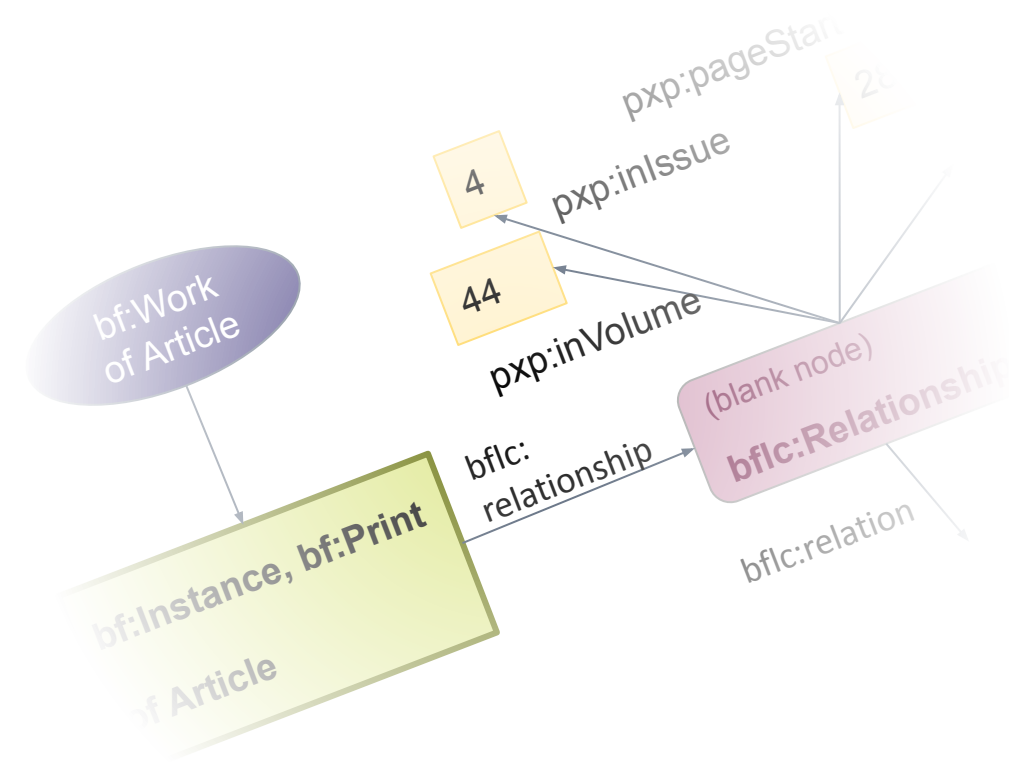


# BIBFRAME for academic publishing in psychology

A case study at Leibniz Institute for Psychology (ZPID)



**Tina Trillitzsch, Leibniz Institute for Psychology (ZPID)**  
[ttr@leibniz-psychology.org](mailto:ttr@leibniz-psychology.org), <https://orcid.org/0000-0001-7239-4844>

Semantic Web in Libraries 2022 (SWIB22), Nov. 29, 2022

**SWIB22**  
Semantic Web in Libraries

# 1. Intro & Outline

- ZPID: publicly funded Open Science institute – infrastructure for psychology: for **researchers** from/in German-speaking countries, but also for **practicing** psychologists, **teaching** (students & professors), information for **laypeople**.
- most well known service: **PSYINDEX** - curated reference database of psychological **literature** and diagnostic **tests** – both **scholarly** articles and books, and selected **popular** science materials.
- **The plan:** New cataloging software, new search portal, more authority records, Linked Open Data – modeled in Bibframe.

## Outline for this talk:

- What we want: **the future**
- What we have: status quo – **now/before**
- **Getting from before to “the future”:** models, software, tools
- Examples for psychology-/academia-specific **Bibframe** modeling issues & how we solved them: author contributions (position in sequence/first author, affiliations), “aboutness” - psychology-specific keyword indexing, modeling parts/journal articles, “isness” - “it’s a Journal Article”, modeling journals.

## 2. What we want and what we have, & how to get there

### a) What we want

- New, modern, stable & maintainable software for **cataloging & indexing**, meeting our specific needs: **PSYNDEX2**
- New, modern search portal offering what users have come to expect from commercial literature search engines, but **non-commercial**, public, open & with detailed, **high-quality topic indexing**: **PsychPorta**
- ZPID's **Open Science** mission: Publish **Linked Open Data** that can be reused, recombined into new applications, is interoperable
- **Authority records** for people (authors/editors), organizations (author affiliations, publishers,...), places, conferences to bundle publications by contributing parties.
- Publications modeled such that “**versions**” are bundled together in one place, allowing users to choose which one to access (probably the free one :) – > **Bibframe Works/Instances**
- **Integrate** some ZPID services currently in separate “**silos**”

## 2. What we want and **what we have**, & how to get there

### b) **What we have**

- **Aging cataloging software** for **PSYINDEX** - Cuadra Star system (*not* based on MARC21!) – frequent errors, ancient user management, data models, file formats & DB structures; changes: external company – > slow turnaround
- **PsychAuthors** (web database of author profiles) & **PsychArchives** (domain-specific repository) needing integration
- **PubPsych**, aging web-based search portal, no API, limited open data reusability, hard to maintain
- **Specific requirements** for PSYINDEX cataloging and indexing: great width & depth!
  - Books, chapters from books, journal articles, journals, dissertations, test instruments; both **published & gray** literature, print and electronic (& old analog media, microfiches...).
  - Deep **psychology-specific indexing**: topics & classifications, weighted topics, study methodology, studied population, linking preregistration, research data.

## 2. What we want and what we have, & **how to get there**

### c) **How we get there**

#### **In-house Development:**

##### **Currently:**

- **PSYINDEX2** cataloging software (PHP/Symfony, RDBMS)
- In parallel, tightly coupled: **BIBFRAME**-based **data model** for publications + our own extensions
- Converting controlled vocabularies to SKOS

##### **2023-2024:**

- **PsychPorta** web search portal, based on Bibframe Linked Open Data
- **Migration** of existing data, regular **conversion** to LOD, triplestore hosting
- LOD support tools & pages, SPARQL endpoint

#### **Tools we use:**

- [Skosmos](#) for hosting controlled SKOS vocabularies (browsing & API for indexing in PSYINDEX)
- [Annif](#) for automated PSYINDEX keyword suggestions
- [Sparql Anything](#) for conversion & data migration (SPARQL *CONSTRUCT* queries generate RDF)
- [Apache Jena](#) for LOD hosting (TBD Triplestore) & SPARQL endpoint (Fuseki)
- Elasticsearch index powering PsychPorta search

### 3. Bibframe Section:

Some psychology-/academia-specific Bibframe modeling issues & how we solved them

- author **contributions** (position in sequence/first author, affiliations, authority links),
- psychology-specific keyword **indexing** and classifications (“**aboutness**”)
- parts/wholes & **serial relationships**: example: **journal article** to journal
- describing kinds of “**isness**” – bibliographic level or “it’s a Journal Article!”, some genre
- modeling continuing/serial publications, example: **journal**
- Finally: Reflections on Bibframe – or rather “airing of grievances”

### 3. Bibframe: **Contributions** – **Author position and affiliations**

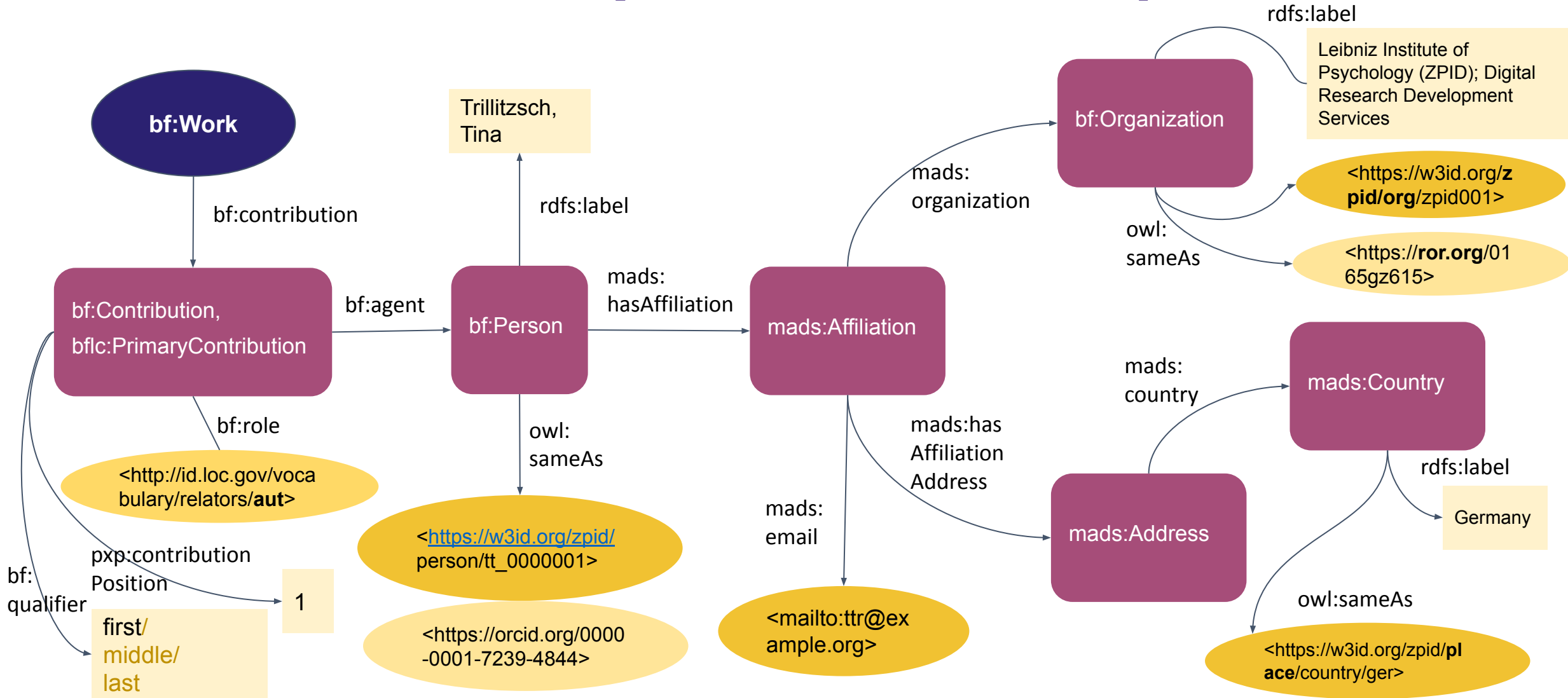
#### Problems:

- Qualified information about contributors
- Author affiliation, correspondence email, address/country
- Author position in sequence (to differentiate first and last, & for ordered display)
- Agents, affiliated organizations, countries need **BOTH names** (when work created) **AND links to authority records**

#### Solutions:

- Work > bf:contribution >> bf:Contribution > bf:agent, bf:role
- [MADS/RDF Ontology](#) (LoC)  
mads:hasAffiliation >> mads:Affiliation  
> mads:organization >> bf:Organization  
> mads:email  
> mads:hasAffiliationAddress >> mads:Address > mads:country >> mads:Country
- pxp:contributionPosition “1” etc. (our own subproperty of bf:qualifier & schema:position) AND bf:qualifier for “first”, “middle”, and “last”
- Blank nodes with rdfs:label + owl:sameAs links to authority entities (our own, ORCID, ROR)

### 3. Bibframe: Contributions – Author position and affiliations, authority links





### 3. Bibframe: **psychology-specific subject indexing/Keywords**

#### **New classes for weighted topics, study-specific keywords:**

- We mark some terms/topics as “**weighted**” (more important) for *this* Work. – > our own *bf:Topic* subclass **pxc:WeightedTopic** plus **owl:sameAs** link to the canonical SKOS concept.
- A place for our classifications & vocabularies describing the **study** within a Work: **methodology** used, **age & location** of studied population – > subclasses of *bf:Classification* (**pxc:ControlledMethod**), *bflc:DemographicGroup* (**pxc:AgeGroup**) and *bf:GeographicCoverage* (**pxc:PopulationLocation**).

```
<W> bf:subject [a bf:Topic,pxc:WeightedTopic;
  rdfs:label "Ontologies"@en,"Ontologien"@de;
  owl:sameAs
  <https://w3id.org/zpid/vocabs/terms/35365>;
  bf:source <https://w3id.org/zpid/vocabs/terms/>];

  bf:classification [a bf:Classification,
pxc:ControlledMethod;
  rdfs:label "..."@en; bf:code "10310";
  owl:sameAs <...>; bf:source <.../methods/>];

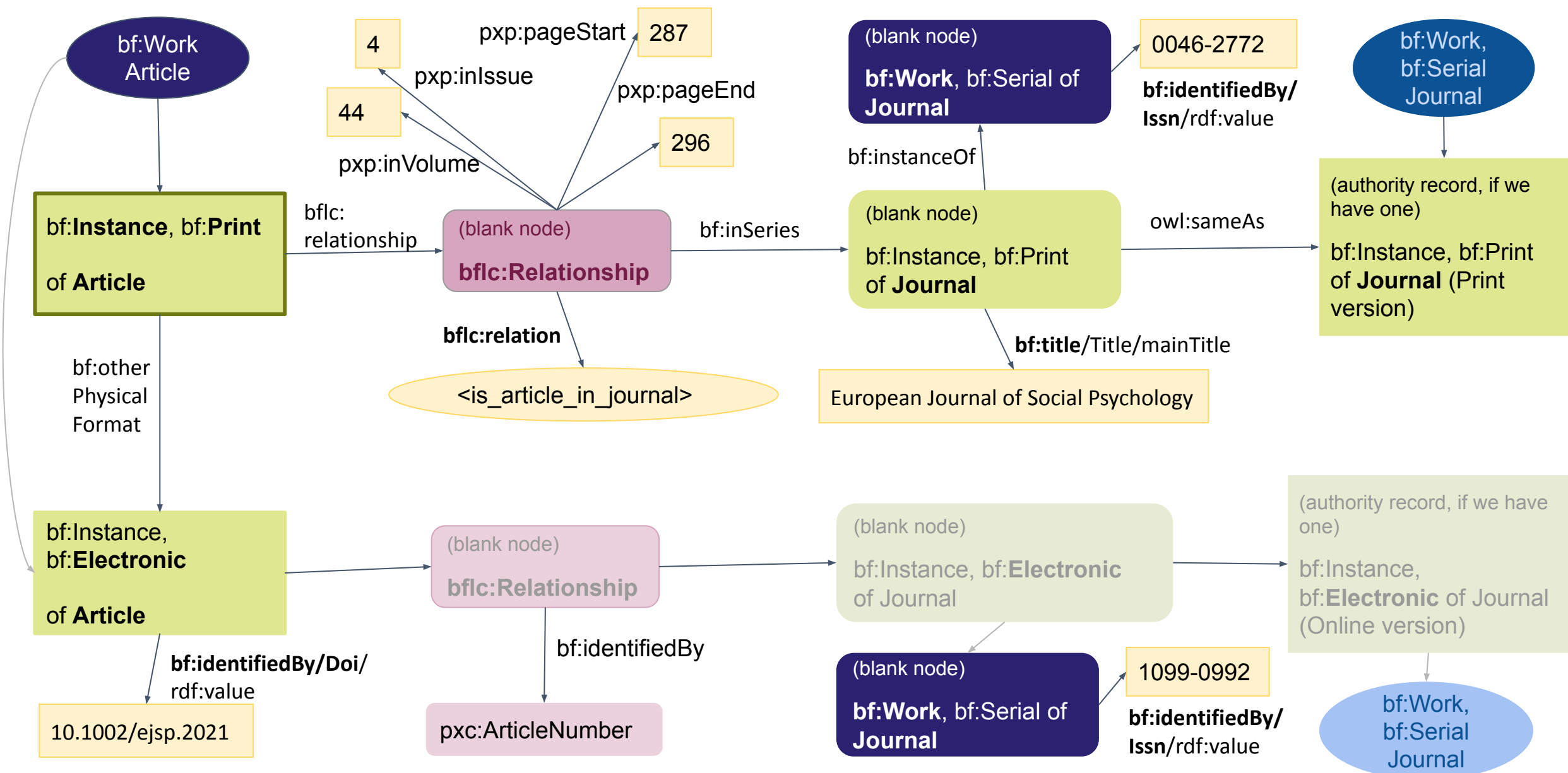
  bflc:demographicGroup [a bflc:DemographicGroup,
  pxc:AgeGroup;
  rdfs:label "Adulthood"@en; owl:sameAs <...>;
  bf:source <.../AgeGroups>];

  bf:geographicCoverage [a bf:GeographicCoverage,
pxc:PopulationLocation;
  rdfs:label "Germany"@en; owl:sameAs <...>;
  bf:source <.../countries>];
.
```

### 3. Bibframe: **Parts and Wholes/Serial Relationships** - Chapters & **Journal Articles**

#### **Part-Whole & Series Relationships: Instance to Instance!**

- We disagree with LoC conversion specs & RDA textbooks that say **part-whole** relationships are **between Works** – only Instance to Instance make sense:
- How else to say that, e.g., a **Print chapter**'s Instance is part of to the **Print Instance** of a **book**, not its *Electronic Instance*?
- Still possible to infer part-whole/serial relationship “**upwards**” from between **Instances** to those between **Works** – but not vice versa!
- We needed to make **qualified statements** about **location** of parts **within** the whole or serial, e.g., article in journal (page range, article no, volume, issue – our own subproperties of bf:part) – > we use **bflc:relationship/bflc:Relationship** and bflc:relation/Relation instead of direct bf:relatedTo (or subproperties)
- We wanted more specific relations than just the bf:relatedTo subproperties offers – > we use our own SKOS **vocabulary of more specific bflc:Relations** like “*article in journal*”



### 3. Bibframe: **“Isness”** – Stating *what* a resource actually is. Bibliography Level, Genre.

#### **Bibliographic level/publication type, or “I just want to say it’s a Journal Article or a Book!”**

- We needed to state whether something is a **component part** or a standalone item, and what kind: book **chapter**, **journal article**, book... but found no way in Bibframe.
- We argue: **Instance-level statement**, like like part/whole and serial relationships - a scholarly paper (Work) might be issued as an article in a journal, a book chapter, an “author version” PDF in a repository, or all three – 3 Instances of the same Work!
- So, **not bf:Work > bf:genreForm** (which we reserve for **purpose-based isness**: thesis, textbook, scholarly paper...).
- **“Leader 07”** in MARC21 (“bibliographic level”) seemed promising: was turned into **bf:Instance > bf:issuance** (& recently, subclasses of bf:Work – bf:Monograph, bf:Serial, bf:Integrating); however, “monographic component part” (07 a) **wasn’t** carried over in either. Worse: Since 3R RDA, “(mode of) issuance” only to be used for **“single unit”** or **“multiple unit”**!
- **Solution**: Our own, extended Mode of Issuance SKOS vocabulary for Instances – includes “single unit” and “multiple unit” values as top-level concepts.

### 3. Bibframe: “**Isness**” – Stating *what* a resource actually is. Bibliography Level, Genre.

#### Combined SKOS vocabulary hierarchy for setting both “bibliographic level” & mode of issuance of Instance at once:

- **single unit**

- Monograph/Item (standalone)
  - **Authored Book**
  - **Edited Book**
- Component Part
  - **Journal Article**
  - **Chapter/Book Section**
- Integrating (Continuing)
  - Loose-leaf
  - Website
  - Repository

Top level can be inferred and exported through **skos:narrower** relationship, resulting in 3R-conformant **bf:issuance** value of (single unit or multiple unit)

- **multiple unit**

- Successive (Continuing)
  - Serial
    - Journal
    - Magazine
    - Yearbook
  - Series
- Multipart Monograph (Set)

#### Combines:

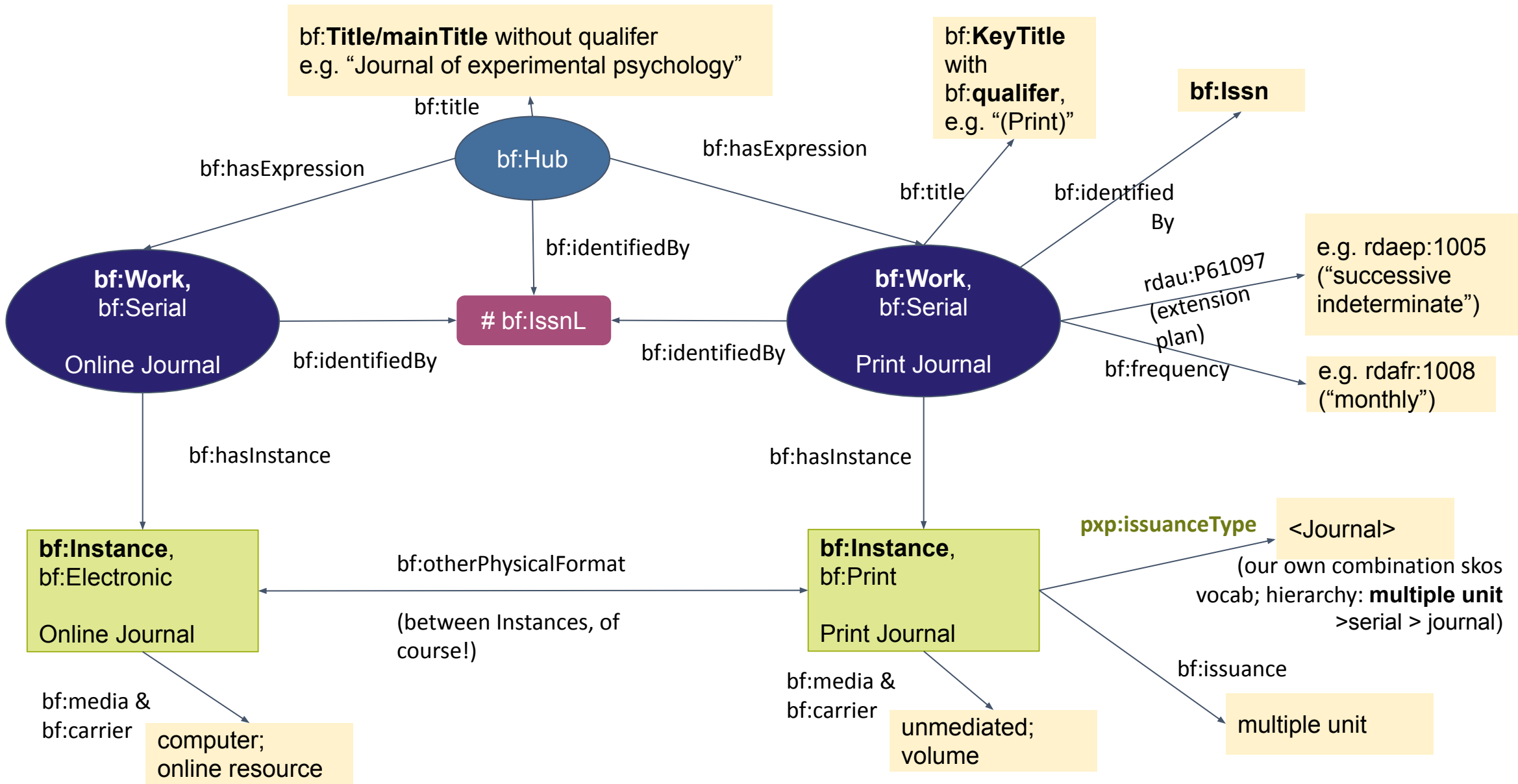
- ❖ new issuance according to 3R RDA (**single/multiple unit**),
- ❖ previously allowed issuance values (serial, integrating, monograph, multipart monograph),
- ❖ missing **monographic component parts** from MARC Leader 07 plus **subtypes** from PSYINDEX field for “bibliographic level”.

### 3. Bibframe: **Modeling Journals** (and other “diachronic works”)

#### Facts and Realizations:

- **One journal** is really one or **more versions** (e.g. Print, Online), each with their own **ISSN**, **each** its own **WEM-locked Work/Instance pair**
- 3R RDA recently decided: ISSN, **frequency of issuance** and new property “**extension plan**” belong to the **Work** part of that pair – following IFLA LRM model for “diachronic works” (=integrating or serial resources like websites, journals, book series, yearbooks).
- (recent?) policy at Library of Congress linked data services seems to be (mostly): A **bf:Hub** with a “collective title” that links the **Works** of all versions via **bf:hasExpression**
- links between Works/versions via **bf:otherPhysicalFormat**
- Issn.org has a slightly [different model](#) with a little Bibframe sprinkled in (not Hub, but a schema:Periodical acting as “**cluster**” for all versions via schema:hasPart).
- Issn.org adds a a “hashed” (referenceable) **bf:IssnL** blank node to cluster, linking to it from all versions (in addition to each Work’s individual bf:Issn).

Putting it all together & adding some corrections (from our view): – >



### 3. Bibframe: **Reflections on Bibframe** / "Airing of Grievances"

The following are a few things we found challenging about Bibframe.

But keep in mind:

We still find Bibframe to be the **best model for our purposes!**



### 3. Bibframe: **Reflections on Bibframe** / “Airing of Grievances”

- You **have to know MARC21** to understand Bibframe – labels, descriptions of classes/properties. **We didn’t**; we have an obscure Cuadra “*STAR XML*” format with unsystematic field names (evolved over many years).
- To understand BF classes and properties & how to properly **combine** them, we needed to **study in depth**:
  - MARC2BF conversion specs & scripts, find out **which MARC field** to look for in the MARC html documentation
  - reference **implementations** like Library of Congress’ linked data (and implementations from many SWIB presentation slides from past years – thank you!)
- Bibframe often feels **underspecified** – “*use with Work or Instance*”,... ok, I guess, I’ll do what I think is best, then – but how can results be **interoperable**?
- LoC implementation and thus, **Bibframe is**, in practice, still pretty much **WEM-locked** (always exactly 1 work and 1 instance together).
- But we *want* to use the benefits of separating **Work** and having multiple **Instances!** We *care* where properties go, please tell us!
- Sometimes we *are* told (conversion specs & scripts), but **decisions seem arbitrary** (due to WEM lock) & we **disagree** with them.

# 4. Summary

- What we want: **the future**
- What we have: status quo – **now/before**
- **Getting from “before” to “the future”**: models, software, tools
- **Examples** for psychology/academia-specific **Bibframe** modeling issues & how we solved them, PLUS some airing of Bibframe grievances ;)

Thank you for listening!

Questions?

Examples of our Bibframe solution patterns in Turtle notation:

<https://github.com/leibniz-psychology/zpid-bibframe-implementation>