Preprints - Crossref’s view of the expanding territories
Crossref - scholarly infrastructure

- Founded to fight link-rot and ensure that the citation record is clear and up-to-date, functioning consistently across publishers
- The metadata is useful, freely available, human & machine accessible
- Works are connected to the full history of the published results
- Contributors are given credit for their work (ORCID)
- Everyone can identify the provenance and get context of a work
Metadata:

- Literature
- Associated research outputs
- Associated research entities
- Events surrounding it
Literature

Associated research outputs

Article

- datasets
- protocols
- preprints
- peer reviews
- software
- materials
- conf papers
- translations...
Preprints

42,541 preprints
Volume of registered preprints in Crossref

- Preprints registered:
  - Yr and a half ago: 13k
  - 1yr ago: 25k
  - Half a yr ago: 38k
  - This month: 50k

Date registered:
- Yr and a half ago
- 1yr ago
- Half a yr ago
- This month
Preprints by publisher (May 5, 2018)

- bioRxiv: 24,571
- PeerJ Preprints: 8,974
- Preprints.org: 4,211
- JMIR Preprints: 2,090
- ChemRxiv: 729
- Therapoid (Open Therapeutics): 2
Preprints metadata

• Repository name & hosting platform
• Contributors & ORCID
• Title
• Dates (posted, accepted)
• License
• Funding
• Abstract
• Relations
• References
Out of 42,541 records, the following metadata have been registered:

- License: 9710, 23% (PeerJ Preprints, ChemRxiv)
- Funder: 0, 0%
- ORCID: 18239, 43% (bioRxiv, PeerJ Preprints, Preprints.org, ChemRxiv)
- References: 1740, 4% (JMIR)
Metadata deposited (all Crossref records)
12,983 articles published from preprints
10.20844/preprints201608.0191.v1 is a preprint of 10.3390/data1030014
“Hey Crossref, which papers in my journals have preprints?”

Let me check the REST API...
“Hey Crossref, what are my most cited preprints?”

Let me check the Citedby count in the REST API...
It’s all about relations: relationship types connect the article with its resources
Research nexus: Clusterflock

Clusterflock: an algorithm optimizing distance-based clusters in orthologous gene families that share an evolutionary history

- Preprint: https://doi.org/10.1101/045773
- Supporting data: http://dx.doi.org/10.5524/100247
- Code: https://github.com/narechan/clusterflock
- Docker hub: https://hub.docker.com/r/narechan/clusterflock-0.1
- Video demo: https://youtu.be/ELZTVOiqKn8
- Peer reviews: https://doi.org/10.5524/review.100507 and https://doi.org/10.5524/review.100508
Literature → Activities surrounding it

Article

- shares
- discussions
- recommendations
- mentions
- citations
- reviews…
Most highly cited preprints

5. Cited by 28 - [https://doi.org/10.1101/021592](https://doi.org/10.1101/021592) Salmon provides accurate, fast, and bias-aware transcript expression estimates using dual-phase inference. August 30, 2016. (10.1038/nmeth.4197, 103 citations)
Crossref metadata reaches:

- Funders
- Institutions
- Archives & repositories
- Research councils
- Data centers
- Professional networks
- Patent offices
- Indexing services
- Publishing vendors
- Metrics providers
- Reference manager systems
- Lab & diagnostics suppliers
- Sharing platforms
- Data analytics systems
- Literature discovery services
- Educational tools
- PID providers, registration agencies
Thank you

Jennifer Lin, PhD
jlin@crossref.org
@jenniferlin15
orcid.org/0000-0002-9680-2328