MARC to **BIBFRAME**: Evaluating the extraction of bibliographic families

Michalis Sfakakis

Hellenic General Council of Libraries

&

Database and Information Systems group,
Department of Archives, Library Science and Museology, Ionian University,
Corfu, Greece

European BIBFRAME Workshop 2018 | 18-19 September 2018, Fiesole (Florence), Italy

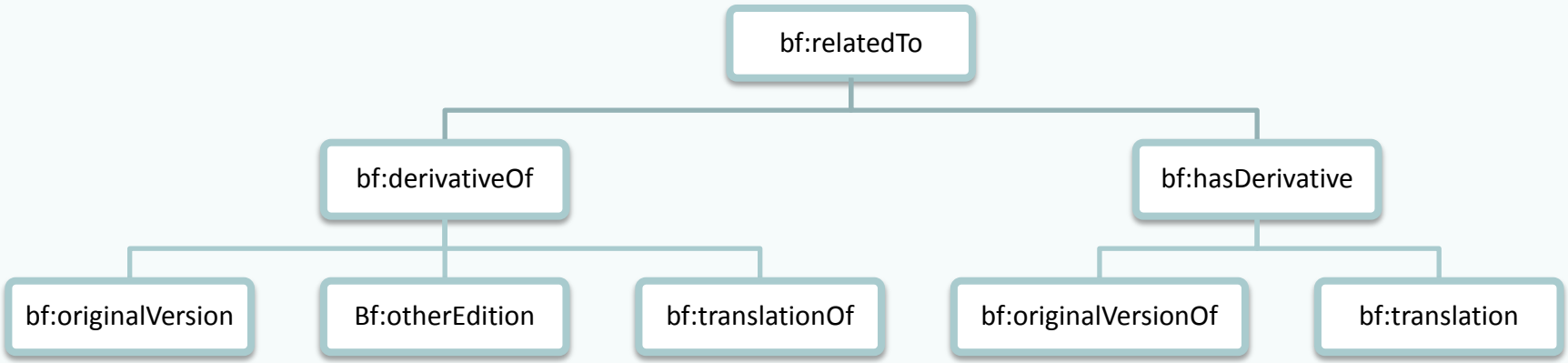
Bibliographic Family

- ‘***a set of related bibliographic works that are somehow derived from a common progenitor***’
  (coined by Prof. *Smiraglia)*
  - *Works or Expressions* within the same bibliographic family *may share* the *same intellectual content* and be related to the progenitor through different types of relationships

- Expresses how a Work (its ideas) is influenced by or influences other works in time

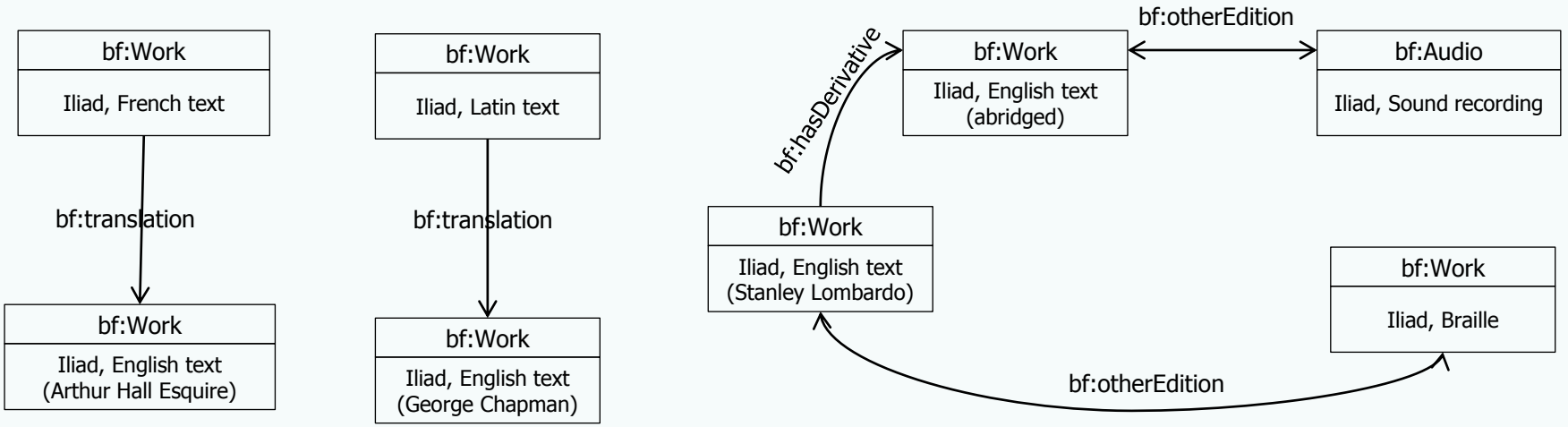
BIBFRAME: Derivations in Bibliographic Families

- Derivative relationships are mostly evolved in translations, adaptations, abridgements, dramatizations
- BF property hierarchy for derivations



Derivation Examples in BIBFRAME

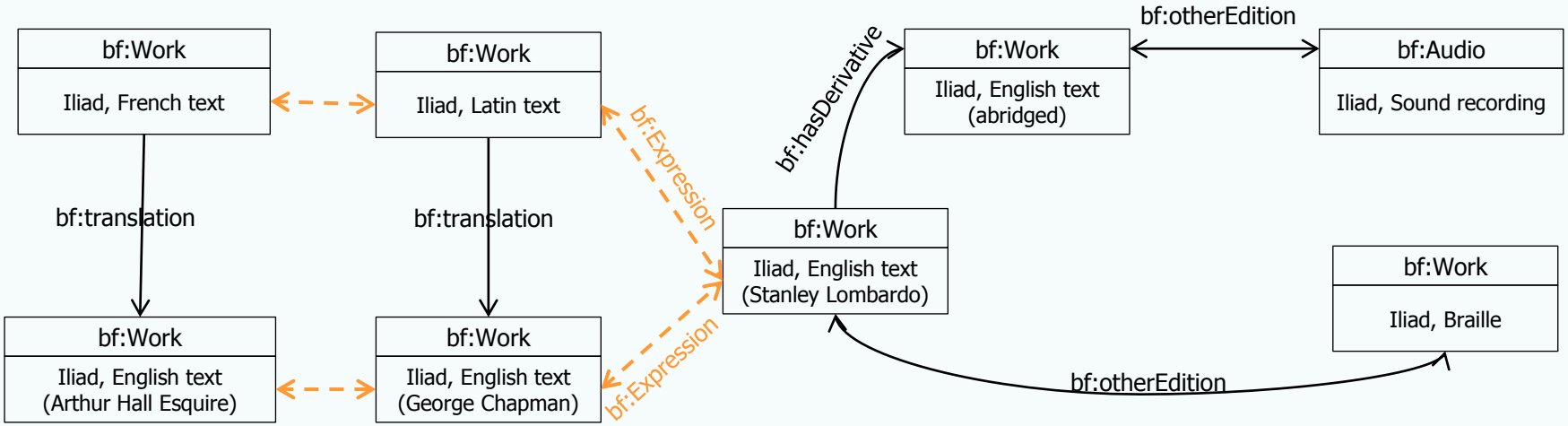
- Homer. Iliad



✘ … Some relationships are lost!

A “Connected” BF Representation

- Homer. Iliad family



Converting MARC records to BIBFRAME

- Moving from Records to Entities, Properties and Relationships
- Converting MARC records to BF
  - How effectively could MARC to BF conversion track down Content Relationships and Bibliographic Families?
- Assessment of conversion process

Assessing Conversion from MARC to BF Framework

- Create a **Gold** (or Reference) BF representation
- Create a **Test** representation from the 256 records using **marc2bibframe2** converter
  - Extracting all Entities and Relationships from MARC records
  - Matching same Entities and properties (effectiveness is not always easy! Converter does not consolidate same entities!)
- Compare Entities and Relationships between these representations

Gold BF Representation for Derivations

- A set of *popular works* (classical and well-known editions) were consider and selected by experts
  - Rich publication history, such as translations, adaptations, abridgements, dramatizations
  - Have more derivations forming large bibliographic families

- The publishing history of each family was studied to identify variant derivation types, such as:
  - Chapman's translation and adaptations in the *Odyssey* family
  - Lombardo's *Odyssey* translation in multiple content types (original text, sound recording, Braille)

Assessment Methodology - Datasets

Family	MARC records (publications)
Cien años de soledad	15
Crime and Punishment	29
Don Quijote	11
Faust	28
Iliad	25
Karamazov Brothers	21
Madame Bovary	32
Odyssey	20
The Scarlet letter	24
Tom Sawyer	31
Wuthering Heights	20
Total	**256**

- 11 popular works forming 11 families
- 256 related MARC records for respective publications

Gold BF Ontology for Derivations Analytics

Family	bf:translation	bf:hasDerivative	Total
Cien años de soledad	5	0	5
Crime and Punishment	15	12	27
Don Quijote	5	1	6
Faust	17	4	21
Iliad	20	14	34
Karamazov Brothers	15	9	24
Madame Bovary	13	4	17
Odyssey	12	9	21
The Scarlet letter	10	8	18
Tom Sawyer	8	8	16
Wuthering Heights	6	8	14
Total	**126**	**77**	**203**

Some Preliminary Indications

- Promising results for translations
- Could be more accurate when preprocess MARC records or post process convertor results

- Entity Matching is critical
- Matching for additional Works created from the Uniform Titles is challenging

- Other derivations and expressions are not covered

Hellenic General Council of Libraries

DBIS database & information systems group ionian university

Thank you!