

BRIF and Variant Databases


Raymond Dalglish
University of Leicester



The three considerations

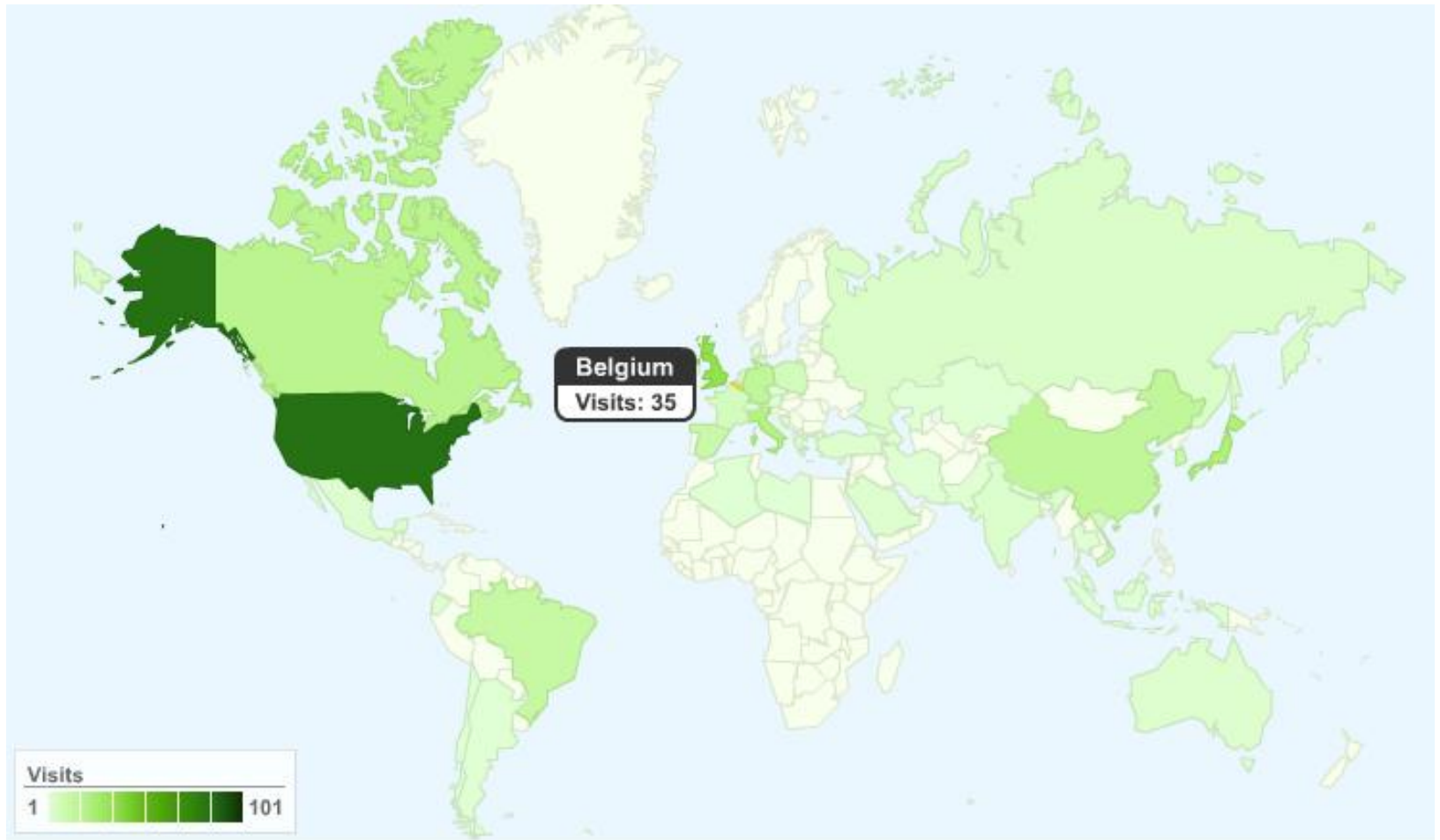
- 1. the quantitative use of a database
- 2. the quality and the importance of research results involving its use
- 3. the scientific and management efforts of those who set up, make available and maintain the database

1. Quantitative use

- What do you actually want to quantify?
- It's difficult to measure real use of a database in a quantitative fashion
- Page-hit counters only measure the number of visits to a page, but no more 
- Google Analytics provides more specific information categorised by dates and locations of the page requests

405 visits from 48 Countries Dec 2010

Nature Precedings : doi:10.1038/npre.2011.5586.1 : Posted 25 Jan 2011



The tracking process

- Tracking pages with Google Analytics requires code to be embedded into web pages
- Straightforward if you host the pages yourself and understand HTML
- More problematic if you have your database hosted for you, say using LOVD at LUMC
- Hit tracking could probably be added to LOVD as a feature

2. Quality and importance

- Is there an objective measure of quality?
- Who will judge the quality?
- Is publication of an account of the database a proxy for a measure of quality?
 - what is the citation impact of the journal?
 - is the number of times cited a measure of quality?
- Is citation of the database URL a valid proxy?
- What if there is no published account?

External quality measures

- ISI formerly featured web sites on the basis that *“The primary goal of this selection process is to ensure that the sites selected are highly authoritative, and that the information they cover is, in fact, scholarly and reliable.”*



Citation impact and citations

- Is publication of an account of the database a proxy for a measure of quality?
 - is the citation impact of the journal important?
 - is the number of times cited a measure of quality?
- Two accounts of the osteogenesis imperfecta database published: Human Mutation (6.887)
 - Dagleish 1997: 96 citations
 - Dagleish 1998: 57 citations

Citing the URL

- Journals sometimes restrict the number of references: URL cited instead of publication
- The cited URL will only be indexed in PubMed if it appears in the abstract
- Only 3 PubMed entries include the URL for the osteogenesis imperfecta database
- Many more accounts of the URL can be found if you mine entire publications: mechanism?

Building and maintaining

- Creating a sequence variant database is a major undertaking
 - Management: ongoing but limited effort
 - setting up the database
 - managing the hardware & software and data backups
 - Data entry: an ongoing major effort
 - identifying the data from journals and abstracts
 - checking and converting data
 - data entry and maintenance

Timescales

- Databases should be thought of as continuous publications: *c.f.* Online Mendelian Inheritance in Man
- There needs to be a measure of the effort taken over many years to build and maintain a database
- Eventually, databases will be handed on to others, or will decay through neglect

How long a contribution?

- Database managers will probably have a commitment of many years to a database
- Curators might be recruited for a one-off period of 3–4 months to enter data
- Submitters might occasionally submit data over several years: this task may be handed on to other persons from the same lab
- Is intense activity over a short period equivalent to lesser activity over a longer period?

Identifying the database

- Copyright protection applies to databases:
 - Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996
 - The Copyright and Rights in Databases Regulations 1997 (UK)
- A method for uniquely identifying the database does not appear to exist in either account of the legislation
- Databases need formal identifiers: something equivalent to ISSN (periodicals) or DOI (anything(?) digital): enquiries have been made

Identifying the contributors

- The identity of everybody who makes a contribution to the database must be unambiguous
 - which John Smith, Wei Li, Vimal Patel *etc.*?
- There is a need for a unique identifier ID such as proposed by ORCID: Open Researcher and Contributor ID (<http://www.orcid.org/>)

Acknowledgements

This work has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 200754 - the GEN2PHEN project

