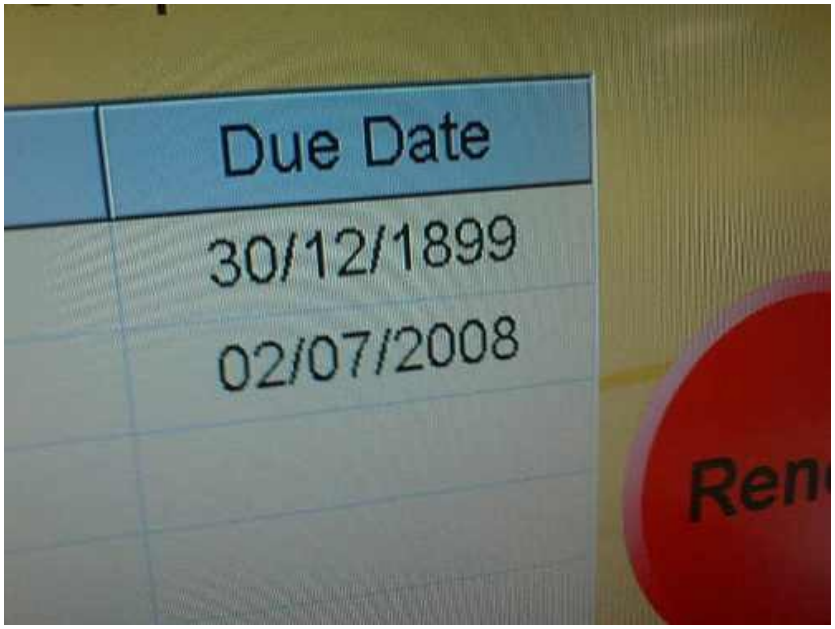
HOCHSCHULE DER MEDIEN

# Ausleihdaten aus Bibliotheken als Linked Open Data publizieren und nutzen




# Publishing and consuming library loan information as linked open data

Prof. Magnus Pfeffer, Stuttgart Media University
pfeffer@hdm-stuttgart.de

# Overview

- Motivation: Use cases
- Types of data
- Loan transaction data
- Modelling the data in RDF
  - Thinking from data
  - Thinking from usage
- Conclusion
- Looking ahead

# Use case: Retrieval

HOCHSCHULE DER MEDIEN

- Loans as a quality indicator
  - More loans → more interest → higher relevance
- „Loan relevance“ is orthogonal to traditional relevance criteria
  - Desireable with result sets of almost identical content
  - Example search:
    - „Introduction to algebra“
    - „Mathematics for students of social sciences“

# Retrieval: Example

HOCHSCHULE DER MEDIEN

## Product Details

**Hardcover:** 656 pages
**Publisher:** Simon & Schuster; First Edition edition (October 24, 2011)
**Language:** English
**ISBN-10:** 1451648537
**ISBN-13:** 978-1451648539
**Product Dimensions:** 9.3 x 6.3 x 1.7 inches
**Shipping Weight:** 2.4 pounds (View shipping rates and policies)
**Average Customer Review:** ★★★★½ (520 customer reviews)

**Amazon Best Sellers Rank:** #2 in Books (See Top 100 in Books)
#1 in Books > Business & Investing > Industries & Professions > **High-Tech**
#1 in Books > Computers & Technology > Business & Culture > **Biographies**
#1 in Books > Biographies & Memoirs > Professionals & Academics > **Business**

# Implemenation ideas



- Aggregation
  - Aggregate loan data on item level
  - Normalize loan data from different locations
  - Aggregate loan data on title level
- User interface
  - Displaying loan statistics in the short result list
  - Sort short result list by loan statistics
  - Using loan statistics in ranking

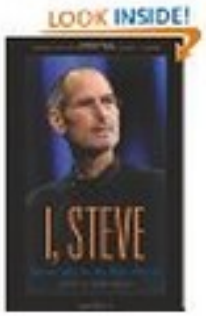
# Use case: Resource discovery

- Assumption: Items loaned together are correlated
  - May not hold true in all instances
  - But certainly on an aggregated level
- Present „similar titles“ based on correlation information

# Resource discovery: Examples

HOCHSCHULE DER MEDIEN

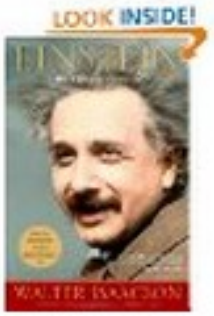
## Customers Who Bought This Item Also Bought




I, Steve: Steve Jobs in His Own Words by George Beahm

(20)

$8.76




Einstein: His Life and Universe by Walter Isaacson

(311)

$12.89


The Presentation Secrets of Steve Jobs: How to B... by Carmine Gallo

(74)

$14.93

BibTip Andere Benutzer fanden auch interessant:

- Apple: die Geburt eines Kults / Moritz, Michael , 2011
- Steve Jobs / Isaacson, Walter , 2011

# Implementation ideas

HOCHSCHULE DER MEDIEN

- Analysing loan history
  - Same patron
  - Similar start of loan
    - Or: overlapping loan periods

    → Groups of titles → pairs of correlated titles
- Aggregation
  - Summation of correlation counts
  - Generation of correlation groups for each title
- Presentation
  - Top-n most correlated titles
    - Consider minumum correlation to suppres spurious results

# Types of data in library information systems

# Types of data: master data

HOCHSCHULE DER MEDIEN

- Properties
  - Changes and edit are rare
  - Slowly growing dataset
  - Stable identifiers
- Examples:
  - Business information systems
    - Product information
    - Customer contact information
    - Vendor information
  - Library information systems
    - Catalogue entries
    - Patron contact information

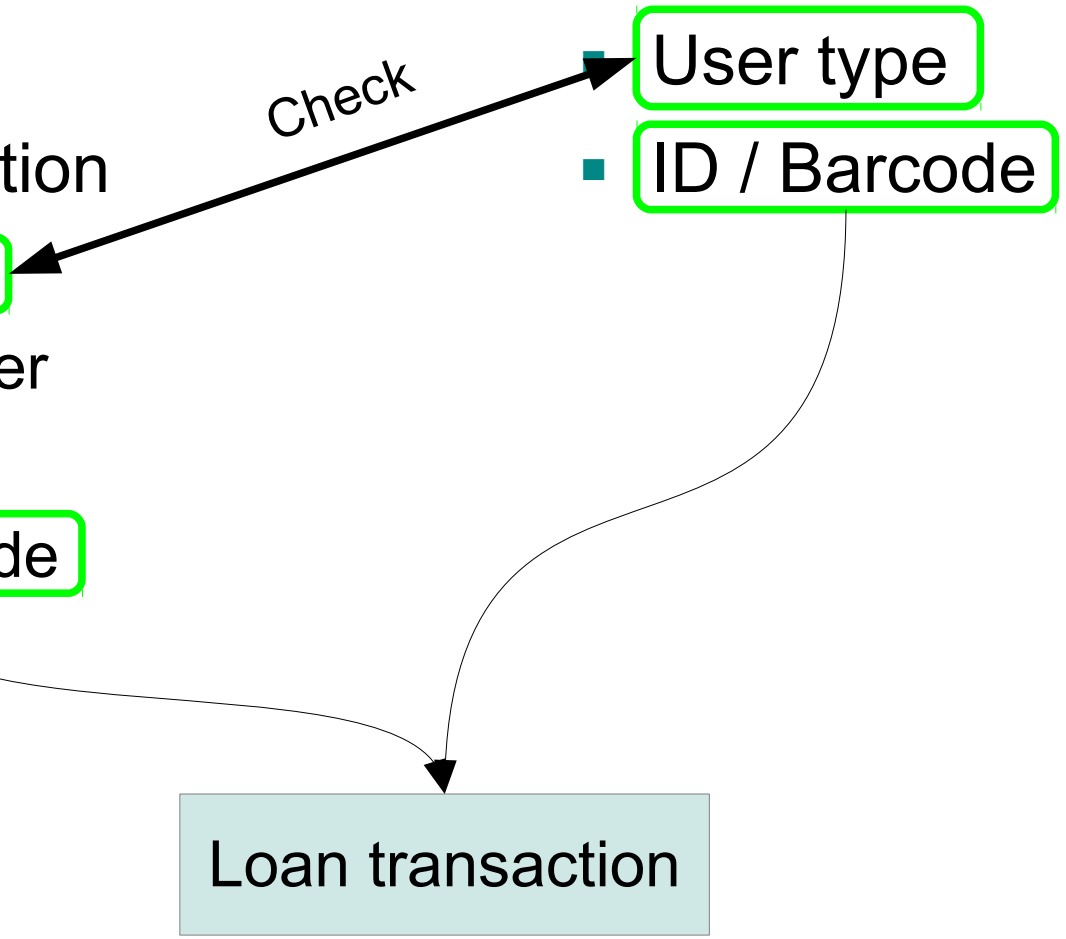
# Types of data: dynamic data

HOCHSCHULE DER MEDIEN

- Properties
  - Subject to changes
  - Quickly growing dataset
  - Usually a combination of master data entries
  - Usually no own identifier
- Examples:
  - Business information systems
    - Sales transaction details
  - Library information systems
    - Media purchase transaction details

# Loan transaction: ILS view

- Bibliografic entry
  - Series
  - Author
  - Title, …
  - Item information
    - Item type
    - Call number
    - Location
    - ID / Barcode
- Patron information
  - Name
  - Adress
  - User type
  - ID / Barcode



# Loan transaction: ILS data

HOCHSCHULE DER MEDIEN

- Current loan
  - User ID
  - Item ID
  - Timestamps (start)
    - Order / Hold request
    - Pickup ready
    - Pickup by patron
  - Loan due time
  - Loan extensions
    - Timestamps / Numbers
  - Overdue escalation / overdue messages sent

# Loan transaction: ILS data

- Completed loan
  - As before
  - Timestamps (end)
    - Item returned by patron
    - Return to stacks

**Privacy**

To protect the privacy of the patrons, the information on completed loans is usualy anonymised after a short period of time.

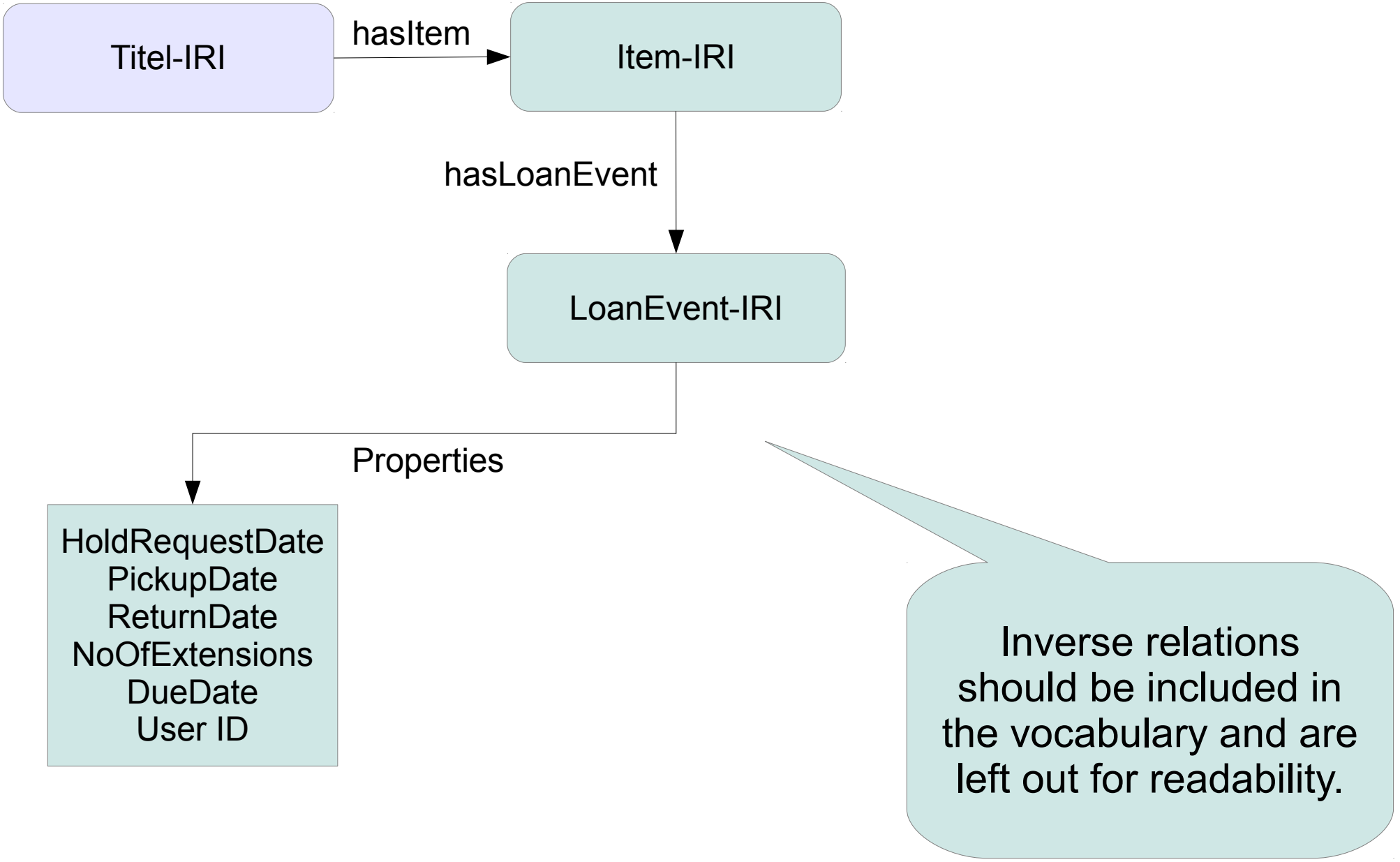
# Modelling loan transaction data in RDF

# Approach 1: Consider the data

HOCHSCHULE DER MEDIEN

- Loans as **events**
- Minimum elements
  - Start Time
  - End Time
  - Item ID
  - **Anonymised** User ID or User Type
- Additional elements
  - Differentiated timestamps
  - Number of extensions

# Event-based model



# Properties



- Easy implementation
  - Existing data can be used 1:1
- Highly granular data
  - Each loan event needs an individual IRI
  - RDF consumers need to aggregate data themselves

    → complex graphs

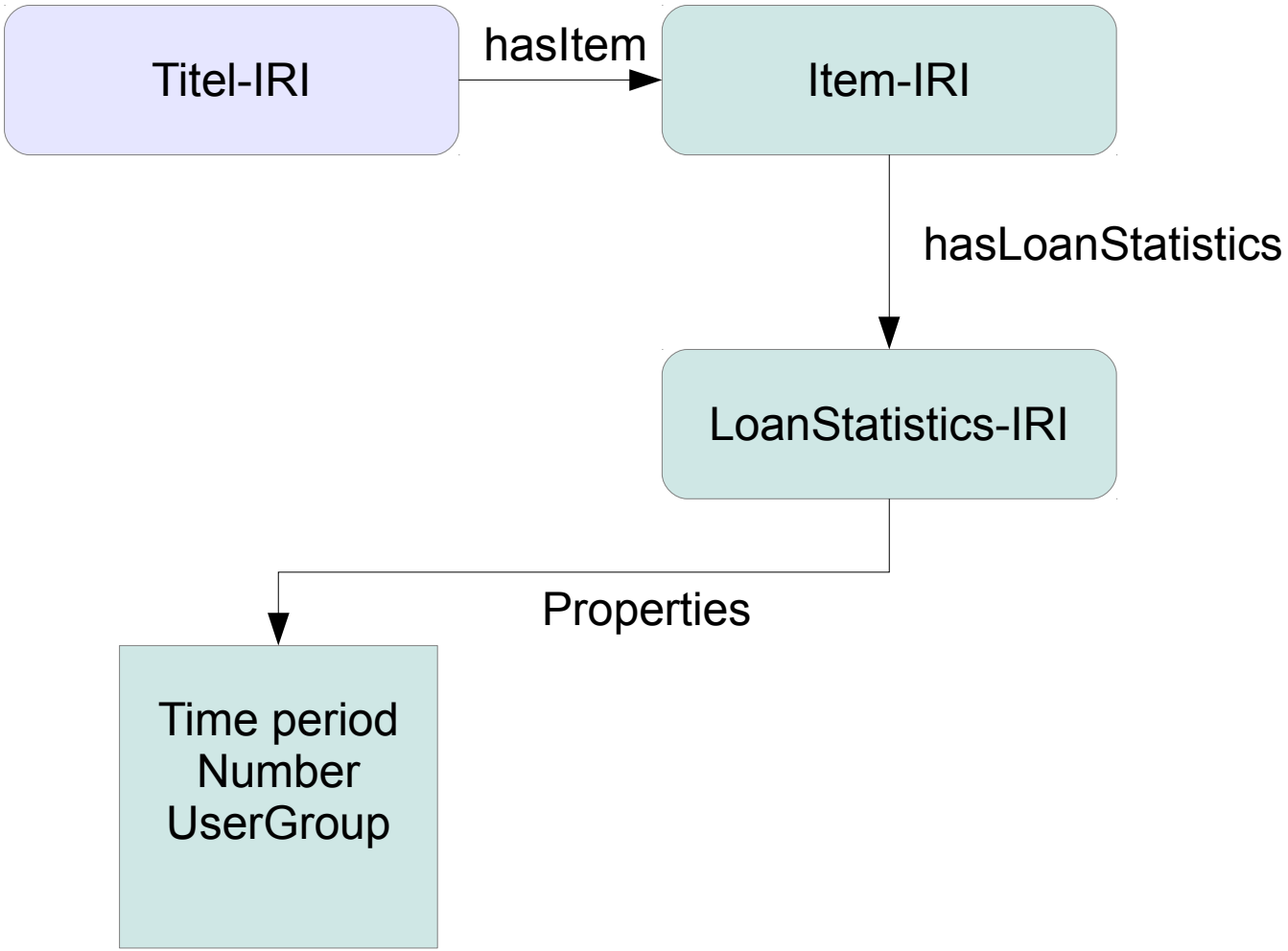
    → costly queries

# Approach 2: consider the application

HOCHSCHULE DER MEDIEN

- Loans as **statistics**
  - Loans per year / month / week
  - Differentiation by user type or location

# Statistics-based model



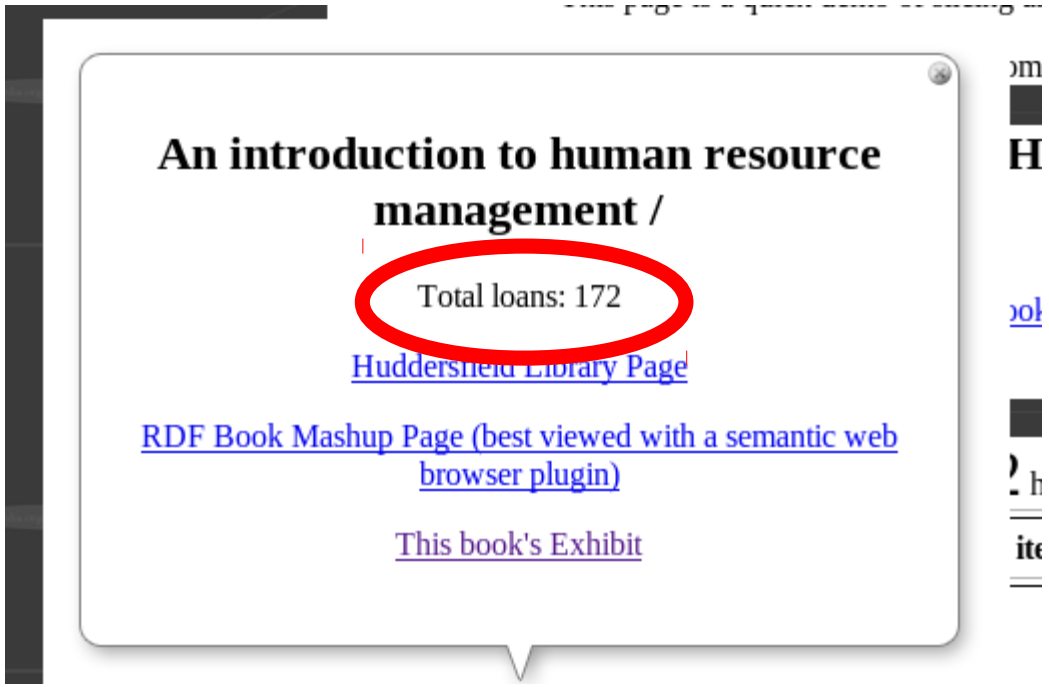
# Properties



- Harder implementation
  - Need to decide on statistical units
  - Need to normalize different loan conditions
- Less granular data
  - Less IRIs
  - Simple queries
  - Information on correlated titles is lost

# So that's it?

- University of Huddersfield Library (UK)
  - Published circulation data as open data
    - http://library.hud.ac.uk/data/usagedata/
    - Used in a semantic catalog prototype



# Loan conditions at UB Mannheim

HOCHSCHULE DER MEDIEN

- Closed stacks
  - Orders or hold requests from the OPAC
  - 4 weeks default loan period
    - Extensions possible
    - 2 weeks if there are other requests
- Textbook collection
  - No online orders or hold requests
  - 2 week default loan period
    - No Extensions possible
- Open access areas
  - No online orders or hold requests
  - 6 month default loan period – staff only

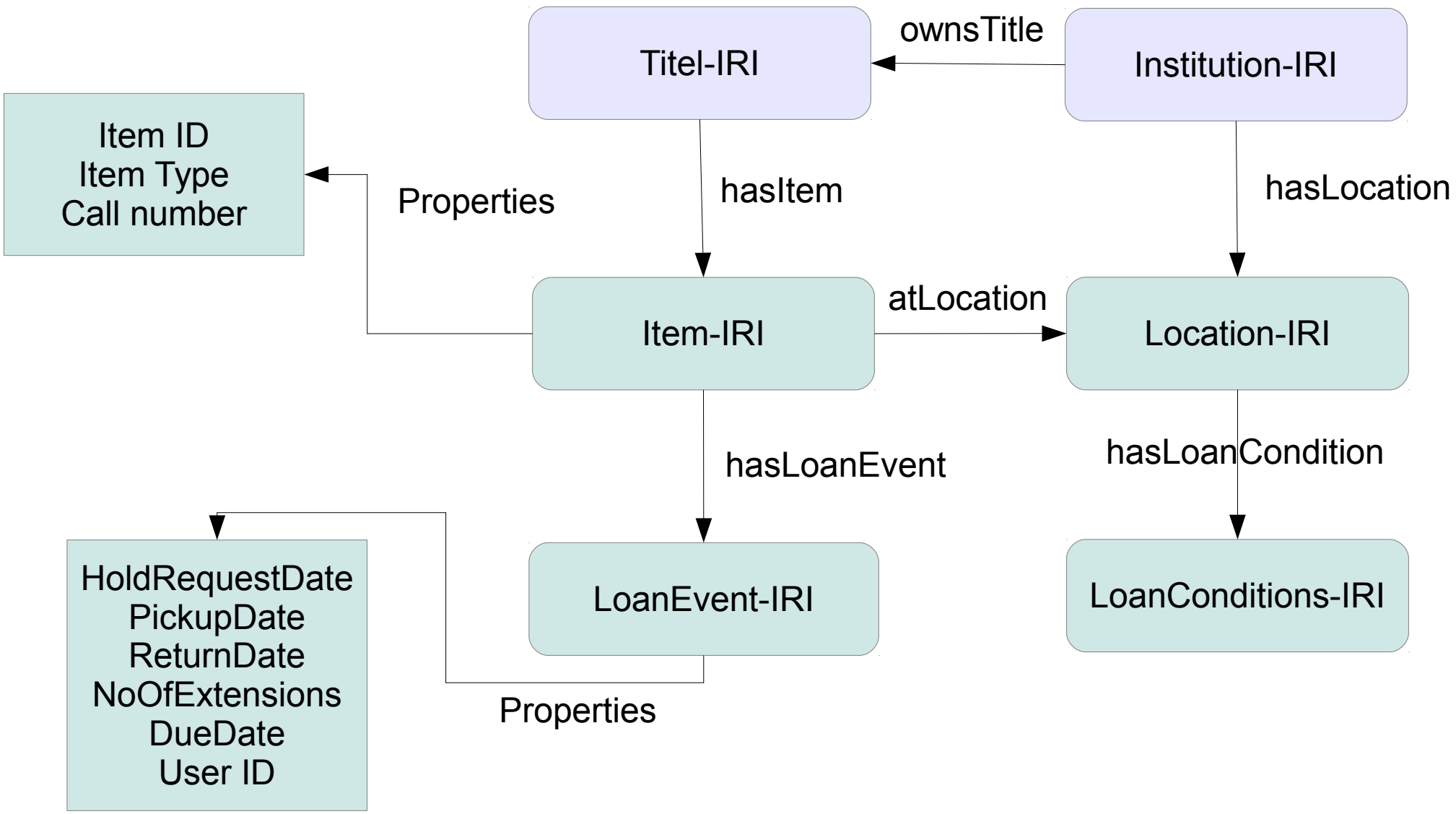
And there are several additional collections
–
with even more diverse conditions

# Normalizing loan data

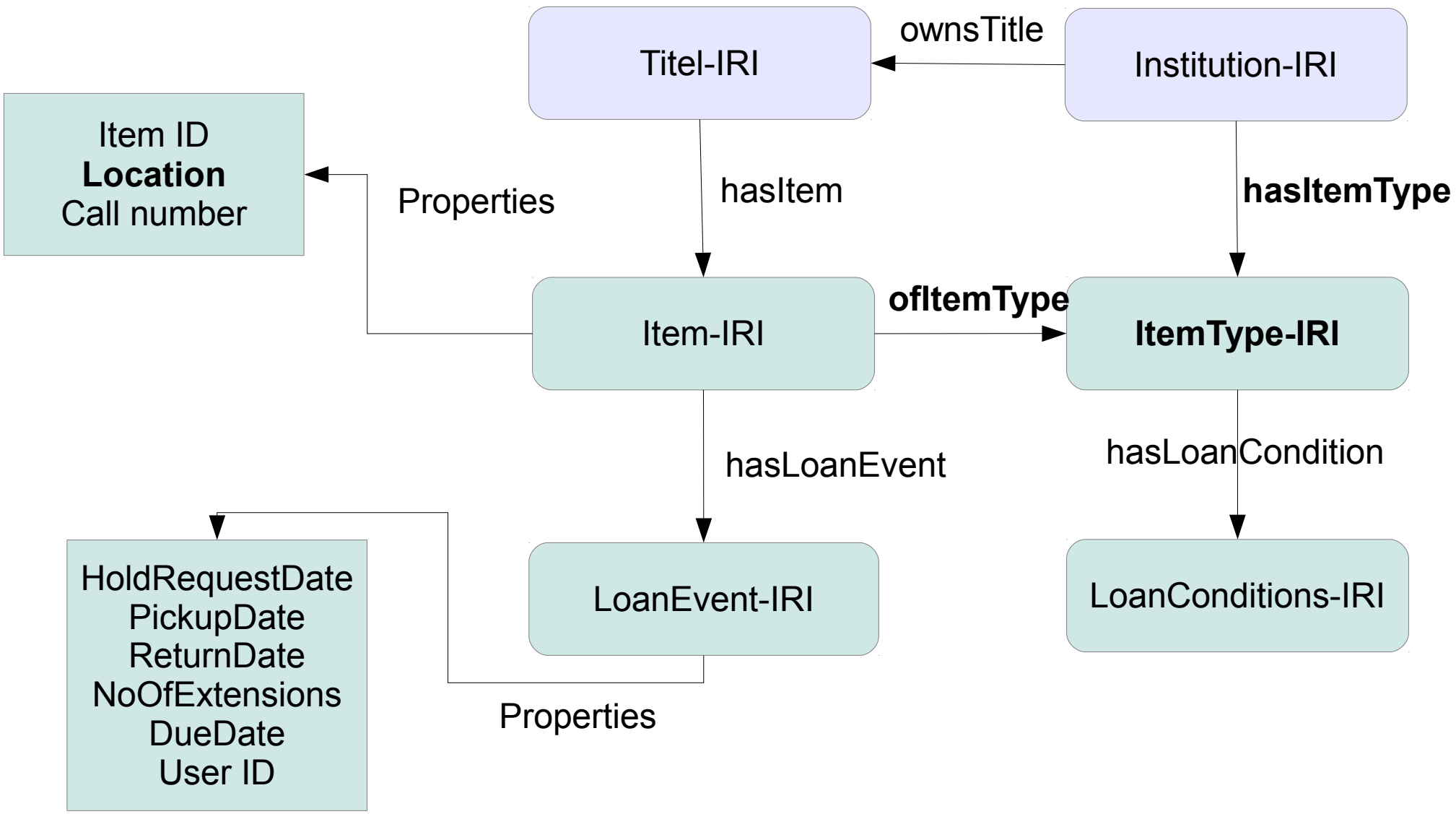
HOCHSCHULE DER MEDIEN

- Different default periods
  - Is a 8 week loan “more” than a two week loan?
- Multiple hold requests on loaned items possible
  - These items are on loan permanently – is this the same as an item on year-long loan by a single staff patron?
- Loan-and-return items
  - Patrons cannot browse books from the closed stacks
  - Browsing is done on the counter and discarded items are returned promptly – should these be counted as loans?

# Model revisited



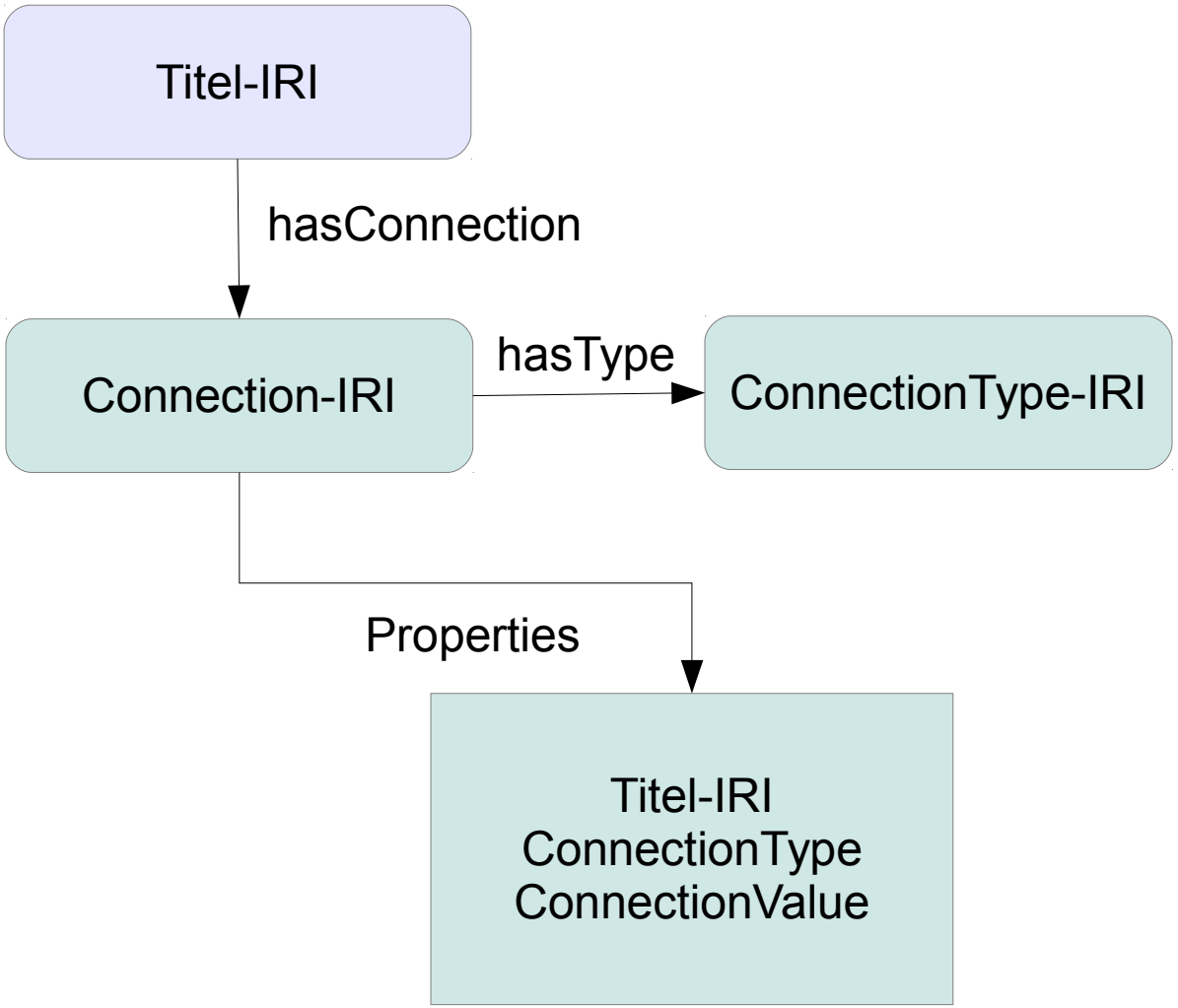
# Model revisited



# Modelling title correlation

# Correlation model

- Correlation connects two titles
  - Type and strength of connection differ



# Implementation



- Aggregation based on RDF published events possible
  - But extremely complex
  - Anonymized data can make it impossible

→ Should be published separately from loan events

# Conclusion

# Conclusion

HOCHSCHULE DER MEDIEN

- There are evident usage scenarios for loan data
  - Retrieval
  - Resource Discovery
- Comparing loan statistics between institutions is hard
  - Wildly diverging loan conditions based on user and item type and/or location
  - Little consensus in the community
- Modelling in RDF is possible
  - Complex model with many IRIs
  - The model theoretically satisfies both use cases
    - But for performance and privacy reasons loan data and correlation data should be published independently

# Looking ahead

HOCHSCHULE DER MEDIEN

- We are currently
  - Creating a complete vocabulary definition
  - Converting several months of loan data for UB Mannheim
    - As granular events
    - As aggregated sums according to the german library statistics (DBS )
- We will
  - Evaluate the run-time complexity of the aggregation based on RDF
  - Create and publish correlation scores based on circulation data mining

# And even further...

HOCHSCHULE DER MEDIEN

- Presentation
  - We are thinking about a Javascript semantic plugin
    - For browsers or embedding into OPAC pages
    - Show links, aggregate additional information, etc.
- Data
  - Loan data is interesting, but relatively spars
  - The real action happens in the OPAC
    - Interest indicators from clicks to full view
    - Correlation data from search sessions
  - Harvesting this information is possible by anonymous session tracking

Thank you for listening.

# Discussion