

BIBFRAME Shapes: Validating our Approach



presented by

Nancy Lorimer (Stanford University)
Sinopia representative to BIG

17 September 2024
BIBFRAME Workshop in Europe

Outline

- Background
 - Charge
 - Governance
 - Potential Work
 - Subgroups
- Work Plan 2024 (Partial)
- Validation Workflow
- Documentation/Sharing
 - PCC Wiki
 - BIG Github repository
- Next Steps





Background



What is BIG??

The

BIBFRAME

Interoperability

Group

- established by the Program for Cooperative Cataloging Policy Committee (PCC PoCo) to address interoperability issues in BF internationally
- full members include:
 - Cornell University
 - IndexData
 - Library of Congress
 - National Library of Sweden
 - National Library of Finland
 - OCLC
 - PCC
 - Share-VDE
 - Linked Data for Production/Sinopia
 - Stanford University
 - University of Alberta
 - University of California, Davis
 - University of Pennsylvania



BIG Terms of Reference--Charge

Work collaboratively on the development and maintenance of interoperable BIBFRAME data guidelines to support production level implementation, to address issues restricting interoperability, and to inform development of associated toolings and infrastructure. It is not responsible for further development of the BIBFRAME ontology itself.

While members may use open and/or proprietary tools to support BIBFRAME data creation and exchange locally, this group is primarily focused on interoperability for unrestricted metadata reuse.





BIG Terms of Reference--Governance

- Chairs:
 - Past Chair: Melanie Wacker (Columbia University)
 - Current Chair: Ian Bigelow (University of Alberta)
 - Incoming Chair: Xiaoli Li (University of California, Davis)
- Terms of reference state:

"Initially the group reports to PCC PoCo on its activities until it is functioning independently. PoCo will provide a mechanism for wider oversight and transparency, to leverage the expertise of other PCC groups as needed, and to coordinate with LC until such time as its functioning as an independent body. After an initial period pending a decision by BIG and PoCo approval, the group will operate as a standalone international group, developing new governance guidelines as needed."

 - originally, BIG was given 2 years to operate as a standalone group; in April 2024, this deadline was extended by one year at the request of BIG





BIG Terms of Reference--Possible Work

- *Define a standard BIBFRAME “shape” to support data reuse including conversion to and from other formats*
- *Explore defining core BIBFRAME elements necessary for data exchange*
- *Surface issues regarding the use of the Official RDA with BIBFRAME and propose strategies for their resolution*
- *Collaborate and communicate with other groups working in the area of BF interoperability to ensure the ability to reuse BIBFRAME created in one community in other BIBFRAME stores.*
- *Examine the work accomplished by the Communication Working Group and apply to this charge where appropriate*
- *Gather use cases as necessary to inform decision making, expanding on the efforts of the Use Case Working Group and others*
- *Provide an avenue for other interested parties to contact the BIBFRAME Interoperability Group and/or reach out to other stakeholders*





Subgroups

- BIBFRAME Interlingua Group
 - developing shape guidelines for validation of a base BF descriptions (aka "BF Interlingua")
 - properties
 - shapes of subgraphs
 - violation levels
- DCTAP/SHACL Group
 - developing the validation structure
 - structure of spreadsheet for DCTap
 - DCTap to SHaCL conversion

DCTAP = Dublin Core Tabular Application Profiles
SHACL = Shapes Constraints Language





2024 Work Plan



2024 Work Plan (Partial)

- Iterative refinement of the DCTAP/SHACL validation for BIG based on testing with data from BIG members in collaboration with the BIBFRAME Interlingua Subgroup
 - *Develop standardized data flow/handoff between Interlingua Group and DCTap/SHACL Group*
 - *Continue to develop and standardize DCTap spreadsheet template*
 - *Identify documentation and training needs (readme documents and related) to support creation/maintenance of spreadsheet template(s).How?Where to keep documentation and other outputs*
- Tabular application profiles for BIBFRAME Interlingua (BI):
 - *Add subgraphs to textual monograph and serials profiles*
 - *Expand beyond bf:Work and bf:Instance to higher level aggregation/clustering approaches, e.g. bf:Hub and svde:Opus*



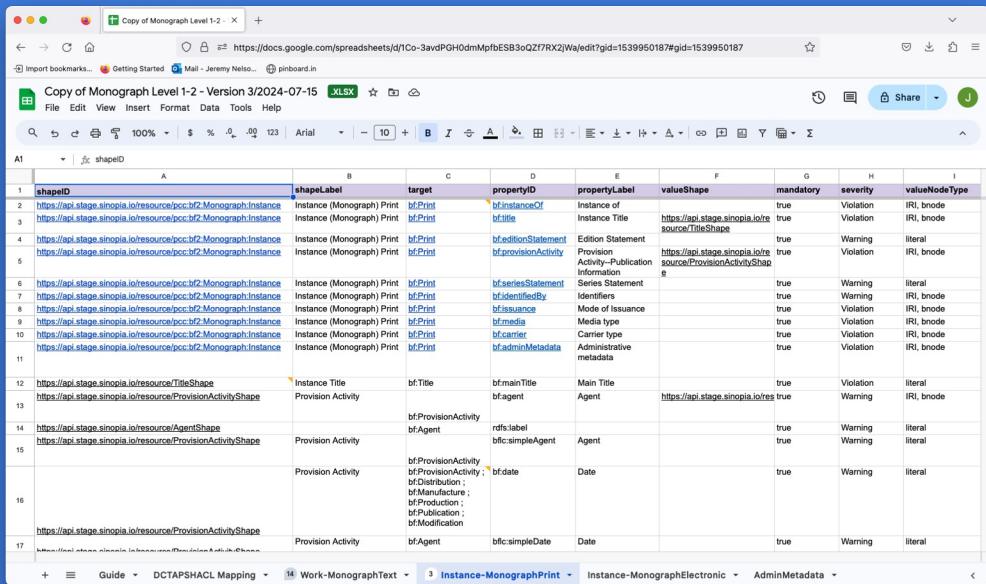
2024 Work Plan

- *Set up test project for data exchange for textual monographs based on Interlingua Group profile and using data validation tool developed by DCTap/SHACL Group*
 - Identify participants
 - Develop project plan
 - Draft testing protocol, including mechanisms/processes for feedback
 - Project documentation
 - Result analysis and iteration
- Documentation of best practices for technical aspects of BIBFRAME Interlingua as identified through the work of the group
 - Where to keep documentation



Validation Workflow

BIG DCTAP-to-SHACL Validation - Step 1



shapell	shapelLabel	target	propertyID	propertyLabel	valueShape	mandatory	severity	valueNodeType
https://api.stage.sinopia.io/resource/pcc:bf2:Monograph:Instance	Instance (Monograph) Print	bf:Print	bf:instanceOf	Instance of		true	Violation	IRI, bnode
https://api.stage.sinopia.io/resource/pcc:bf2:Monograph:Instance	Instance (Monograph) Print	bf:Print	bf:title	Instance Title	https://api.stage.sinopia.io/resource/TitleShape	true	Violation	IRI, bnode
https://api.stage.sinopia.io/resource/pcc:bf2:Monograph:Instance	Instance (Monograph) Print	bf:Print	bf:editionStatement	Edition Statement		true	Warning	literal
https://api.stage.sinopia.io/resource/pcc:bf2:Monograph:Instance	Instance (Monograph) Print	bf:Print	bf:provisionActivity	Provision Activity—Publication Information	https://api.stage.sinopia.io/resource/ProvisionActivityShape	true	Violation	IRI, bnode
https://api.stage.sinopia.io/resource/pcc:bf2:Monograph:Instance	Instance (Monograph) Print	bf:Print	bf:seriesStatement	Series Statement		true	Warning	literal
https://api.stage.sinopia.io/resource/pcc:bf2:Monograph:Instance	Instance (Monograph) Print	bf:Print	bf:identifiedBy	Identifiers		true	Warning	IRI, bnode
https://api.stage.sinopia.io/resource/pcc:bf2:Monograph:Instance	Instance (Monograph) Print	bf:Print	bf:issuance	Mode of Issuance		true	Violation	IRI, bnode
https://api.stage.sinopia.io/resource/pcc:bf2:Monograph:Instance	Instance (Monograph) Print	bf:Print	bf:media	Media type		true	Violation	IRI, bnode
https://api.stage.sinopia.io/resource/pcc:bf2:Monograph:Instance	Instance (Monograph) Print	bf:Print	bf:carrier	Carrier type		true	Violation	IRI, bnode
https://api.stage.sinopia.io/resource/pcc:bf2:Monograph:Instance	Instance (Monograph) Print	bf:Print	bf:adminMetadata	Administrative metadata		true	Violation	IRI, bnode
https://api.stage.sinopia.io/resource/TitleShape	Instance Title	bf:Title	bf:mainTitle	Main Title		true	Violation	literal
https://api.stage.sinopia.io/resource/ProvisionActivityShape	Provision Activity	bf:Agent	bf:agent	Agent	https://api.stage.sinopia.io/res	true	Warning	IRI, bnode
https://api.stage.sinopia.io/resource/AgentShape	Provision Activity	bf:ProvisionActivity	rdfs:label			true	Warning	literal
https://api.stage.sinopia.io/resource/ProvisionActivityShape	Provision Activity	bf:Agent	bf:simpleAgent	Agent		true	Warning	literal
	Provision Activity	bf:ProvisionActivity	bf:Distribution			true	Warning	literal
	Provision Activity	bf:ProvisionActivity ; bf:Production ; bf:Publication ; bf:Modification	bf:date	Date		true	Warning	literal
https://api.stage.sinopia.io/resource/ProvisionActivityShape	Provision Activity	bf:Agent	bf:simpleDate	Date		true	Warning	literal

Tab Separated Values File (TSV)

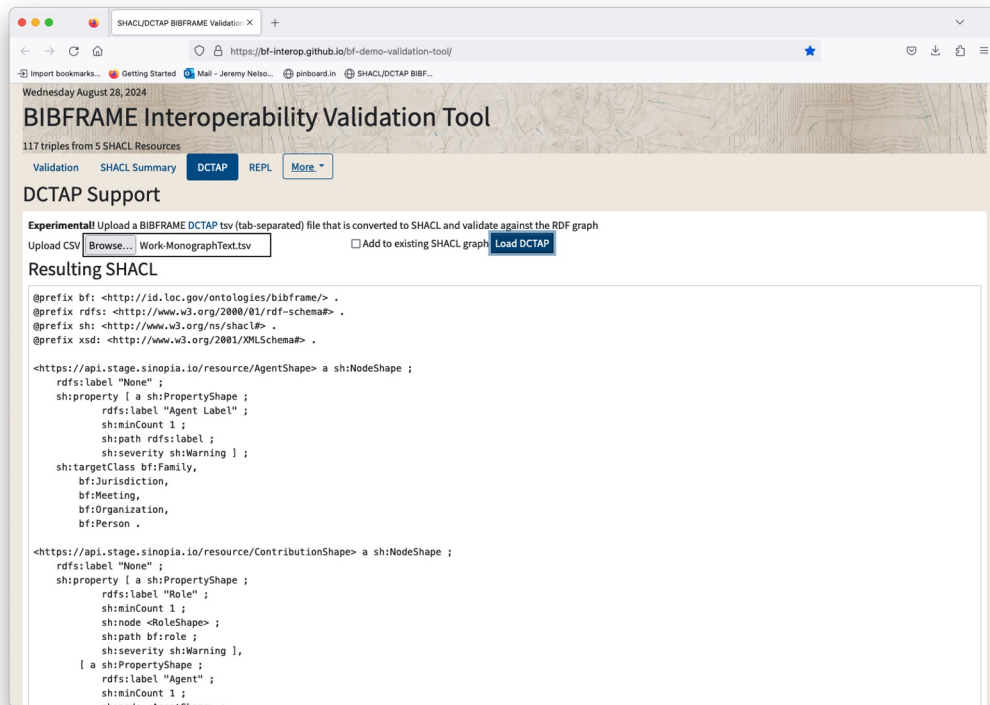
Each sheet is downloaded as TSV

Google Sheets used to construct DCTAP

BIG DCTAP-to-SHACL Validation - Step 2

Tab Separated
File (TSV)

TSV is uploaded to the
validation tool and then
converted to a SHACL
graph



The screenshot shows a web browser window with the URL <https://bf-interop.github.io/bf-demo-validation-tool/>. The page title is "BIBFRAME Interoperability Validation Tool". Below the title, it says "117 triples from 5 SHACL Resources". There are tabs for "Validation", "SHACL Summary", "DCTAP", "REPL", and "More". The "DCTAP Support" section contains an "Experimental" section with the text "Upload a BIBFRAME DCTAP tsv (tab-separated) file that is converted to SHACL and validate against the RDF graph". There are buttons for "Upload CSV", "Browse...", "Work-MonographText.tsv", "Add to existing SHACL graph", and "Load DCTAP". Below this is the "Resulting SHACL" section, which displays a SHACL graph in code format:

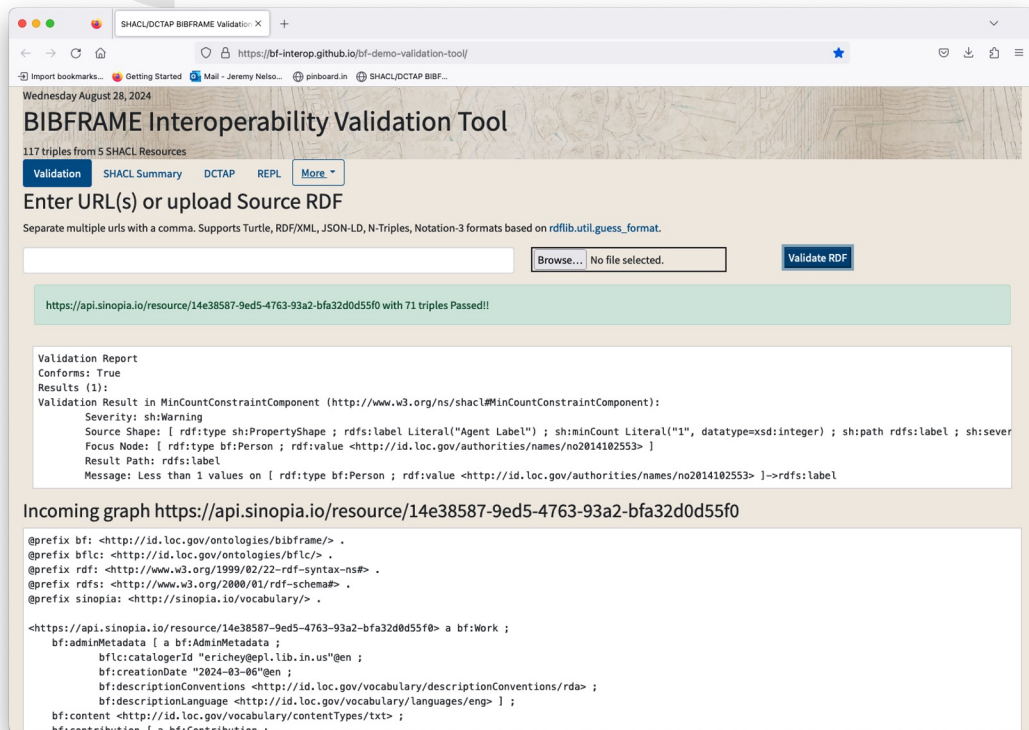
```
@prefix bf: <http://id.loc.gov/ontologies/bibframe/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix sh: <http://www.w3.org/ns/shacl#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<https://api.stage.sinopia.io/resource/AgentShape> a sh:NodeShape ;
  rdfs:label "None" ;
  sh:property [ a sh:PropertyShape ;
    rdfs:label "Agent Label" ;
    sh:minCount 1 ;
    sh:path rdfs:label ;
    sh:severity sh:Warning ] ;
  sh:targetClass bf:Family,
    bf:Jurisdiction,
    bf:Meeting,
    bf:Organization,
    bf:Person .

<https://api.stage.sinopia.io/resource/ContributionShape> a sh:NodeShape ;
  rdfs:label "None" ;
  sh:property [ a sh:PropertyShape ;
    rdfs:label "Role" ;
    sh:minCount 1 ;
    sh:node <RoleShape> ;
    sh:path bf:role ;
    sh:severity sh:Warning ],
  [ a sh:PropertyShape ;
    rdfs:label "Agent" ;
    sh:minCount 1 ;
    sh:node <AgentShape> .
```

Available at <https://bf-interop.github.io/bf-demo-validation-tool/>

BIG DCTAP-to-SHACL Validation - Step 3



Wednesday August 28, 2024

BIBFRAME Interoperability Validation Tool

117 triples from 5 SHACL Resources

Validation SHACL Summary DCTAP REPL More ▾

Enter URL(s) or upload Source RDF

Separate multiple urls with a comma. Supports Turtle, RDF/XML, JSON-LD, N-Triples, Notation-3 formats based on [rdflib.util.guess_format](#).

No file selected.

<https://api.sinopia.io/resource/14e38587-9ed5-4763-93a2-bfa32d0d55f0> with 71 triples Passed!!

Validation Report

Conforms: True

Results (1):

Validation Result in MinCountConstraintComponent (<http://www.w3.org/ns/shacl#MinCountConstraintComponent>):

Severity: sh:Warning

Source Shape: [rdf:type sh:PropertyShape ; rdfs:label Literal("Agent Label") ; sh:minCount Literal("1", datatype=xs:integer) ; sh:path rdfs:label ; sh:severity sh:Warning]

Focus Node: [rdf:type bf:Person ; rdf:value <http://id.loc.gov/authorities/names/n02014102553>]

Result Path: rdfs:label

Message: Less than 1 values on [rdf:type bf:Person ; rdf:value <http://id.loc.gov/authorities/names/n02014102553>]->rdfs:label

Incoming graph <https://api.sinopia.io/resource/14e38587-9ed5-4763-93a2-bfa32d0d55f0>

```
@prefix bf: <http://id.loc.gov/ontologies/bibframe/> .
@prefix bfcl: <http://id.loc.gov/ontologies/bfcl/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix sinopia: <http://sinopia.io/vocabulary/> .

<https://api.sinopia.io/resource/14e38587-9ed5-4763-93a2-bfa32d0d55f0> a bf:Work ;
  bf:adminMetadata [ a bf:AdminMetadata ;
    bf:catalogerId "erichey@pl.lib.in.us"@en ;
    bf:creationDate "2024-03-06"@en ;
    bf:descriptionConventions <http://id.loc.gov/vocabulary/descriptionConventions/rda> ;
    bf:descriptionLanguage <http://id.loc.gov/vocabulary/languages/eng> ] ;
  bf:content <http://id.loc.gov/vocabulary/contentTypes/txt> ;
  bf:contribution [ a bf:Contribution ;
```

Clicking on the Validation tab, a user can validate BIBFRAME RDF based on the SHACL graph by:

- Entering a URL
- Uploading a RDF file

Violations and Warnings based on the DCTAP rules will be displayed to the user



Test Preparation

- reformat the spreadsheets to accommodate requirements for DCTAP
 - initial tests showed that original spreadsheets were intended for humans, not machines
- expand the spreadsheet to include subgraphs
 - initially only top-level work/instance properties
- edit the SHACL to accommodate changes & subgraph information
- collect sample data
- provision of "how to" instructions & other documentation



Documentation/Sharing

BIG Wiki on PCC Wiki Site

Confluence Spaces People Search ? Log in

Program for Cooperative Cataloging (PCC)

SPACE SHORTCUTS

- PCC ISNI Pilot
- PCC Identity Management

PAGE TREE

- Advisory Committee on Equity, Diversity, Inclusion, Belonging, and Accessibility
- BIBFRAME Interoperability Group (BIG)**
 - 2023 Work Plan
 - 2024 Work Plan
 - Agendas
 - Membership
 - Presentations
- EMCO
 - Identity Management Advisory Committee
 - Linked Data Advisory Committee (LDAC)
 - Meetings
 - PCC AI and Machine Learning

Pages / Program for Cooperative Cataloging (PCC) Home

BIBFRAME Interoperability Group (BIG)

Created by Melanie Wacker, last modified on May 10, 2024

"In the interest of working collaboratively on Linked Data, the Program for Cooperative Cataloging (PCC) is initiating formation of this group per the interests and concerns expressed at the BIBFRAME Data Exchange Meeting organized by the PCC in September 2021. During this meeting different implementation decisions of the BIBFRAME ontology were identified as major obstacles to successful BIBFRAME data exchange. This group is the outgrowth of that meeting and is not intended to supersede any existing groups working on BIBFRAME or linked data in general. It will focus on supporting efficient and interoperable use of the BIBFRAME standard through establishing and sharing best practices between participants. While this group will be formed under the auspices of the PCC, it is intended that the leadership will rotate among the members of this group. " (Terms of Reference)

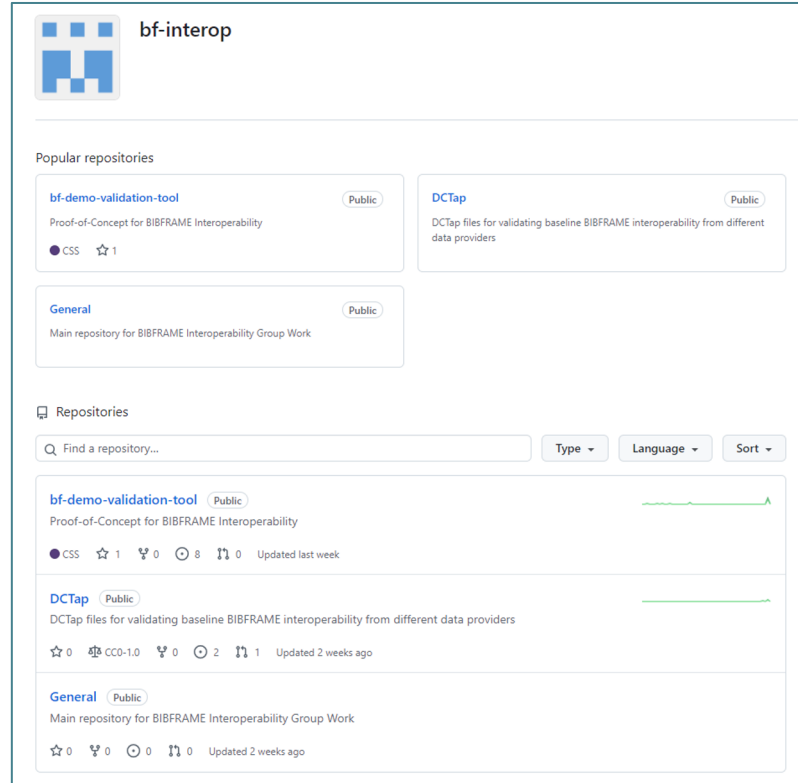
- Agendas
- 2023 Work Plan
- 2024 Work Plan
- Membership
- Presentations

No labels

All content on the LYRASIS Wiki is licensed under the CC BY (Attribution) license, unless otherwise noted.

Github Repository

github.com/bf-interop



The screenshot shows the GitHub repository page for 'bf-interop'. At the top left is the repository's profile picture, a blue square with a white grid pattern. To its right is the repository name 'bf-interop'. Below this is a section titled 'Popular repositories' containing three repository cards. The first card is for 'bf-demo-validation-tool', described as a 'Proof-of-Concept for BIBFRAME Interoperability', with a 'Public' badge, 'CSS' icon, and '1 star'. The second card is for 'DCTap', described as 'DCTap files for validating baseline BIBFRAME interoperability from different data providers', with a 'Public' badge. The third card is for 'General', described as the 'Main repository for BIBFRAME Interoperability Group Work', with a 'Public' badge. Below the popular repositories is a 'Repositories' section with a search bar and filters for 'Type', 'Language', and 'Sort'. The search results list the same three repositories with their respective icons (stars, forks, issues, pull requests) and update dates.

bf-interop

Popular repositories

- bf-demo-validation-tool** (Public)
Proof-of-Concept for BIBFRAME Interoperability
● CSS ☆ 1
- DCTap** (Public)
DCTap files for validating baseline BIBFRAME interoperability from different data providers
- General** (Public)
Main repository for BIBFRAME Interoperability Group Work

Repositories

Find a repository... Type Language Sort

- bf-demo-validation-tool** (Public)
Proof-of-Concept for BIBFRAME Interoperability
● CSS ☆ 1 🍴 0 🔄 8 🛠️ 0 Updated last week
- DCTap** (Public)
DCTap files for validating baseline BIBFRAME interoperability from different data providers
☆ 0 📄 CC0-1.0 🍴 0 🔄 2 🛠️ 1 Updated 2 weeks ago
- General** (Public)
Main repository for BIBFRAME Interoperability Group Work
☆ 0 🍴 0 🔄 0 🛠️ 0 Updated 2 weeks ago



URIs

- SHACL graph uses Sinopia URIs
 - result of its initial development in the tool

- Looking for a more general URI subdomain/subdirectory
 - would prefer something more neutral or more connected specifically with BF itself





Next Steps



Next Steps for Testing

- complete test preparation
 - complete and test edited SHACL
 - collect sample data
 - *Library of Congress, National Library of Finland, National Library of Sweden, OCLC, Share-VDE, Stanford, UCDavis*
 - Native BIBFRAME descriptions vs BIBFRAME converted from MARC
 - document metadata origins tested (e.g., BF created with an LC MARC2BF converter; BF from Sinopia; BF with SVDE extension; BF created from MARC by OCLC)
- complete initial documentation for tool use
- carry out testing & tweak as necessary
- share work with consultants for testing and validation of assumptions





Testing--The Big Questions

- Will the shapes and validation tools work across varied flavors of BIBFRAME we are testing with?
- Are the shapes specified by BF Interlingua sufficient for interchange? Or do they need to be enhanced further?
- How much do local extensions & flavors affect the efficacy of the validation tools?





Issues for the Future

- codify & expand interoperability scope
 - expand to other formats
 - provide general guidelines for use of extensions
 - document origin of metadata
- Set up consultant group and coordinate with them to receive feedback and input on testing
- Subdomains/Subdirectory for SHACL
- Deeper dive into BIBFRAME extensions
- Official RDA & BIBFRAME
- Coordinate with external groups (committees, working groups, etc.) working in related areas



Questions?

**Thanks to all BIG members who contributed to these slides
(especially Ian Bigelow, Jeremy Nelson & Xiaoli Li)**