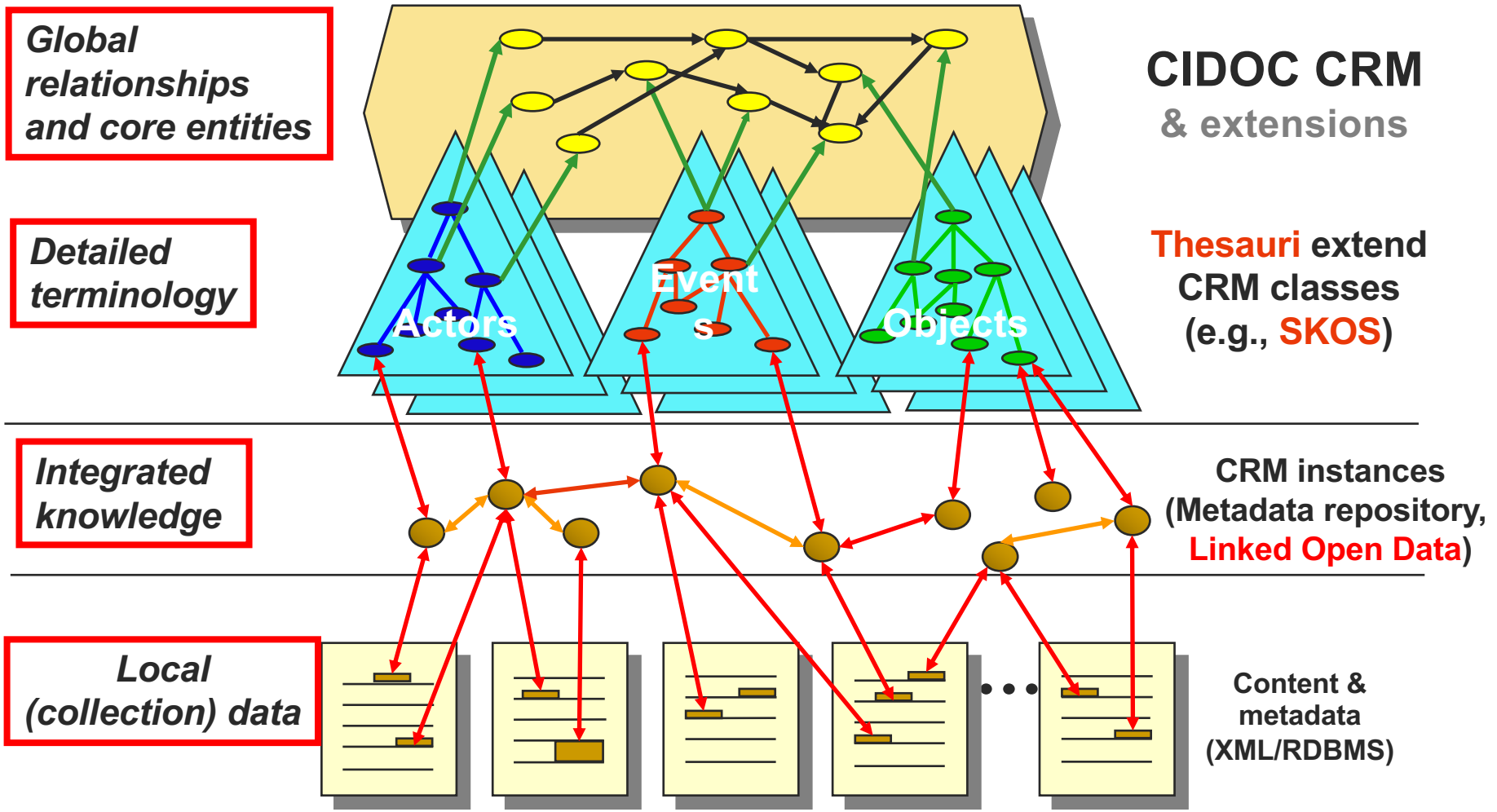
The Scholarly Process



Source: http://www.cidoc-crm.org/sites/default/files/FRBR_for_TPDL.ppt, p58

The CIDOC CRM

Information Integration: Relations & Terms



Source: http://www.cidoc-crm.org/sites/default/files/FRBR_for_TPDL.ppt, p59

Standards und Schnittstellen (Ausschnitt)

- *Referenzontologie*
  CIDOC-CRM (ISO 21127) + Erweiterungen (insbesondere zu FRBR)
- *Deskriptive Metadaten*
  TEI Profile, MARC, MODS, EAD, LIDO, …
- *Normdaten*
  Vielfach lokale Ausprägungen. GND gewinnt an Bedeutung.
  Darüber hinaus z.B. Geonames, Getty Vokabulare (AAT, TGN), Iconclass, …
  VIAF und Wikidata als Koreferenzierungsprojekte.
- *Schnittstellen zur Präsentation*
  Textliche Beschreibungsdaten: OAI-PMH, OData, SPARQL
  Bild, AV-Daten, 3D: IIIF International Image Interoperability Framework
- *Technische Datenformate*
  Je nach Datentyp teils hoch standardisiert (z.B. TIFF), teils noch weitgehend unstandardisiert (z.B. Remote Sensing).

Gutenbergbibel (B42), Johannes Gutenberg, Johannes Fust, Peter Schöffer, um 1454, SUB Göttingen.
http://www.gutenbergdigital.de/ausstellung2018/

Vielen Dank!