



Project E: Citation

- Understanding the problem space
- Progress so far
- How you can contribute : afternoon session
- Lessons learned and challenges ahead

- Acknowledgements:
 - *Liz Lyon, Monica Duke (Bath)*
 - *Carole Goble (Manchester)*
 - *Jonathan Rees (Creative Commons)*
 - *Myles Axton, Timo Hannay (Nature PG)*
 - *Andrew Kasarskis (Sage)*
 - *Greg Hannum (UCSD)*
 - *Barend Mons (Netherlands)*

nature
cell biology

Editorial

Nature Cell Biology **11**, 1273 (2009)
doi:10.1038/ncb1109-1273a

Sharing data



Reference datasets should be accessible independently of scientific papers in a citable form, allowing attribution.



PLoS COMPUTATIONAL
BIOLOGY

An official journal of the International Society for Computational Biology

Scholar Factor (SF)

Philip E. Bourne^{*}, J. Lynn Fink

Correspondence

Nature Biotechnology **27**, 984 - 985 (2009)
doi:10.1038/nbt1109-984b

nature
biotechnology

Accreditation and attribution in data sharing

Gudmundur A Thorisson¹

1. Department of Genetics, University of Leicester, UK.

Credit where credit is overdue

EDITORIAL

A universal tagging system that links data sets with the author(s) that generated them is essential to promote data sharing within the proteomics and other research communities.

Why citation of network models?



...and data-driven models in other disciplines...

- **Career credit** : attribution, research assessment & reward
- **Research integrity** : academic rigour, provenance, audit tracking, data quality, validation and verification
- **Sustainability** : preservation of the scientific record over the long-term
- **Discovery and access** : Return-On-Investment from public funds, economic benefits
- **Re-use** : societal benefits, new knowledge, better medicine and healthcare

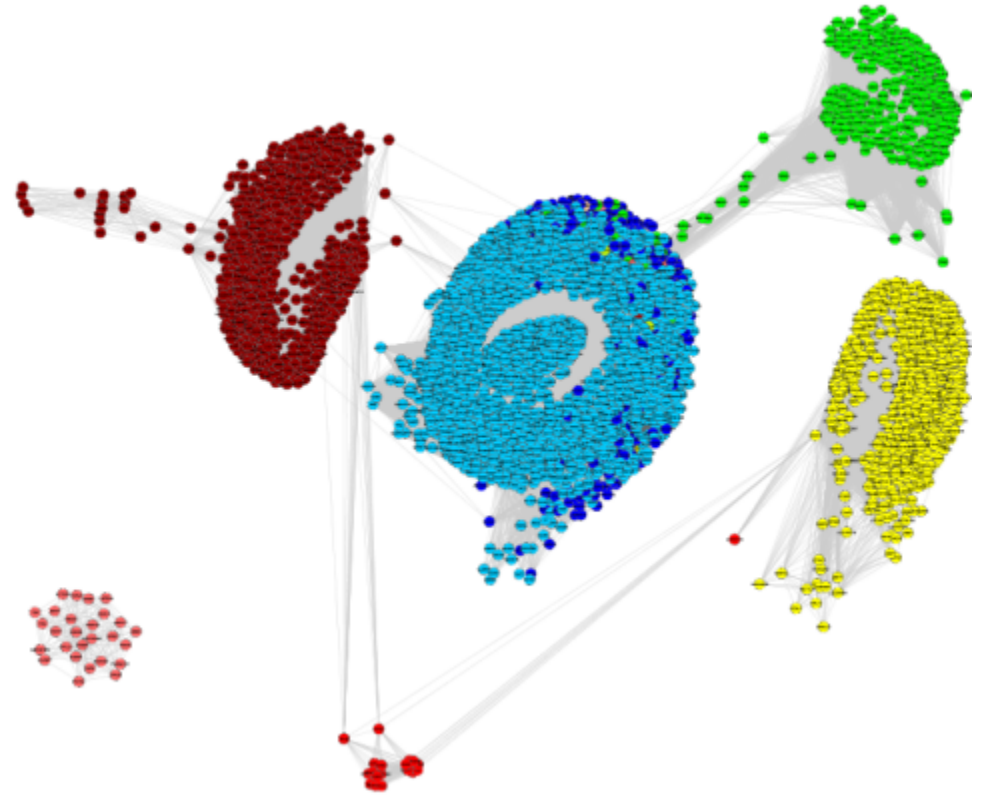
What are we citing?

- Journal
- Article
- Workflow
- Visualisation
- Model
- Data
- Annotation
- Concept

Macro



Micro / Nano



Attribution granularity

How? Functionality? Policy?

- Descriptive metadata – standards
- Persistent identification - URIs
- Identifier-agnostic framework
- Resilient resolution service
- Multi-directional linking e.g. to peer-reviewed paper, to datasets
- Version control
- Integrated in workflow
- ??



crossref.org

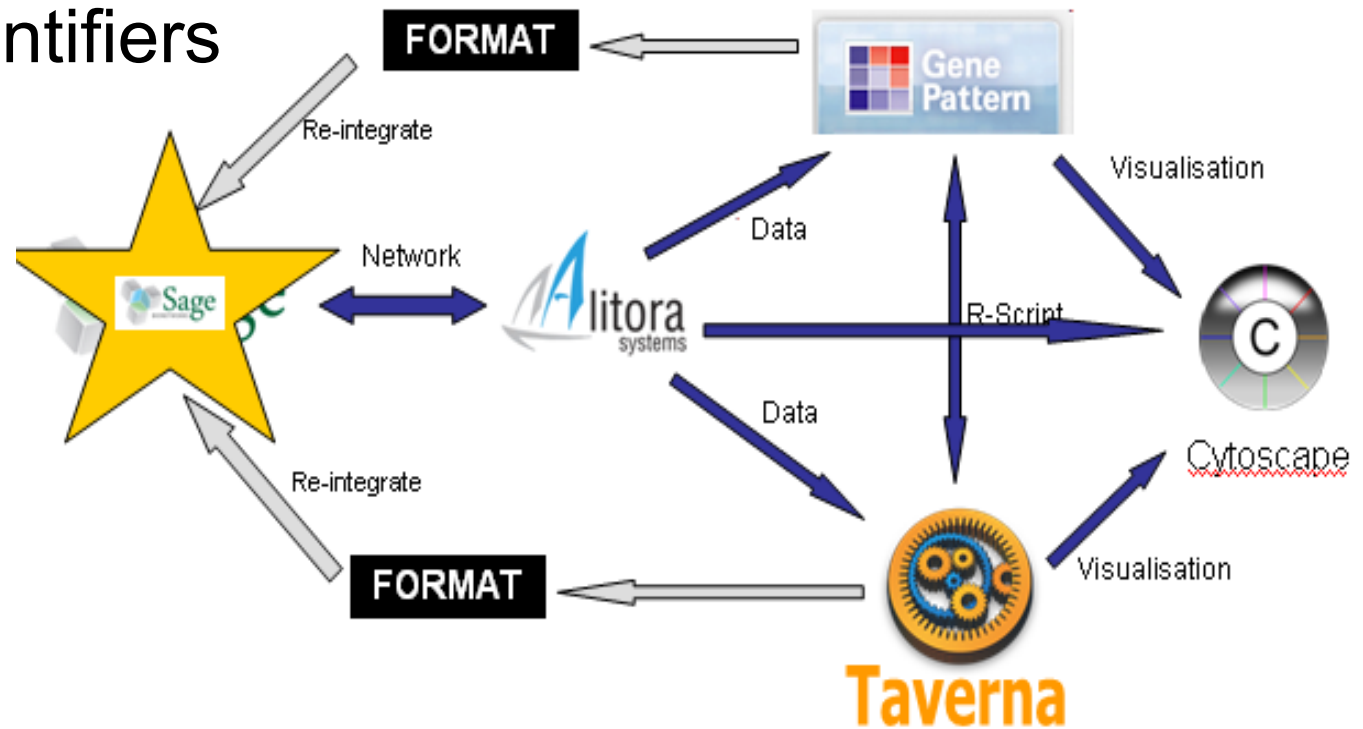
Examples: [10.1594/PANGAEA.119754](https://doi.org/10.1594/PANGAEA.119754)

http://dx.doi.org/10.1594/WDC/CCSRNIES_SRES_B2

The Sage Pipeline



- Local lab workflows
- Distributed data repositories
- Complex visualisations
- Emerging standards
- Multiple identifiers
- ???



Citation Overview Paper



- Draft on wiki for comment
- http://sage.fhcrc.org/sagewiki/index.php/Project_1/Draft_Overview
 - *Drivers*
 - *Nature of the data*
 - *Identifiers and Resolution*
 - *Content of Citation*
 - *Following Citations*
 - *Bi-directional linking*
 - *Versioning*
 - *Approaches: Citation vectors, TrackBack, DataCite, NanoPublication, ?*
 - *DataStore Initiatives: PANGAEA, Dryad, Dataverse, ?*



Afternoon session



- Requirements gathering exercise
- 10 Questions covering:
 - Workflows and process?
 - Data formats / data types?
 - Citation data and metadata?
 - Linking?
 - Versioning?
 - Metrics?
 - Barriers to implementation?
- We need your contribution.....



Project E: Summary

- Lessons learned
 - We need a (much) better understanding of Sage workflows and practitioner requirements
 - Citation service functionality will be dependent on community adoption of common standards
- Challenges ahead
 - Service development in partnership with relevant agencies and organisations
 - Influence funders and policy makers to ensure citation of models and associated data is rewarded
 - Development of new metrics for research assessment

