Data Sharing and Publication

Mary A. Majumder, JD, PhD
Center for Medical Ethics and Health Policy

Baylor
College of
Medicine

Two Questions

Do citizen scientists have an ethical obligation to share data*?

*protocols, tools…

Do they have an ethical obligation to publish findings in scientific journals?

Are journals receptive to citizen science?

Data Sharing

Pro

- Principles and codes

"Citizen science project data and meta-data are made publicly available and where possible, results are published in an open access format."

EUROPEAN
CITIZEN SCIENCE
ASSOCIATION

Ten principles of citizen science

Citizen science is a flexible concept which can be adapted and applied within diverse situations and disciplines. The statements below were developed by the *'Sharing best practice and building capacity'* working group of the **European Citizen Science Association**, led by the Natural History Museum London with input from many members of the Association, to set out some of the key principles which as a community we believe underlie good practice in citizen science.

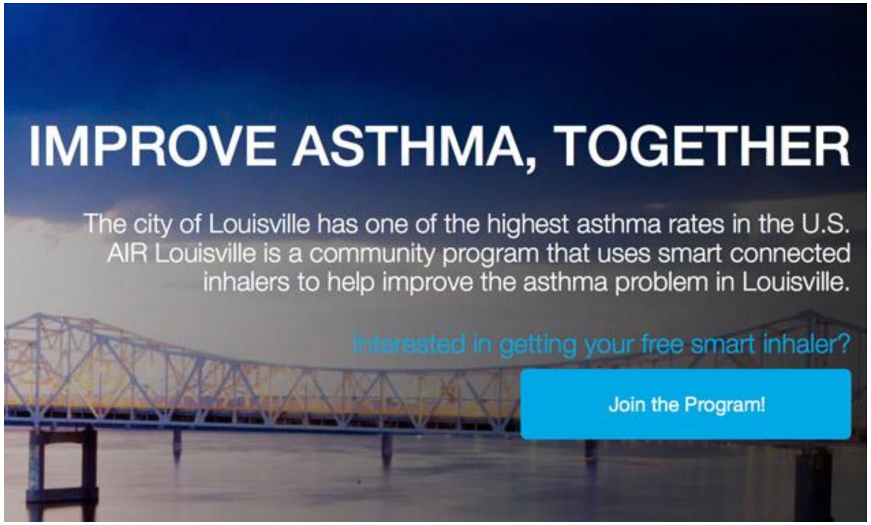
1. **Citizen science projects actively involve citizens in scientific endeavour that generates new knowledge or understanding.**
   Citizens may act as contributors, collaborators, or as project leader and have a meaningful role in the project.
2. **Citizen science projects have a genuine science outcome.**
   For example, answering a research question or informing conservation action, management decisions or environmental policy.
3. **Both the professional scientists and the citizen scientists benefit from taking part.**
   Benefits may include the publication of research outputs, learning opportunities, personal enjoyment, social benefits, satisfaction through contributing to scientific evidence e.g. to address local, national and international issues, and through that, the potential to influence policy.
4. **Citizen scientists may, if they wish, participate in multiple stages of the scientific process.**
   This may include developing the research question, designing the method, gathering and analysing data, and communicating the results.
5. **Citizen scientists receive feedback from the project.**
   For example, how their data are being used and what the research, policy or societal outcomes are.
6. **Citizen science is considered a research approach like any other, with limitations and biases that should be considered and controlled for.**
   However unlike traditional research approaches, citizen science provides opportunity for greater public engagement and democratisation of science.
7. **Citizen science project data and meta-data are made publicly available and where possible, results are published in an open access format.**
   Data sharing may occur during or after the project, unless there are security or privacy concerns that prevent this.
8. **Citizen scientists are acknowledged in project results and publications.**
9. **Citizen science programmes are evaluated for their scientific output, data quality, participant experience and wider societal or policy impact.**
10. **The leaders of citizen science projects take into consideration legal and ethical issues surrounding copyright, intellectual property, data sharing agreements, confidentiality, attribution, and the environmental impact of any activities.**

September 2015, London

Data Sharing

Pro

- Principles and codes
- “Citizen”
  - From the Latin *civitas* (city) – a part of a community, a contributor to a larger enterprise



IMPROVE ASTHMA, TOGETHER

The city of Louisville has one of the highest asthma rates in the U.S. AIR Louisville is a community program that uses smart connected inhalers to help improve the asthma problem in Louisville.

Interested in getting your free smart inhaler?

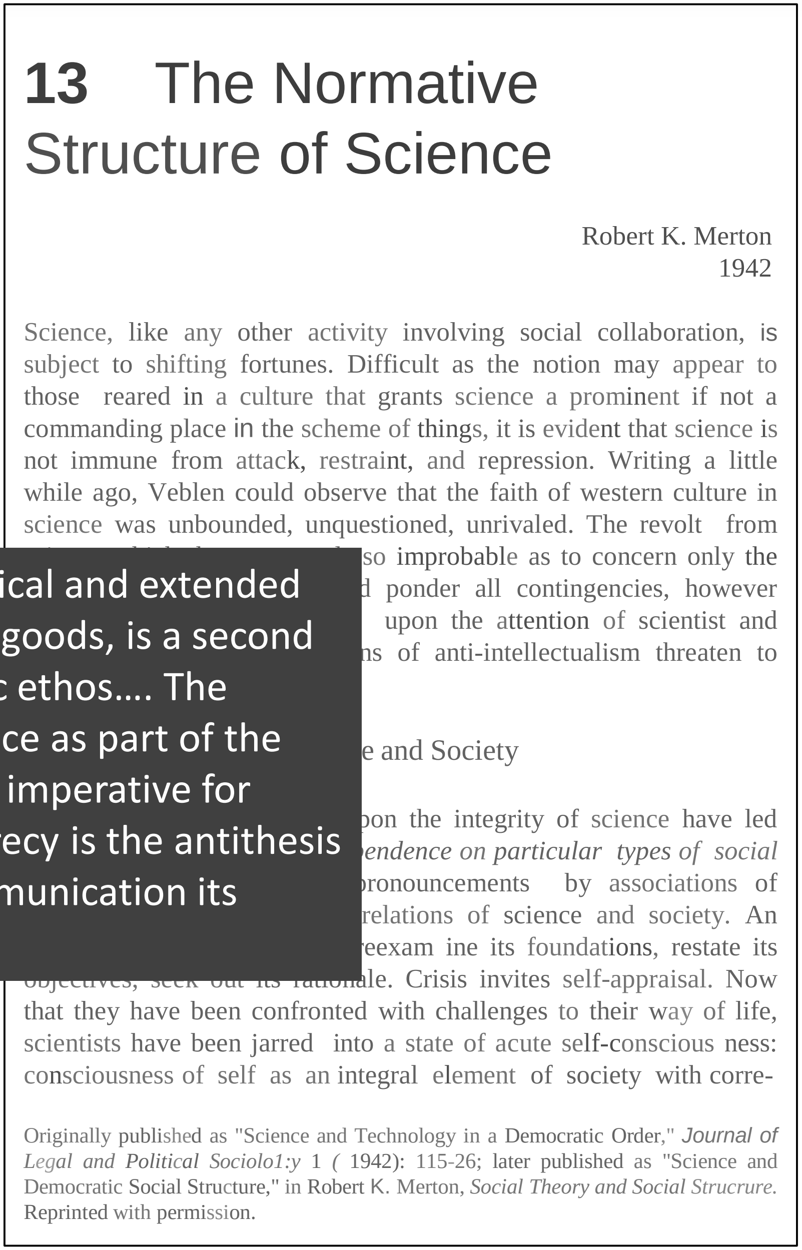
Baylor
College of
Medicine

Data Sharing

Pro

- Principles and codes
- “Citizen”
- “Science”

“‘Communism,’ in the nontechnical and extended sense of common ownership of goods, is a second integral element of the scientific ethos…. The institutional conception of science as part of the public domain is linked with the imperative for communication of findings. Secrecy is the antithesis of this norm; full and open communication its enactment.”



13 The Normative Structure of Science

Robert K. Merton
1942

Science, like any other activity involving social collaboration, is subject to shifting fortunes. Difficult as the notion may appear to those reared in a culture that grants science a prominent if not a commanding place in the scheme of things, it is evident that science is not immune from attack, restraint, and repression. Writing a little while ago, Veblen could observe that the faith of western culture in science was unbounded, unquestioned, unrivaled. The revolt from

so improbable as to concern only the
d ponder all contingencies, however
upon the attention of scientist and
ns of anti-intellectualism threaten to

e and Society

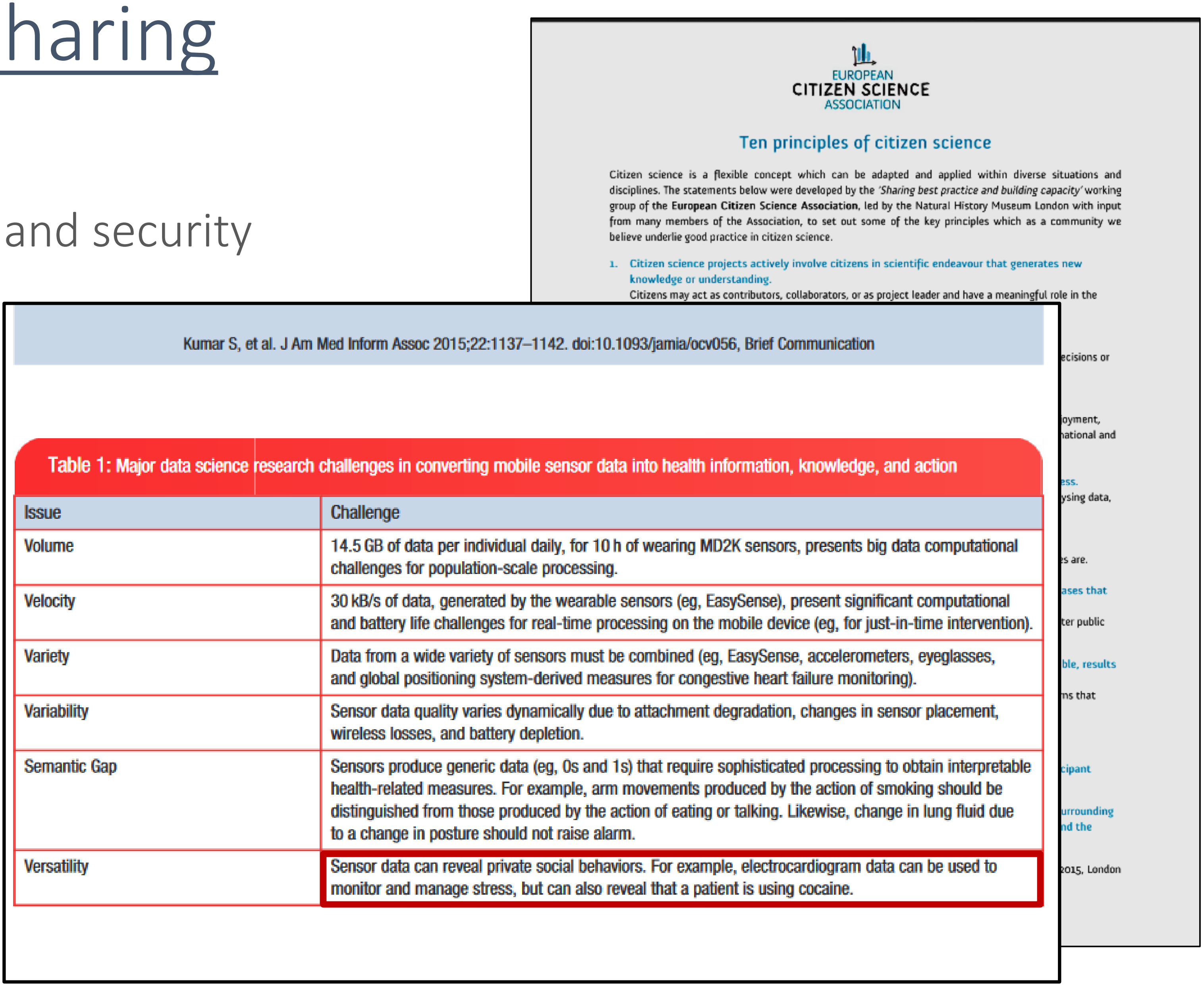
pon the integrity of science have led
endence on particular types of social
pronouncements by associations of
relations of science and society. An
reexam ine its foundations, restate its
ale. Crisis invites self-appraisal. Now that they have been confronted with challenges to their way of life, scientists have been jarred into a state of acute self-conscious ness: consciousness of self as an integral element of society with corre-

Originally published as "Science and Technology in a Democratic Order," *Journal of Legal and Political Sociolo1:y* 1 (1942): 115-26; later published as "Science and Democratic Social Structure," in Robert K. Merton, *Social Theory and Social Strucrure.* Reprinted with permission.

Data Sharing

Caveats

- Privacy and security



EUROPEAN
CITIZEN SCIENCE
ASSOCIATION

Ten principles of citizen science

Citizen science is a flexible concept which can be adapted and applied within diverse situations and disciplines. The statements below were developed by the *'Sharing best practice and building capacity'* working group of the **European Citizen Science Association**, led by the Natural History Museum London with input from many members of the Association, to set out some of the key principles which as a community we believe underlie good practice in citizen science.

1. **Citizen science projects actively involve citizens in scientific endeavour that generates new knowledge or understanding.**
   Citizens may act as contributors, collaborators, or as project leader and have a meaningful role in the

ecisions or

loyment,
ational and

ess.
ysing data,

s are.

ases that

ter public

ble, results

ns that

cipant

urrounding
nd the

2015, London

Kumar S, et al. J Am Med Inform Assoc 2015;22:1137–1142. doi:10.1093/jamia/ocv056, Brief Communication

Table 1: Major data science research challenges in converting mobile sensor data into health information, knowledge, and action

Issue	Challenge
Volume	14.5 GB of data per individual daily, for 10 h of wearing MD2K sensors, presents big data computational challenges for population-scale processing.
Velocity	30 kB/s of data, generated by the wearable sensors (eg, EasySense), present significant computational and battery life challenges for real-time processing on the mobile device (eg, for just-in-time intervention).
Variety	Data from a wide variety of sensors must be combined (eg, EasySense, accelerometers, eyeglasses, and global positioning system-derived measures for congestive heart failure monitoring).
Variability	Sensor data quality varies dynamically due to attachment degradation, changes in sensor placement, wireless losses, and battery depletion.
Semantic Gap	Sensors produce generic data (eg, 0s and 1s) that require sophisticated processing to obtain interpretable health-related measures. For example, arm movements produced by the action of smoking should be distinguished from those produced by the action of eating or talking. Likewise, change in lung fluid due to a change in posture should not raise alarm.
Versatility	Sensor data can reveal private social behaviors. For example, electrocardiogram data can be used to monitor and manage stress, but can also reveal that a patient is using cocaine.

Baylor
College of
Medicine

Data Sharing

Caveats

- Privacy and security
- Quality/infrastructure cost

Big Bad Data: Law, Public Health, and Biomedical Databases

Sharona Hoffman and Andy Podgurski

The accelerating adoption of electronic health record (EHR) systems will have profound impacts on clinical care. It will also have far-reaching implications for public health research and surveillance, which in turn could lead to changes in public policy, statutes, and regulations. The public health benefits of EHR use can be significant. However, researchers and analysts who rely on EHR data must proceed with caution and understand the potential limitations of EHRs.

Much has been written about the risk of EHR privacy breaches.[1] This paper focuses on a different set of concerns, those relating to data quality. Unlike clinical trial data, EHR data is not recorded primarily to meet the needs of researchers. Because of clinicians' workloads, poor user-interface design, and other factors, EHR data is surprisingly likely to be erroneous, miscoded, fragmented, and incomplete. Although EHRs eliminate the problem of cryptic handwriting, other kinds of errors are more common with EHRs than with paper records. Moreover, automated processing of EHR data can eliminate some opportunities for checks by humans. In addition, if causation is at issue, analysts must grapple with the complexities of making causal inferences from observational data. Public health findings can be tainted by the problems of selection bias, confounding bias, and measurement bias. These and other obstacles can easily lead to invalid conclusions and unsound public health policies.

The paper will highlight the public health uses of EHRs. It will also probe the shortcomings of EHR information and the challenges of collecting and analyzing it. Although some of the problems we discuss apply to all research, including that based on paper records, they will become all the more troubling and important in an era of electronic "big data," in which massive amounts of data are processed automatically, without human checks. Finally, we outline several regulatory and other interventions to address data analysis difficulties.

Public Health Benefits of EHRs

The advent of EHRs brings with it a wealth of opportunities for enhanced public health initiatives. EHR systems can report timely data that could facilitate surveillance of infectious diseases, disease outbreaks, and chronic illnesses. Software can extract data from records, analyze them, and electronically submit them to public health authorities, which will likely soon receive unprecedented amounts of information.[2] In fact, the "Meaningful Use" regulations with which providers must comply in order to be granted federal incentive payments for EHR adoption already require that providers be able to submit three types of data to public health authorities: lab results, syndromic surveillance, and immunizations.[3]

EHRs will also greatly facilitate public health research. Large EHR databases can enable researchers to conduct comprehensive observational studies that include millions of records from patients with diverse demographics who are treated in real clinical settings over many years. Researchers could use

Sharona Hoffman, J.D., LL.M., *is the Edgar A. Hahn Professor of Law and Professor of Bioethics and the Co-Director of the Law-Medicine Center at Case Western Reserve University School of Law. She received her B.A. from Wellesley College; J.D. from Harvard Law School; and LL.M. in Health Law from the University of Houston.* **Andy Podgurski, Ph.D.,** *is a Professor of Electrical Engineering and Computer Science at Case Western Reserve University. He received his B.S., M.S., and Ph.D. degrees from the University of Massachusetts.*

56 JOURNAL OF LAW, MEDICINE & ETHICS

Example 1

PLOS | BIOLOGY

COMMUNITY PAGE

CitSci.org: A New Model for Managing, Documenting, and Sharing Citizen Science Data

Yiwei Wang[1]*, Nicole Kaplan[2], Greg Newman[2], Russell Scarpino[2]

Helping you do great science

Create projects, build datasheets, collect data, and view results in real-time.

Data

Data can be downloaded on a project-by-project basis by visiting a project's profile page and clicking on the 'Download' button. Please contact webmaster@citsci.org for any additonal large dataset requests.

To browse the full CitSci.org dataset, click the 'Browse Data' button below. Note, this dataset is large and may load slowly.

Note:only public project data is accessible for downloading purposes.

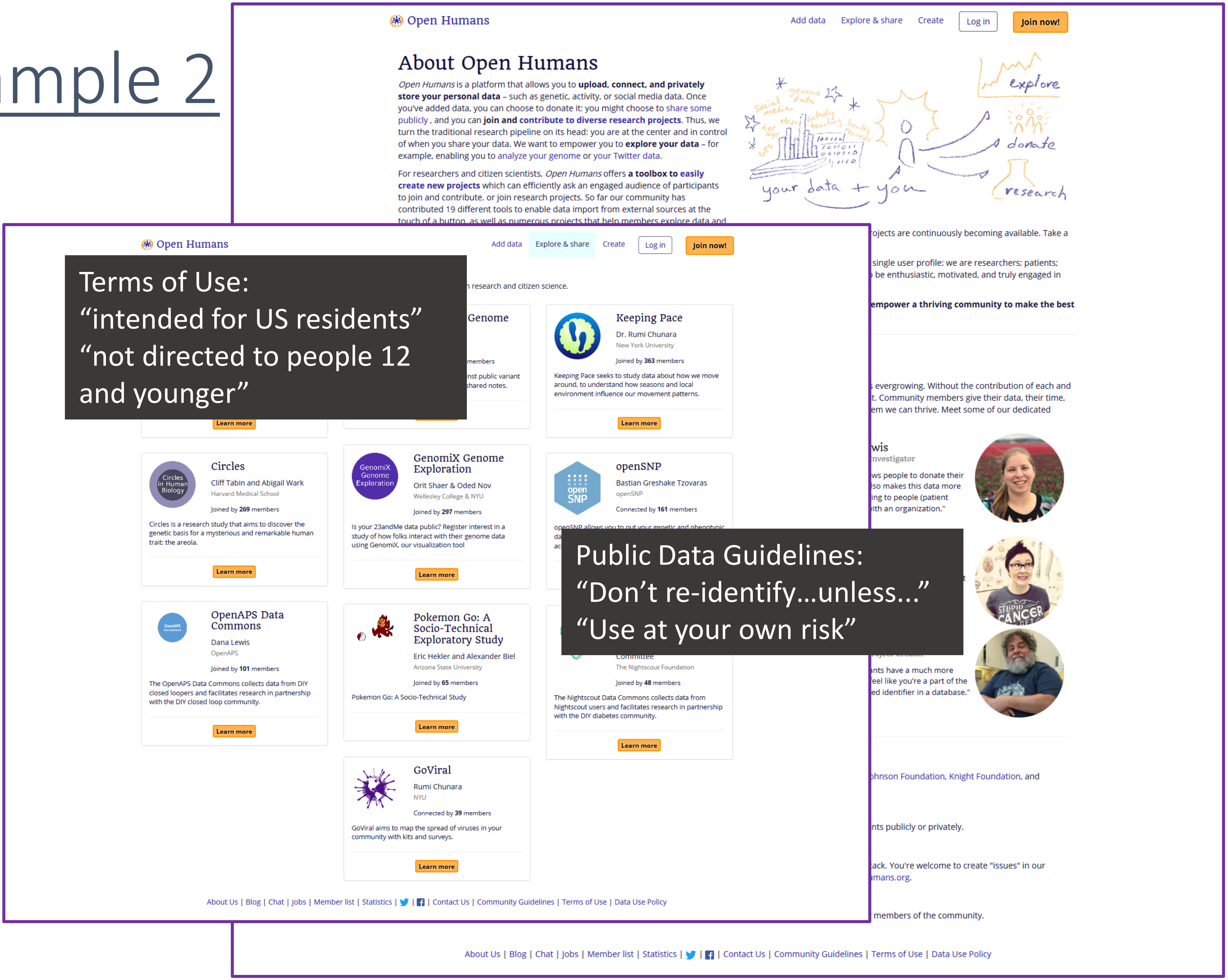
Search: homo sapiens

Show 50 entries

Date	Project	Organism	Scientific Name	Measurement	Value	Unit
No matching records found						

Showing 0 to 0 of 0 entries (filtered from 625,971 total entries)

Example 2



About Open Humans

Open Humans is a platform that allows you to **upload, connect, and privately store your personal data** – such as genetic, activity, or social media data. Once you've added data, you can choose to donate it: you might choose to share some publicly , and you can **join and contribute to diverse research projects**. Thus, we turn the traditional research pipeline on its head: you are at the center and in control of when you share your data. We want to empower you to **explore your data** – for example, enabling you to analyze your genome or your Twitter data.

For researchers and citizen scientists, *Open Humans* offers **a toolbox to easily create new projects** which can efficiently ask an engaged audience of participants to join and contribute. or join research projects. So far our community has contributed 19 different tools to enable data import from external sources at the touch of a button, as well as numerous projects that help members explore data and

ojects are continuously becoming available. Take a

single user profile: we are researchers; patients;
be enthusiastic, motivated, and truly engaged in

empower a thriving community to make the best

s evergrowing. Without the contribution of each and
t. Community members give their data, their time,
em we can thrive. Meet some of our dedicated

wis
nvestigator

ws people to donate their
so makes this data more
ing to people (patient
ith an organization."

nts have a much more
feel like you're a part of the
ed identifier in a database."

ohnson Foundation, Knight Foundation, and

nts publicly or privately.

ack. You're welcome to create "issues" in our
mans.org.

members of the community.

About Us | Blog | Chat | Jobs | Member list | Statistics | | | Contact Us | Community Guidelines | Terms of Use | Data Use Policy

n research and citizen science.

Genome

members

nst public variant
shared notes.

Keeping Pace
Dr. Rumi Chunara
New York University
Joined by **363** members
Keeping Pace seeks to study data about how we move around, to understand how seasons and local environment influence our movement patterns.
Learn more

Circles
Cliff Tabin and Abigail Wark
Harvard Medical School
Joined by **269** members
Circles is a research study that aims to discover the genetic basis for a mysterious and remarkable human trait: the areola.
Learn more

GenomiX Genome Exploration
Orit Shaer & Oded Nov
Wellesley College & NYU
Joined by **297** members
Is your 23andMe data public? Register interest in a study of how folks interact with their genome data using GenomiX, our visualization tool
Learn more

openSNP
Bastian Greshake Tzovaras
openSNP
Connected by **161** members
openSNP allows you to put your genetic and phenotypic

OpenAPS Data Commons
Dana Lewis
OpenAPS
Joined by **101** members
The OpenAPS Data Commons collects data from DIY closed loopers and facilitates research in partnership with the DIY closed loop community.
Learn more

Pokemon Go: A Socio-Technical Exploratory Study
Eric Hekler and Alexander Biel
Arizona State University
Joined by **65** members
Pokemon Go: A Socio-Technical Study
Learn more

Committee
The Nightscout Foundation
Joined by **48** members
The Nightscout Data Commons collects data from Nightscout users and facilitates research in partnership with the DIY diabetes community.
Learn more

GoViral
Rumi Chunara
NYU
Connected by **39** members
GoViral aims to map the spread of viruses in your community with kits and surveys.
Learn more

About Us | Blog | Chat | Jobs | Member list | Statistics | | | Contact Us | Community Guidelines | Terms of Use | Data Use Policy

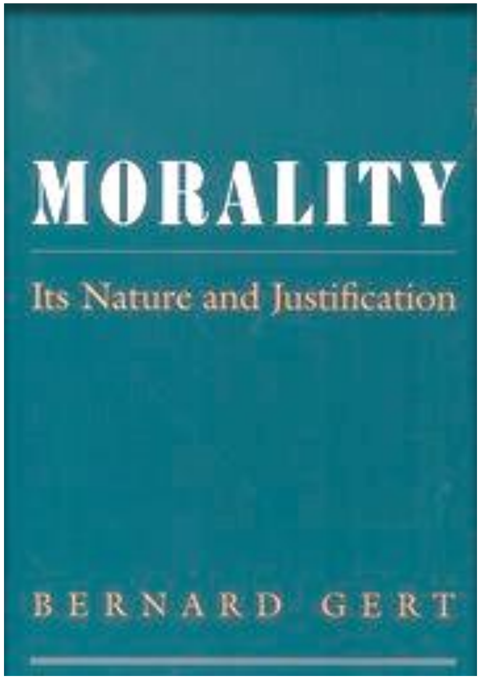
Terms of Use:
“intended for US residents”
“not directed to people 12 and younger”

Public Data Guidelines:
“Don’t re-identify…unless…”
“Use at your own risk”

Data Sharing

Con

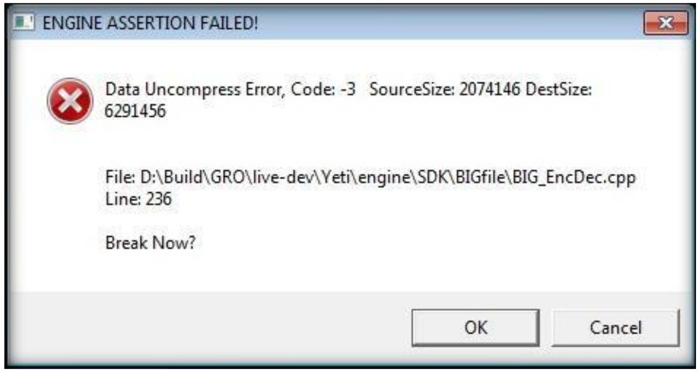
- Existence of general *duty* of beneficence debatable



Data Sharing

Con

- Existence of general *duty* of beneficence debatable
- Potential financial cost to citizen scientists plus time, effort, frustration
    - Citizen scientists unlikely to have funding/support staff
    - Citizen scientists unlikely to benefit (in reputational and financial terms) to same extent as professional scientists

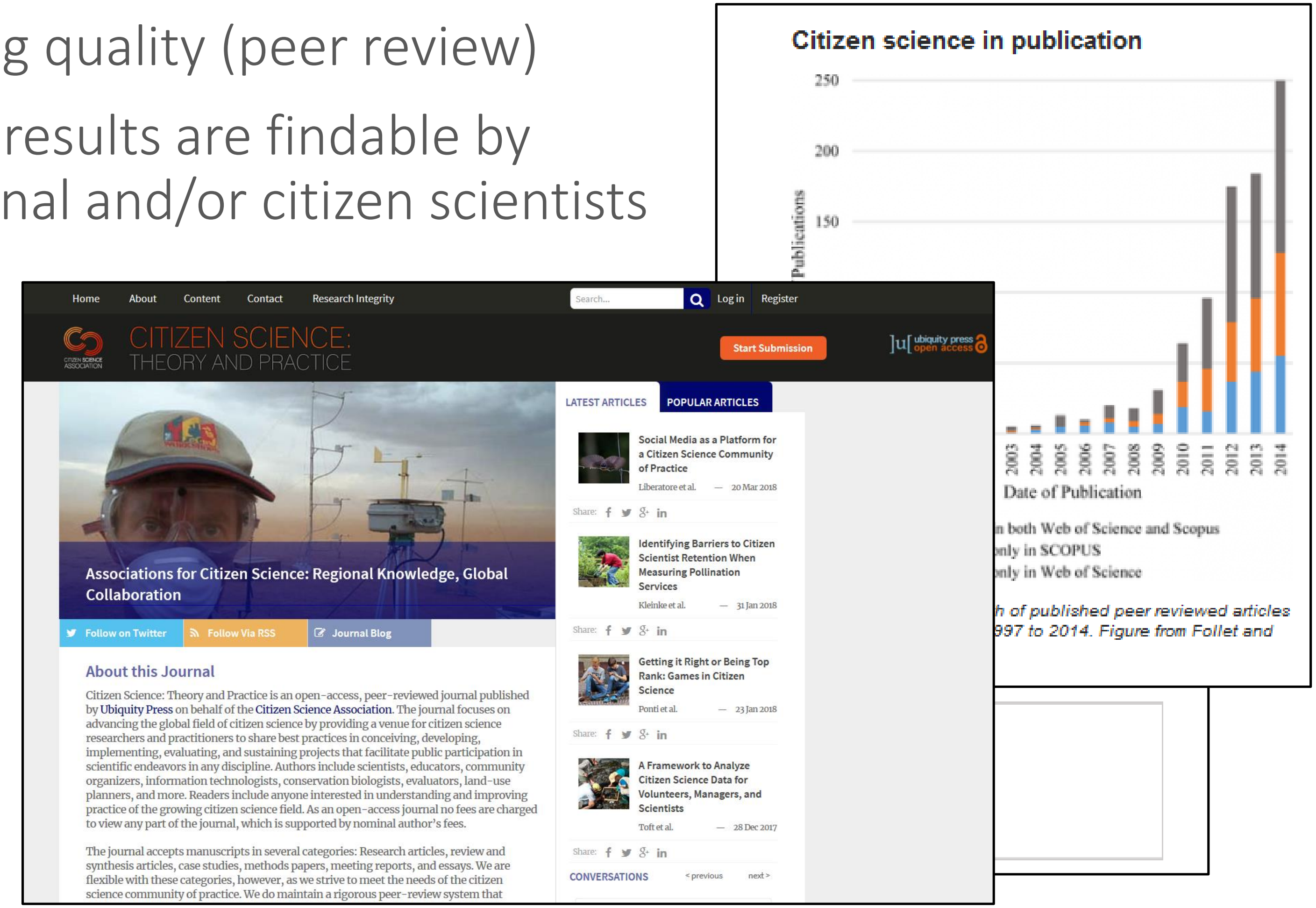


```
ENGINE ASSERTION FAILED!

Data Uncompress Error, Code: -3   SourceSize: 2074146 DestSize:
6291456

File: D:\Build\GRO\live-dev\Yeti\engine\SDK\BIGfile\BIG_EncDec.cpp
Line: 236

Break Now?
```

Journal Publication

Role in

- Enhancing quality (peer review)
- Ensuring results are findable by professional and/or citizen scientists



Associations for Citizen Science: Regional Knowledge, Global Collaboration

Social Media as a Platform for a Citizen Science Community of Practice
Liberatore et al. — 20 Mar 2018

Identifying Barriers to Citizen Scientist Retention When Measuring Pollination Services
Kleinke et al. — 31 Jan 2018

Getting it Right or Being Top Rank: Games in Citizen Science
Ponti et al. — 23 Jan 2018

A Framework to Analyze Citizen Science Data for Volunteers, Managers, and Scientists
Toft et al. — 28 Dec 2017

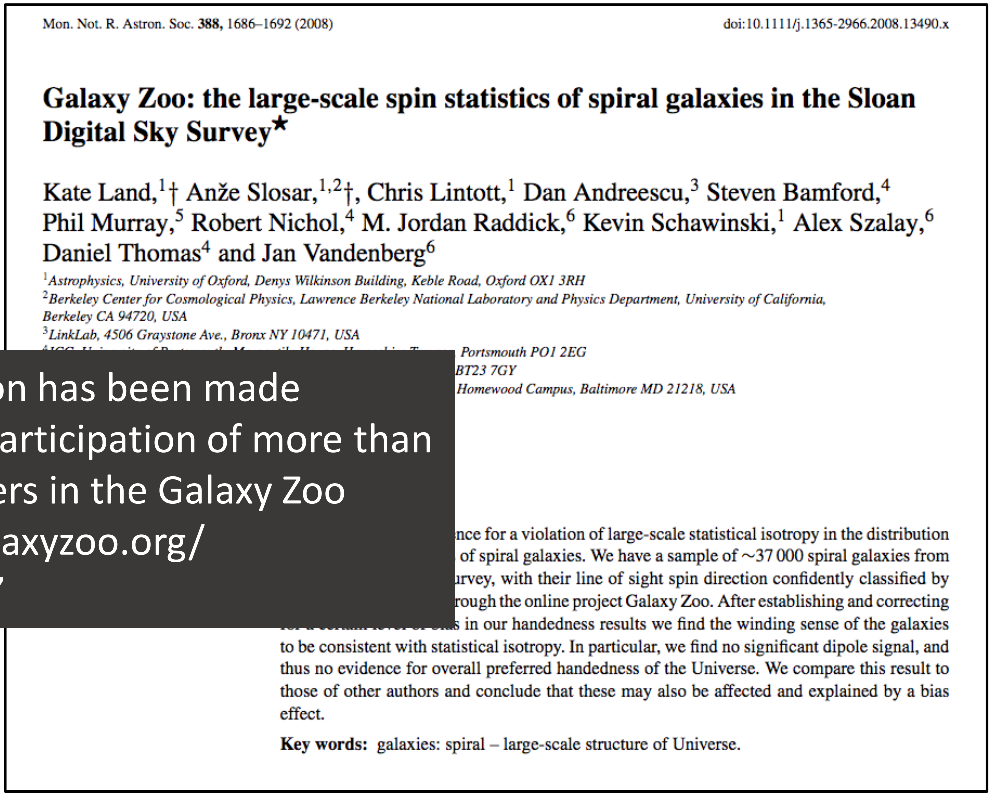
About this Journal

Citizen Science: Theory and Practice is an open-access, peer-reviewed journal published by Ubiquity Press on behalf of the Citizen Science Association. The journal focuses on advancing the global field of citizen science by providing a venue for citizen science researchers and practitioners to share best practices in conceiving, developing, implementing, evaluating, and sustaining projects that facilitate public participation in scientific endeavors in any discipline. Authors include scientists, educators, community organizers, information technologists, conservation biologists, evaluators, land-use planners, and more. Readers include anyone interested in understanding and improving practice of the growing citizen science field. As an open-access journal no fees are charged to view any part of the journal, which is supported by nominal author's fees.

The journal accepts manuscripts in several categories: Research articles, review and synthesis articles, case studies, methods papers, meeting reports, and essays. We are flexible with these categories, however, as we strive to meet the needs of the citizen science community of practice. We do maintain a rigorous peer-review system that

Barriers/Concerns

- Requirements such as IRB review and data sharing
- Overgeneralization of quality concerns/prejudice
- Complex questions re authorship vs. acknowledgement



Mon. Not. R. Astron. Soc. **388,** 1686–1692 (2008) doi:10.1111/j.1365-2966.2008.13490.x

Galaxy Zoo: the large-scale spin statistics of spiral galaxies in the Sloan Digital Sky Survey★

Kate Land,[1]† Anže Slosar,[1,2]†, Chris Lintott,[1] Dan Andreescu,[3] Steven Bamford,[4] Phil Murray,[5] Robert Nichol,[4] M. Jordan Raddick,[6] Kevin Schawinski,[1] Alex Szalay,[6] Daniel Thomas[4] and Jan Vandenberg[6]

[1]Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH
[2]Berkeley Center for Cosmological Physics, Lawrence Berkeley National Laboratory and Physics Department, University of California, Berkeley CA 94720, USA
[3]LinkLab, 4506 Graystone Ave., Bronx NY 10471, USA
Portsmouth PO1 2EG
BT23 7GY
Homewood Campus, Baltimore MD 21218, USA

nce for a violation of large-scale statistical isotropy in the distribution of spiral galaxies. We have a sample of ∼37 000 spiral galaxies from urvey, with their line of sight spin direction confidently classified by rough the online project Galaxy Zoo. After establishing and correcting s in our handedness results we find the winding sense of the galaxies to be consistent with statistical isotropy. In particular, we find no significant dipole signal, and thus no evidence for overall preferred handedness of the Universe. We compare this result to those of other authors and conclude that these may also be affected and explained by a bias effect.

Key words: galaxies: spiral – large-scale structure of Universe.

"* This publication has been made possible by the participation of more than 100 000 volunteers in the Galaxy Zoo project: www.galaxyzoo.org/Volunteers.aspx."

Multi-modal Strategies

CULTURE OF HEALTH

By Meredith Barrett, Veronica Combs, Jason G. Su, Kelly Henderson, Michael Tuffli, and The AIR Louisville Collaborative

DOI: 10.1377/hlthaff.2017.1315
HEALTH AFFAIRS 37,
NO. 4 (2018): 525–534

By Veronica Combs

Just as healthcare is moving far beyond the doctor's office, science is branching out beyond labs and universities. AIR Louisville is one of many projects featured in this public TV series that comes out in spring 2017.

The Crowd and the Cloud is a 4-part public television series exploring citizen science in the age of mobile technology.

The series explains how regular folks — the Crowd — can work together — using the Cloud — to solve some of today's largest challenges. C&C's vision is that citizen science combined with big data and mobile technology can serve as an early warning systems for earthquakes, epidemics, floods, and climate change. The project is funded by the National Science Foundation.

end of the article.

Cross-sector collaborations that bring together public, private, and philanthropic partners present a useful approach for addressing complex health issues,[1] such as infectious disease,[2] obesity and tobacco use,[3] oral health,[4] and natural disaster preparedness.[5] However, the knowledge base for developing effective partnerships and understanding how they create change remains limited. In this study we documented to more than \$56 billion annually in direct and indirect costs,[7] as well as impaired quality of life and increased mortality,[7–9] asthma accounts for more than 10.5 million missed days of school and 14.2 million missed days of work annually.[8] It is influenced by environmental factors such as air pollution, which can increase the odds of disease onset and exacerbate symptoms.[10]

The use of digital health technologies as an aid to the self-management and clinical treatment of