ISTI Open Access - Feature #19931

Reinforce the information about the sources of items' manifestations

Oct 07, 2020 05:35 PM - Leonardo Candela

Status:	New	**Start date:**	Oct 07, 2020	
Priority:	Normal	**Due date:**		
Assignee:	Michele Artini	**% Done:**	0%	
Category:		**Estimated time:**	0.00 hour	
Target version:	ISTI Open Portal Tasks			

Description

The information we report for the manifestations associated to an item is heterogeneous and duplicated, e.g.

- the same url is annotated with diverse repositories / source names;
- sometimes it is reported unknown yet the name of the repository can be inferred from the url;

Some examples are below:

For https://openportal.isti.cnr.it/doc?id=people_____::b81bb986203da10873f301a5b2832b5b there is:

- arXiv.org e-Print Archive con URL http://arxiv.org/abs/1707.08411 -> OK

- an IEEE Transactions on Software Engineering is pointing to arxiv

- Online Research Database In Technology

  - https://backend.orbit.dtu.dk/ws/files/151468323/08405597.pdf
  - https://orbit.dtu.dk/en/publications/2d637490-4f09-4125-ba32-7d476215d1bc

- Dépôt Institutionel de l'Université catholique de Louvain et de l'Université Saint-Louis

  - http://hdl.handle.net/2078.1/218739
  - http://hdl.handle.net/2078.1/228850

- IEEE Transactions on Software Engineering (4 times)

  - https://academic.microsoft.com/#/detail/2962858291 … questo però è Microsoft Academy
  - http://dblp.uni-trier.de/db/journals/corr/corr1707.html#BeekLLV17 … questo però è DBLP
  - https://ieeexplore.ieee.org/document/8405597/
  - http://xplorestaging.ieee.org/ielx7/32/9036010/08405597.pdf?arnumber=8405597

- ieeexplore.ieee.org

  - https://ieeexplore.ieee.org/document/8405597

For https://openportal.isti.cnr.it/doc?id=people_____::f64822d915324f667ea4f02721d06398 there are 3 URLs annotated with "Unknown" ... one is Microsoft Academics while 2 came from Springer.

We should define some rules aiming at reducing all of this, e.g.

- avoid reporting the same URL many times ... it is useless;
- avoid repeating many times the name of the same source ... in case the name of the source is actually pointing to diverse URLs we might report the name of the source one time only and then add a placeholder (e.g. a number) for each URL;
- reduce the "Unknown" cases by inferring the name of the repo / source from the URL;

The solutions will never be perfect yet we will aim at reducing the current "mess".

Another approach to consider might be the following.

This information mainly come from OpenAIRE where giving visibility to the data sources is really important.

From the OpenPortal point of view we might also decouple

- the manifestations stored in People / IOP;
- the manifestations coming from OpenAIRE;

This solution is certainly simple to implement and give more visibility to OpenAIRE.