

SageCite

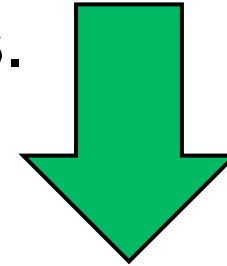
Building a data citation demonstrator for bionetwork models



Problem ?

Global coherent science requires multiple data contributors in a Federation / commons.

How to incentivise data sharing?



Opportunity....

Data citation provides attribution and credit associated with publication.....

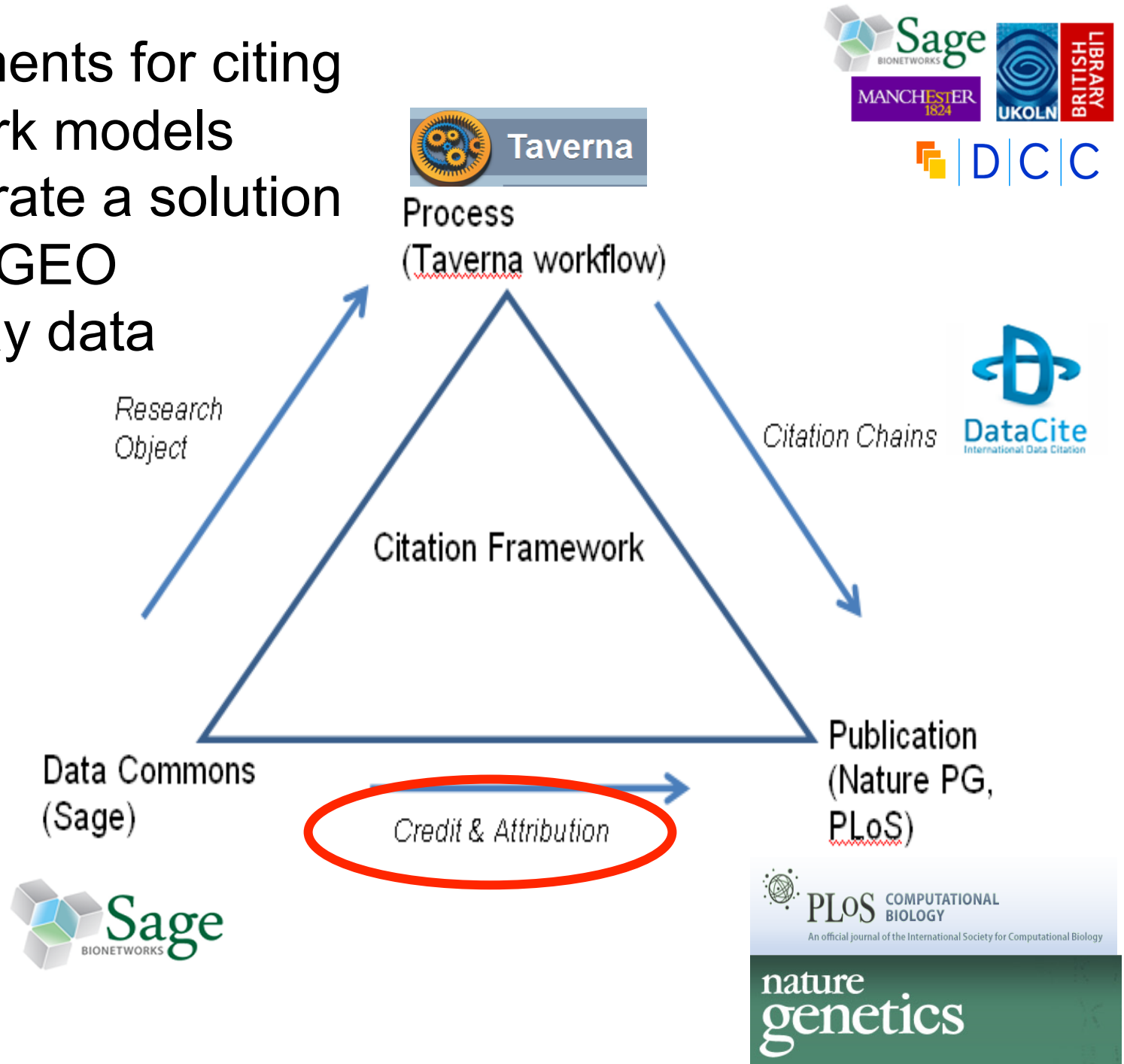


Credit where credit is overdue

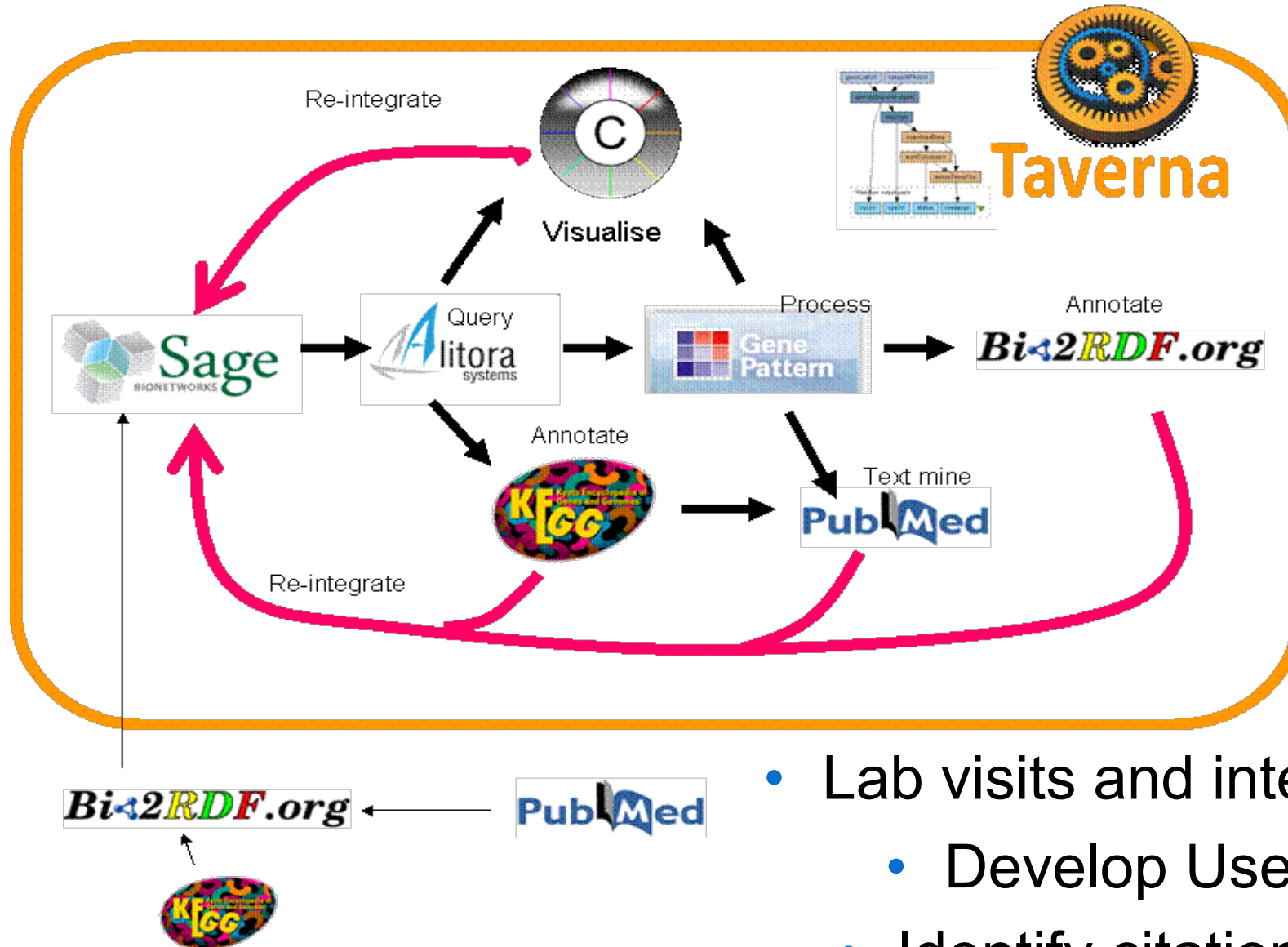
Accreditation and attribution in data sharing

- Requirements for citing bionetwork models
- Demonstrate a solution
- Testbed: GEO microarray data

SageCite

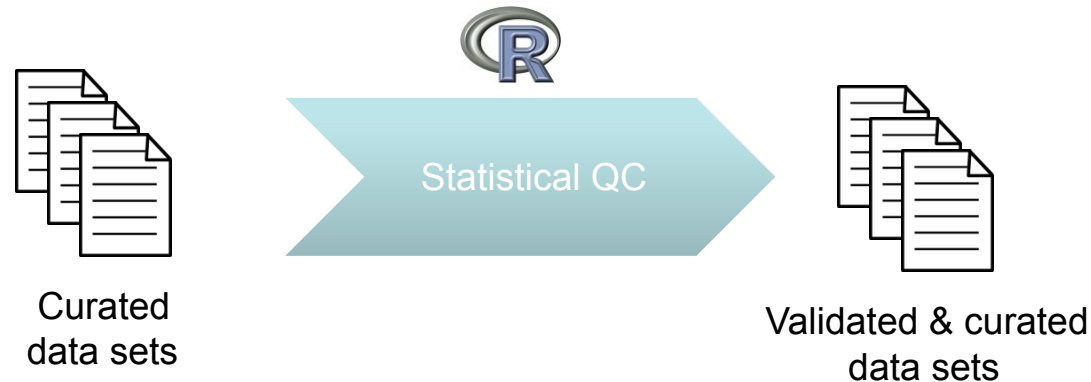


Mapping data pipelines



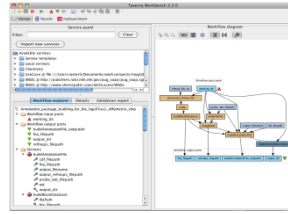
- Lab visits and interviews
 - Develop Use Cases
 - Identify citation points

Sage bionetwork modelling process



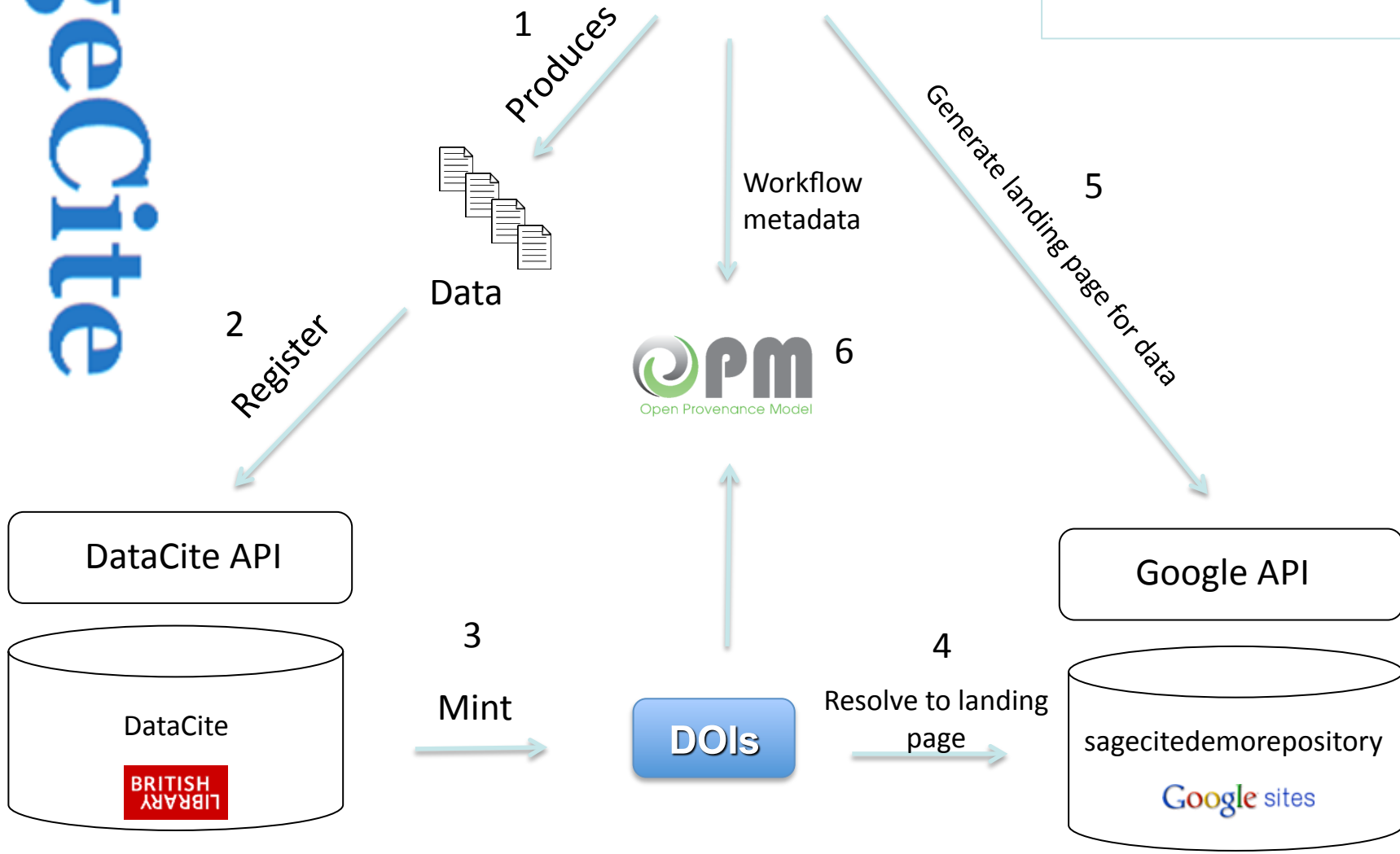
- Actual values in data sets are validated for quality to check for experimental artifacts

SageCite

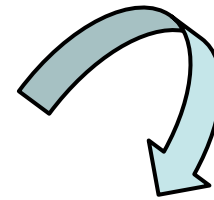
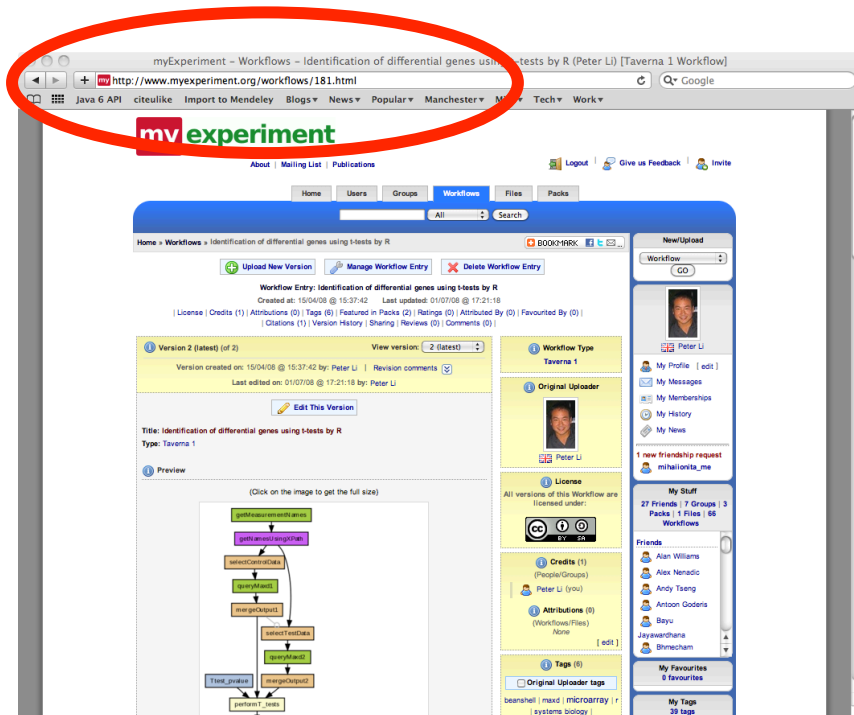


Taverna workflow

The relationships between data via DataCite DOIs with tools are captured by the provenance (OPM) produced by Taverna



Sharing and citation of workflow protocols



BMC Bioinformatics



Open Access

Software

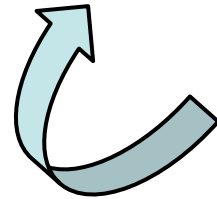
Performing statistical analyses on quantitative data in Taverna workflows: An example using R and maxdBrowse to identify differentially-expressed genes from microarray data

Peter Li^{*1}, Juan I Castrillo², Giles Velarde¹, Ingo Wassink³, Stian Soiland-Reyes⁴, Stuart Owen⁴, David Withers⁴, Tom Oinn⁵, Matthew R Pocock⁶, Carole A Goble⁴, Stephen G Oliver² and Douglas B Kell¹

Availability and requirements

Taverna (version 1.7.1) and its supporting documentation are available from <http://taverna.sourceforge.net>. Supplementary material including the example workflow, its accompanying documentation, and the results of the analyses described in this paper are available from <http://www.mcisb.org/software/taverna/microarray/index.html>. The example microarray data analysis workflow is also available from the myExperiment web site: <http://www.myexperiment.org/workflows/181>.

my experiment



<http://www.myexperiment.org>

Benefits of citing bionetwork models?

- Keeping Research Data Safe (KRDS)
Benefits Taxonomy

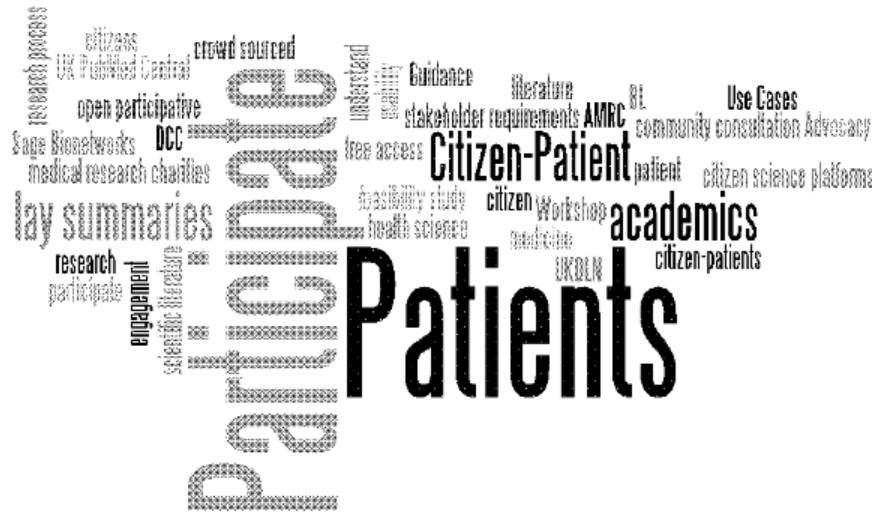
Dimension	Type of benefit	Type of benefit
What?	Direct	Indirect
When?	Near-term	Long-term
Who?	Private	Public

- Value Chain / Impact Analysis tool applied to SageCite research workflow

Activity	Change and Benefit	Impact Weighting (1 low- 5 high)	From When? (year1-5+ etc)	Who Benefits? (researcher, research group, institution, discipline, etc)
----------	--------------------	-------------------------------------	------------------------------	---



Patients Participate!



Citizen-patients producing crowd-sourced lay summaries of UK PubMed Central papers

Blog : <http://blogs.ukoln.ac.uk/patientsparticipate/>

- Feasibility pilot study
- Stem cell research areas
- Centre for Regenerative Medicine, Univ. of Bath UK
- Medical charities involved
- Workshop 16 May London
 - Develop Use Cases
- Deliver advocacy, guidance
 - Report & Recommendations



Questions?

- Monica Duke
- Liz Lyon
- Peter Li
- Carole Goble
- Max Wilkinson
- Tom Pollard
- Jon Izant
- Stephen Friend
- Phil Bourne
- Myles Axton

- Monica Duke
- Liz Lyon
- Lee-Ann Coleman
- Karen Walshe
- Allan Sudlow
- Simon Denegri
- Sophie Petit-Zeman
- Sara Ellis
- Stephen Friend
- Julian Chaudhuri