

Identifying Users and Contributors on the Biomedical Internet

Gudmundur A Thorisson Paul Burton and Anthony J Brookes
University of Leicester, University Road, Leicester, LE1 7RH, UK

Key technologies & projects

Abstract

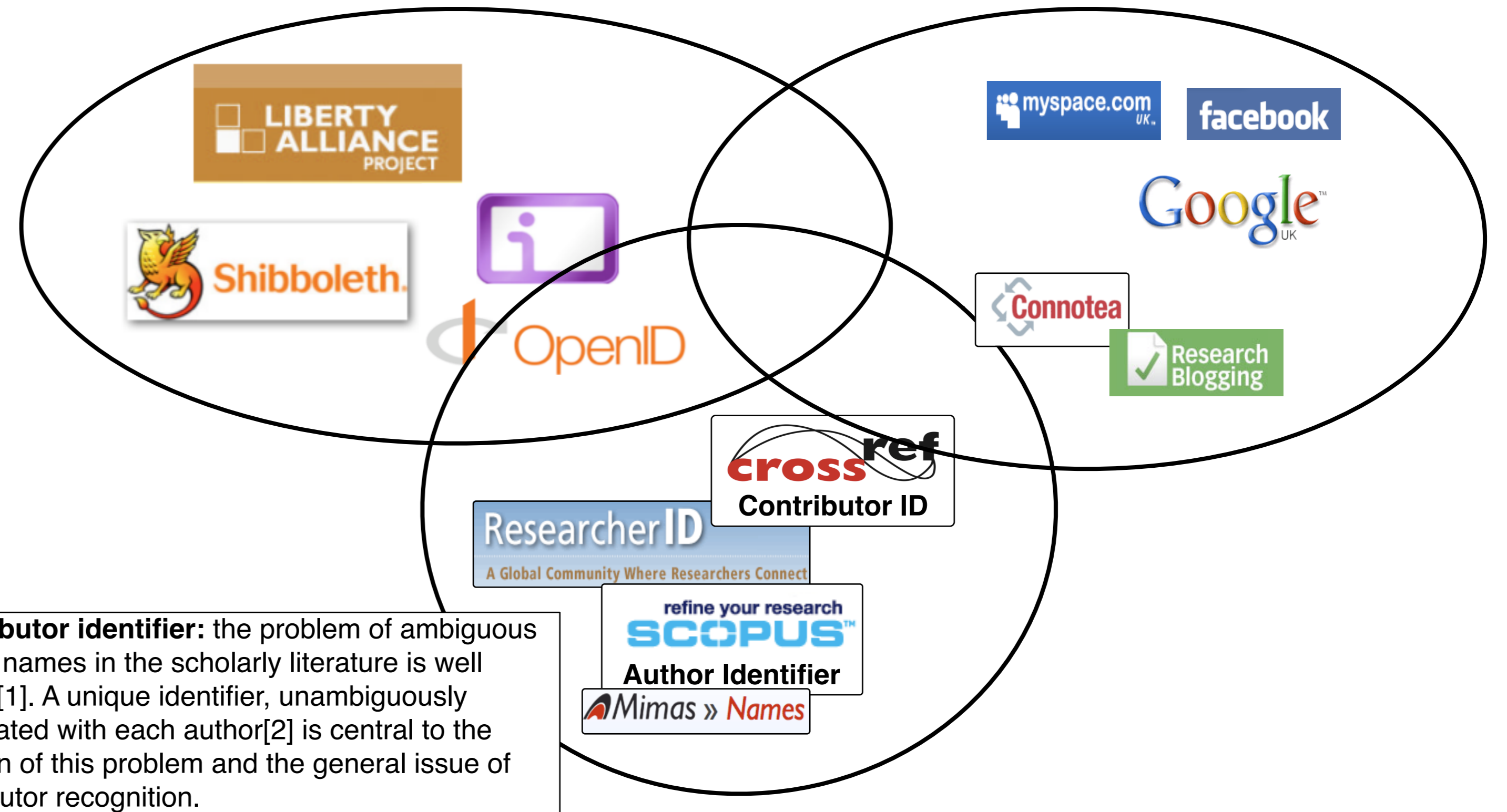
Given the pace and direction of evolution of the biomedical domain and the Internet, there is an urgent need for unambiguous and secure ways for researchers to be identified as they use and contribute to electronic publications and data resources. We consider researcher identification to be a central part of how biomedical databasing, and scientific reporting in general, needs to be developed. At the heart of this lies the concept of a user-centric system for researcher identification – i.e., one or more ‘ID systems’ by which individuals can be unambiguously identified along with various types of information associated with them, and where the individual controls his/her online identity and how/where it is used. At present, key Web 2.0 Internet technologies which can underpin such a system (e.g., OpenID - a decentralized, open authentication protocol), are being widely adopted.

To advance this field, a community of key stakeholders (e.g., GEN2PHEN, P3G, HUGO, HVP) has been assembled and is continually growing. This group is exploring innovative ways to exploit this new Internet ecosystem to support research-related activities and services. A dedicated website has been launched¹ and an international workshop on the subject was held on May 13-14 2009.

[1] <http://www.gen2phen.org/researcher-identification/> (<http://tinyurl.com/qo2lcj>)

Decentralized authentication: single sign-on (SSO) technologies provide users with portable identity profiles for logging into sites across the Web using a single account on an identity provider site.

Online social and collaborative tools: Web 2.0 tools to support online communities and collaborative work are increasingly used by researchers.



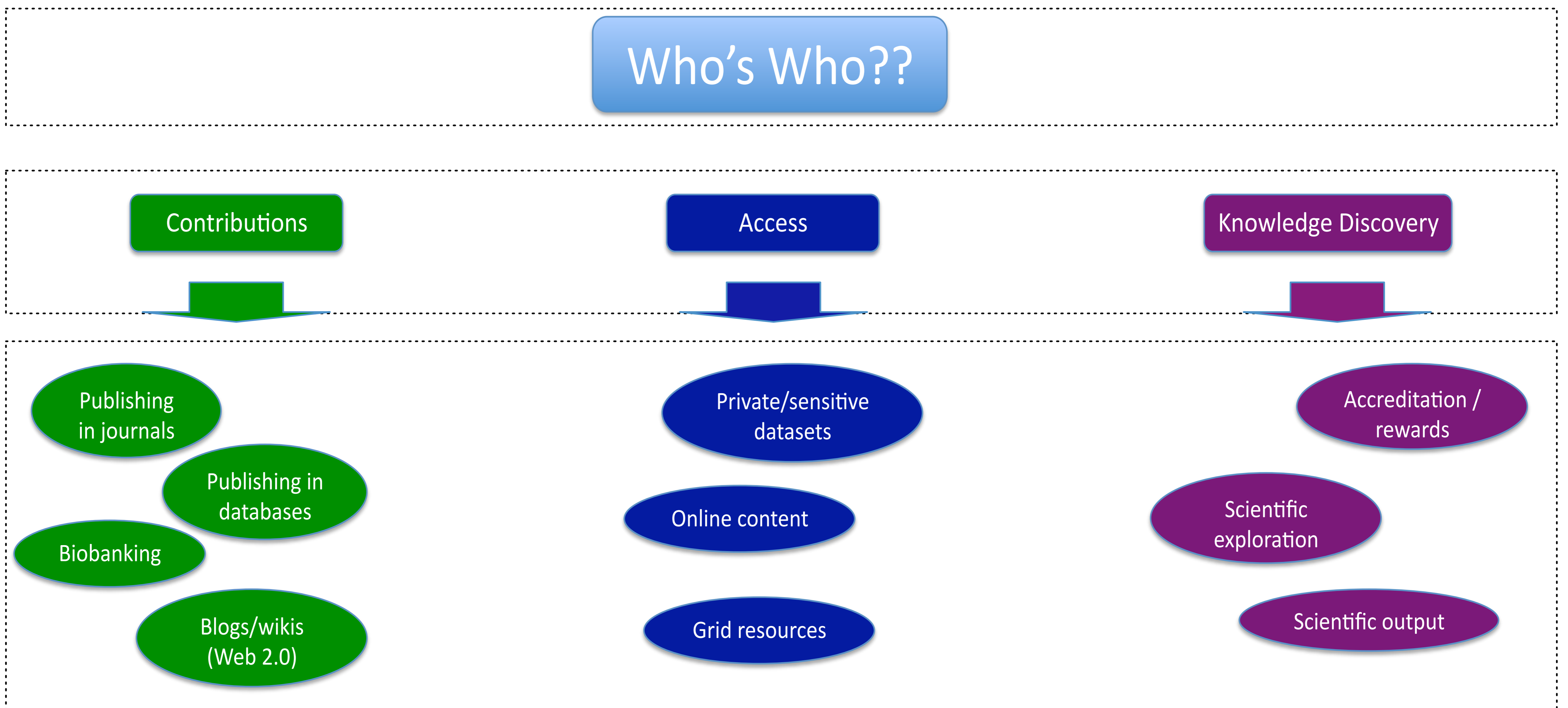
Contributor identifier: the problem of ambiguous author names in the scholarly literature is well known[1]. A unique identifier, unambiguously associated with each author[2] is central to the solution of this problem and the general issue of contributor recognition.

WHAT? The aim is to unambiguously and robustly identify researchers on the Internet.

Who's Who??

WHY? These three areas are the main domains where unambiguous researcher identification could prove immensely useful, and solve a host of existing and emerging problems.

WHEN/WHERE? Within each of the three broad domains identified above, several prominent application areas are highlighted.

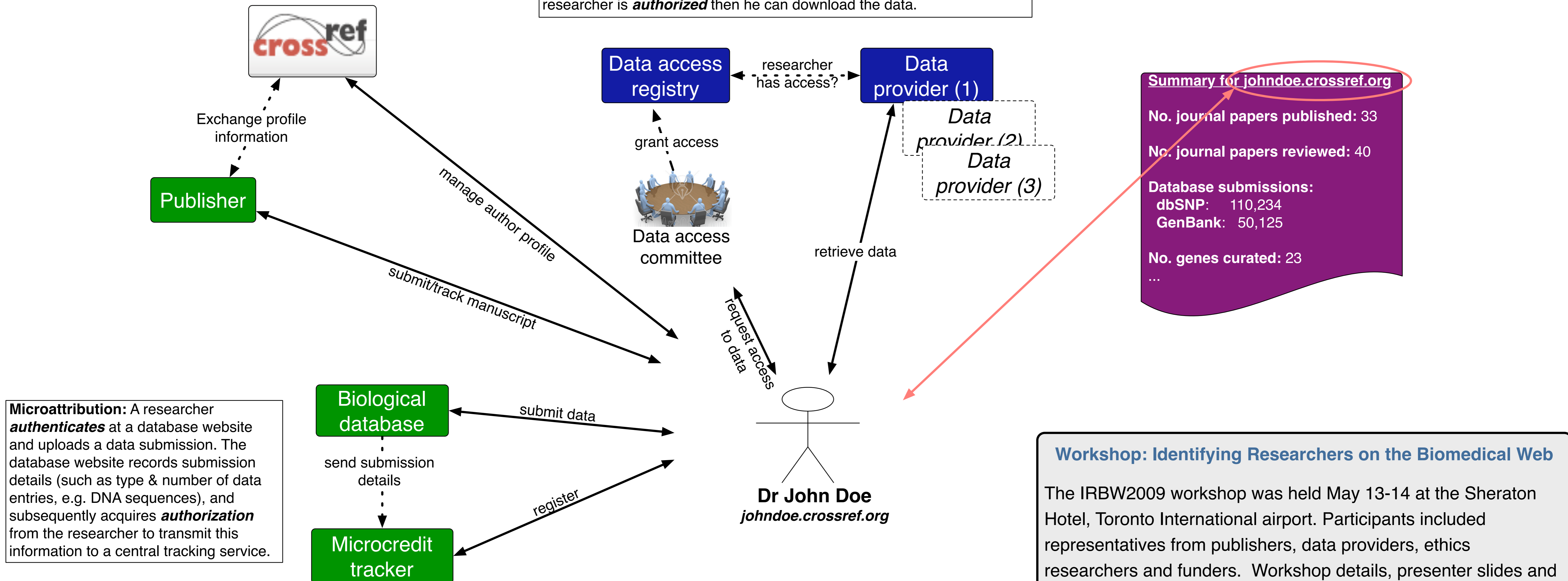


EXAMPLES

Single sign-on (SSO) for MS tracking systems: An author *authenticates* at a journal website in order to submit a manuscript. Instead of repeatedly submitting his address, affiliation and other information, he *authorizes* the journal site to retrieve this information from his central author profile housed at the (future) CrossRef Contributor ID service.

Controlling access to sensitive data: A researcher wishes to access a restricted-access genome-wide association study (GWAS) dataset. He first *authenticates* with the primary data producer and requests access. If the request is approved, access privileges for this dataset are linked to his identity in a global registry. The researcher then *authenticates* at the data provider website, the data provider looks up in the registry, and if the researcher is *authorized* then he can download the data.

Assessing scientific output: Given a researcher's unique identifier, summarize contributions of various sorts, not only traditional journal publications, but also database submissions, blogs etc. associated with that identifier. Optionally calculate metric(s), e.g. the Scholar Factor proposed in [3].



Microattribution: A researcher *authenticates* at a database website and uploads a data submission. The database website records submission details (such as type & number of data entries, e.g. DNA sequences), and subsequently acquires *authorization* from the researcher to transmit this information to a central tracking service.

Workshop: Identifying Researchers on the Biomedical Web
The IRBW2009 workshop was held May 13-14 at the Sheraton Hotel, Toronto International airport. Participants included representatives from publishers, data providers, ethics researchers and funders. Workshop details, presenter slides and other materials will be available at:
<http://www.gen2phen.org/researcher-identification/irbw2009-workshop-may-13-14-toronto> (<http://tinyurl.com/r5gv14>)