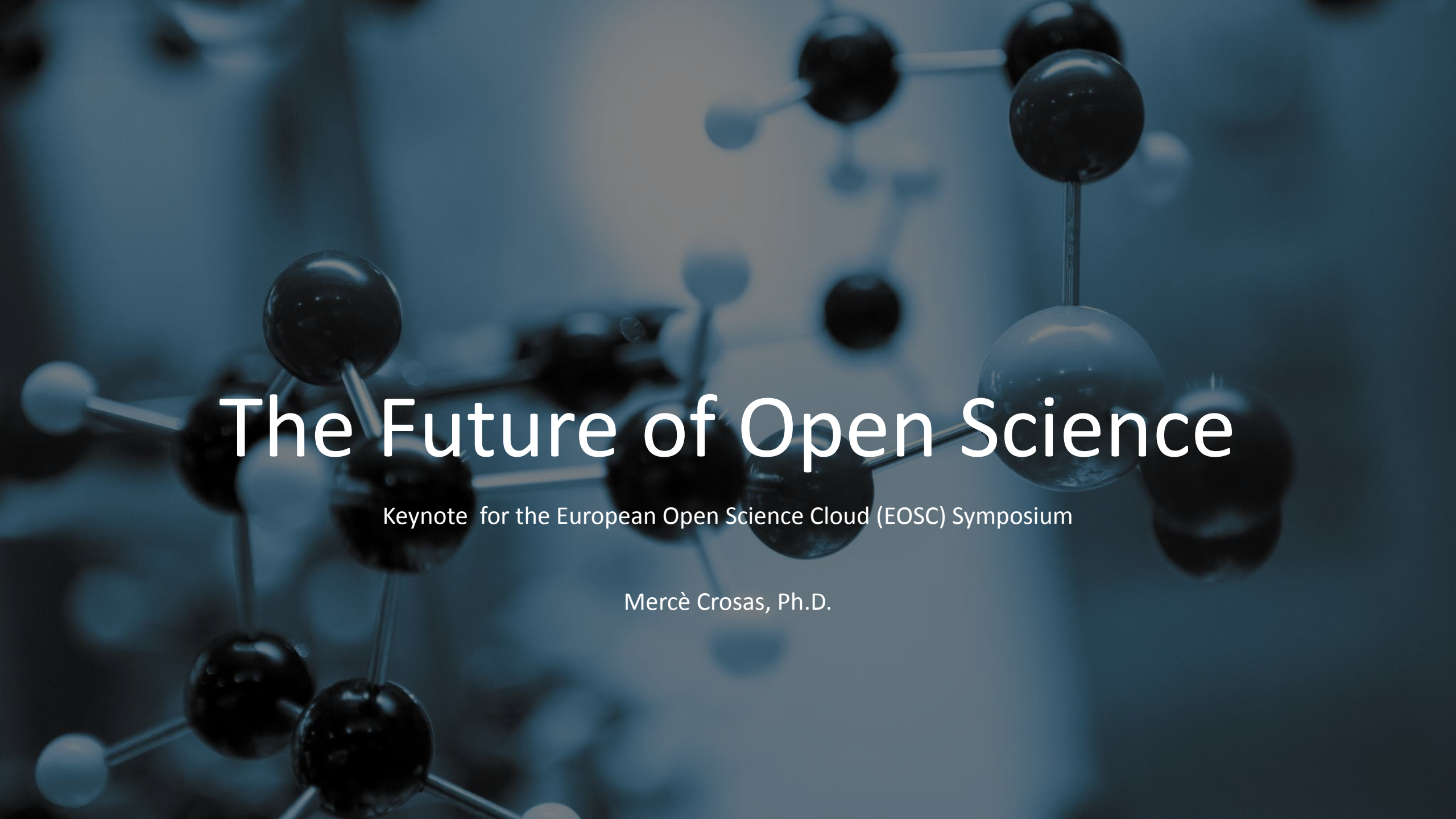


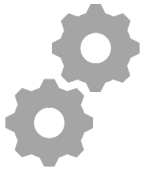
The Future of Open Science

Keynote for the European Open Science Cloud (EOSC) Symposium

Mercè Crosas, Ph.D.

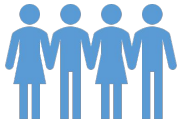
Open Science should support Today's Science

Increasingly collaborative



Increasingly automated

Responsible towards data



Increasingly relevant to society

Open Science should support Today's Science

Increasingly collaborative

Increasingly automated

Responsible towards data

Increasingly relevant to society

Recent studies released by the Organization for Economic Cooperation and Development (2022)

- References:
  - https://doi.org/10.1787/08f79edd-en
  - https://doi.org/10.1787/e11c5274-en

Co-creation during COVID-19: Study Details

- COVID-19 was a testbed for collaborations practices
- Study by OECD analyzed 30 international co-creation initiatives

Distinctive features shown in the study's initiatives:

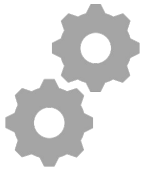
- Collaboration helped grew the initiative from the **bottom-up**
- There was active involvement by **governments**
- Made use of large global and national **networks** and **existing programs and infrastructures**

Co-creation during COVID-19: Lessons Learned

1. **Purpose is the strongest driver of co-creation:** a common shared goal.
2. **Networks & infrastructures** should be strengthened.
3. There is room for building **new collaborations between researchers and producers** to accelerate innovation.
4. Support the wider development and use of digital tools: **platforms for collaboration, data sharing, open challenges.**
5. New approaches to **leverage the wide diversity of actors.**
6. **Government´s involvement as network builders** can help speed up solutions: help **link research with industry.**
7. **User engagement can help uptake of new solutions**

Open Science should support Today's Science

Increasingly collaborative



Increasingly automated

Responsible towards data

Increasingly relevant to society

Study organized by U. S. National Academies of Sciences, Engineering, and Medicine (2022)

Study Committee members and NASEM Staff

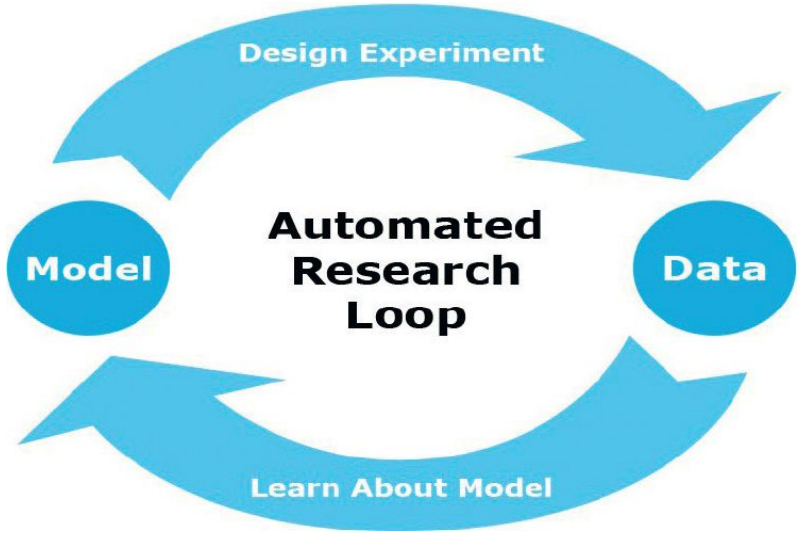
- Dan Atkins (Chair) – University of Michigan
- Ilkay Altintas – San Diego Supercomputer Center
- Shreyas Cholia – Lawrence Berkeley National Laboratory
- Mercè Crosas – Harvard University, Government of Catalunya
- Alfred Hero – University of Michigan
- Rebecca Lawrence – F100 Research Ltd, London
- Bradley Mailin – Vanderbilt University
- Lara Magravite – Sage Bionetworks
- Tapio Schneider – California Institute of Technology
- Tom Arrison, National Academies Study Director
- Emi Kameyama, National Academies Program Officer

Reference:

- https://doi.org/10.17226/26532

What is an Automated Research Workflow (ARW)?

- ARWs are scientific **research processes** emerging across a **variety of disciplines** and fields
- ARWs **integrate computation, laboratory automation, and tools from AI** in the performance of tasks that make up the research process, such as designing experiments-observations-simulations, collecting and analyzing data, and learning from the results to inform further experiments-observations-simulations
- Specific tools and resources vary by field, but the common goal is to **accelerate scientific knowledge generation while achieving greater control and reproducibility**



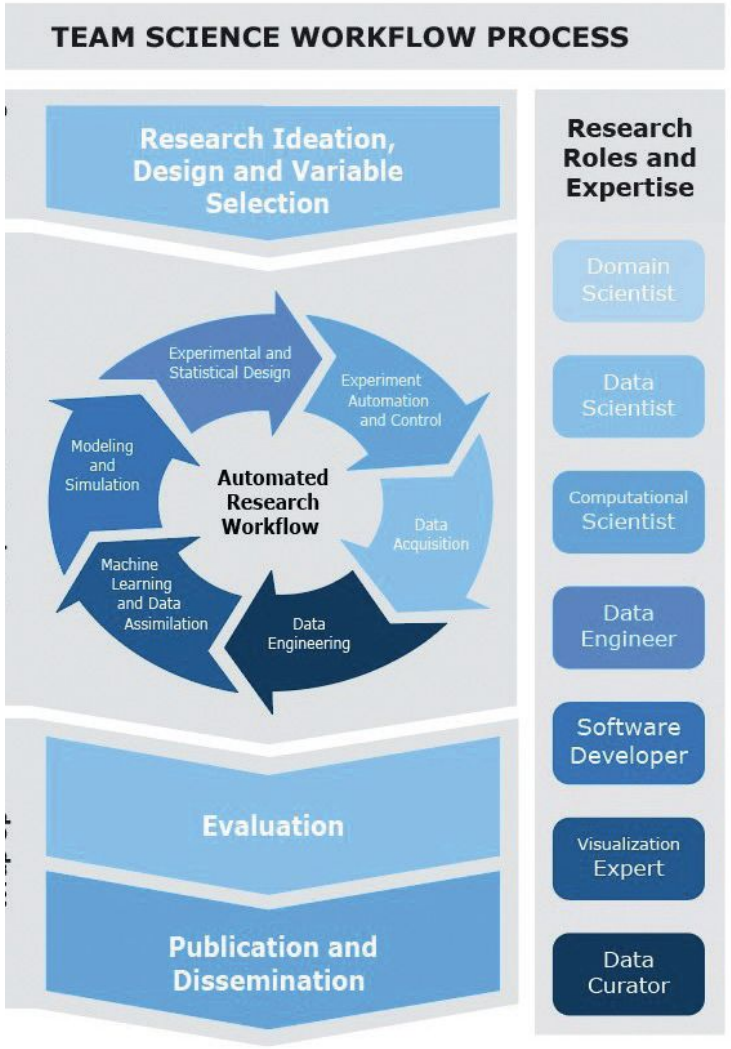
Examples of ARW in action

- **Material Science** - Cut the time required for synthesis and testing of materials from 9 months to 5 days
- **Particle Physics** - Allow experiments to achieve a given sensitivity with ½ the data
- **Drug Discovery** - Identify 57 percent of active compounds performing 2.5% of possible experiments, compared with 20% identified with traditional approach
- **Astronomy** - Automate telescope target selection so that observations are optimally informative given constraints and scientific objectives.
- **Digital Humanities** - Compile information from enormous volumes of words across multitudes of languages over centuries to see patterns in how ideas have spread and changed over time, and to understand the development of human thought.

Multi-disciplinary Team Science

ARW require expertise in multiple disciplines:

- Methods to manage, integrate and interpret large amounts of data
- Modeling and simulation tools executing on scalable computing platforms
- Methods and interfaces for analysis, communication and visualization of the results
- Technologies to make the process FAIR, that is, transparent, repeatable and reproducible



Study Recommendations

1. ARW **design principles** should:
   a. Facilitate openness, reproducibility, and transparency,
   b. Incorporate principles of responsible AI,
   c. Prioritize reuse and sustainability,
   d. Be driven and controlled by the research community.
2. Further progress on **openness, sustainability and sharing of infrastructure, instruments, code, and data** is required.
3. Research funders, institutions, and professional societies should cooperate in **supporting the education and training** required for creating and using ARWs.
4. An enhanced **culture of sharing and multi-disciplinary collaboration, with incentives** to do so is critical to the creation of ARW-based research.
5. Preservation of privacy must be robustly addressed in ARW world.

Open Science should support Today's Science

Increasingly collaborative

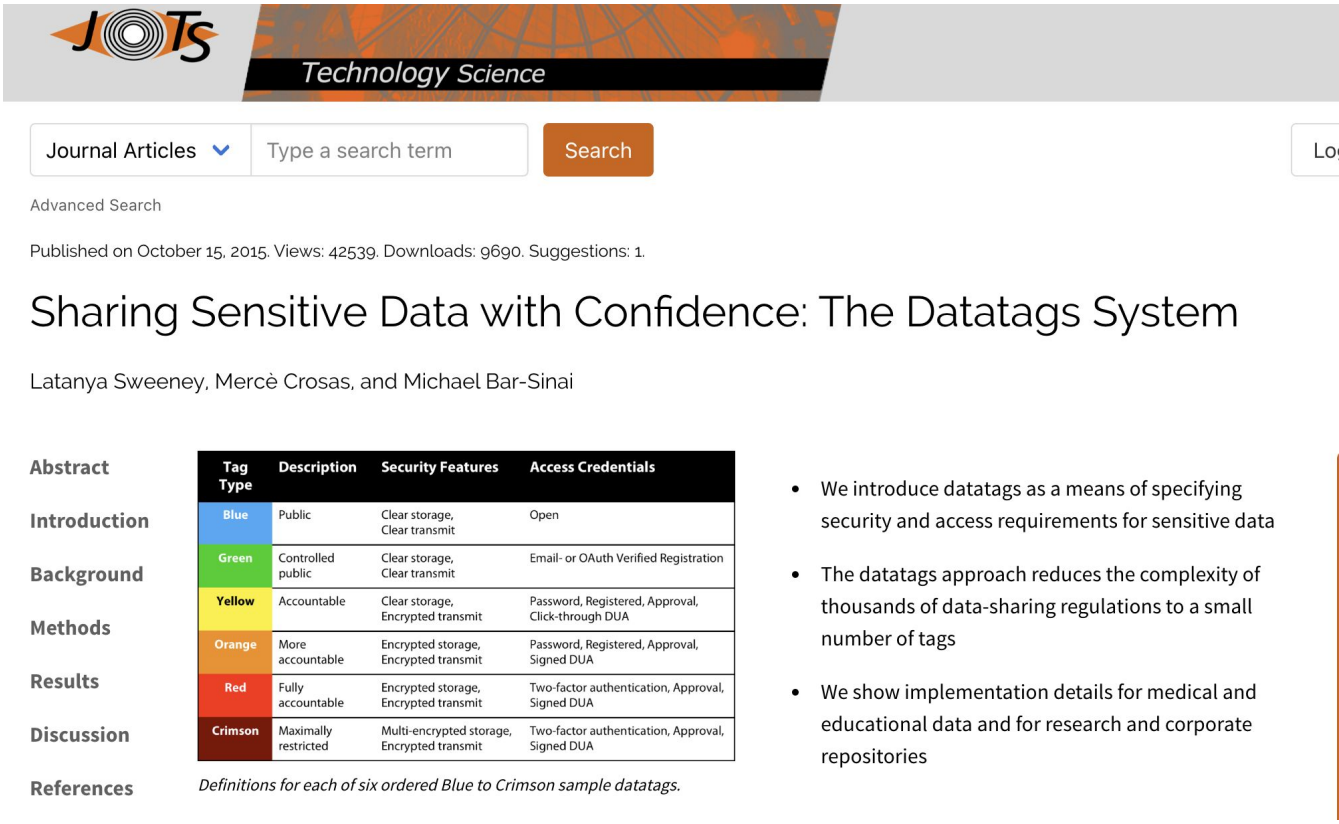
Increasingly automated

Responsible towards data

Increasingly relevant to society

Sharing Sensitive Data with Confidence: The DataTags system

- Sweeney, Crosas, Bar-Sinai (2015)
- Reference:
  - https://techscience.org/a/2015101601/



Technology Science

Journal Articles | Type a search term | Search

Advanced Search

Published on October 15, 2015. Views: 42539. Downloads: 9690. Suggestions: 1.

Sharing Sensitive Data with Confidence: The Datatags System

Latanya Sweeney, Mercè Crosas, and Michael Bar-Sinai

Abstract
Introduction
Background
Methods
Results
Discussion
References

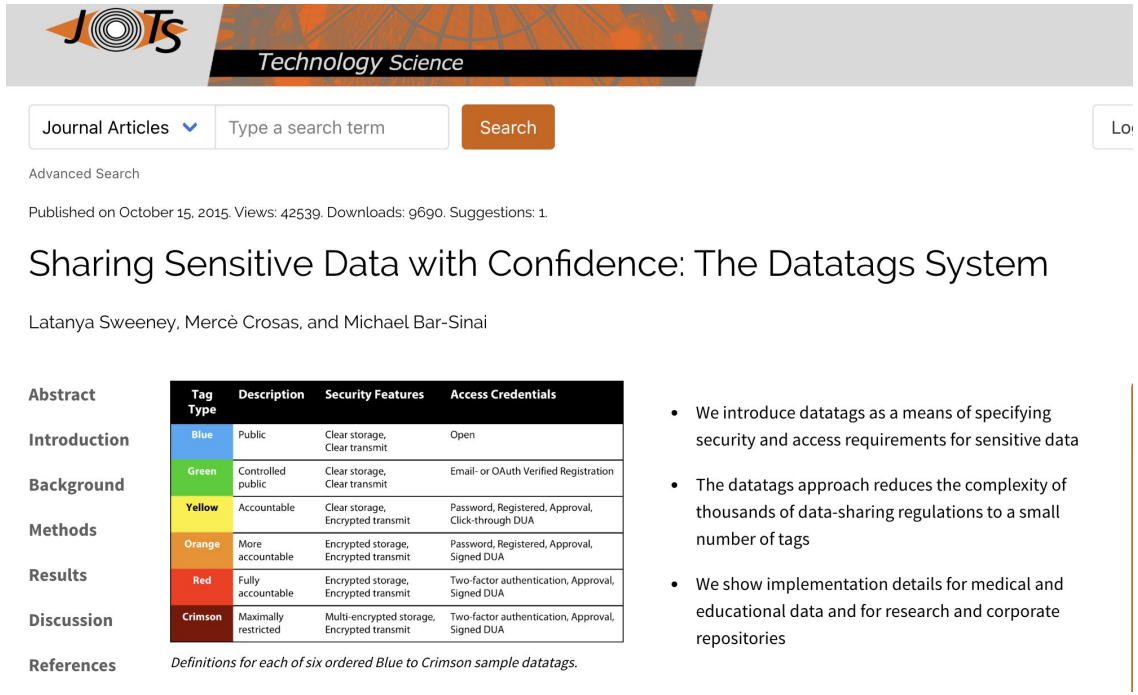
Tag Type	Description	Security Features	Access Credentials
Blue	Public	Clear storage, Clear transmit	Open
Green	Controlled public	Clear storage, Clear transmit	Email- or OAuth Verified Registration
Yellow	Accountable	Clear storage, Encrypted transmit	Password, Registered, Approval, Click-through DUA
Orange	More accountable	Encrypted storage, Encrypted transmit	Password, Registered, Approval, Signed DUA
Red	Fully accountable	Encrypted storage, Encrypted transmit	Two-factor authentication, Approval, Signed DUA
Crimson	Maximally restricted	Multi-encrypted storage, Encrypted transmit	Two-factor authentication, Approval, Signed DUA

Definitions for each of six ordered Blue to Crimson sample datatags.

- We introduce datatags as a means of specifying security and access requirements for sensitive data
- The datatags approach reduces the complexity of thousands of data-sharing regulations to a small number of tags
- We show implementation details for medical and educational data and for research and corporate repositories

What is a DataTags system?

- The **DataTag system** helps you store and share data in a manner that **respects legal commitments and ethical promises**
- A **datatag** is a **set of security features and access requirements** for file handling.



JOTS Technology Science

Journal Articles | Type a search term | Search | Lo

Advanced Search

Published on October 15, 2015. Views: 42539. Downloads: 9690. Suggestions: 1.

Sharing Sensitive Data with Confidence: The Datatags System

Latanya Sweeney, Mercè Crosas, and Michael Bar-Sinai

Abstract
Introduction
Background
Methods
Results
Discussion
References

Tag Type	Description	Security Features	Access Credentials
Blue	Public	Clear storage, Clear transmit	Open
Green	Controlled public	Clear storage, Clear transmit	Email- or OAuth Verified Registration
Yellow	Accountable	Clear storage, Encrypted transmit	Password, Registered, Approval, Click-through DUA
Orange	More accountable	Encrypted storage, Encrypted transmit	Password, Registered, Approval, Signed DUA
Red	Fully accountable	Encrypted storage, Encrypted transmit	Two-factor authentication, Approval, Signed DUA
Crimson	Maximally restricted	Multi-encrypted storage, Encrypted transmit	Two-factor authentication, Approval, Signed DUA

Definitions for each of six ordered Blue to Crimson sample datatags.

- We introduce datatags as a means of specifying security and access requirements for sensitive data
- The datatags approach reduces the complexity of thousands of data-sharing regulations to a small number of tags
- We show implementation details for medical and educational data and for research and corporate repositories

Examples of DataTags Levels

Non-sensitive Data

- **Blue:** completely open
- **Green:** access with registration
- **Yellow:** access with data user agreement

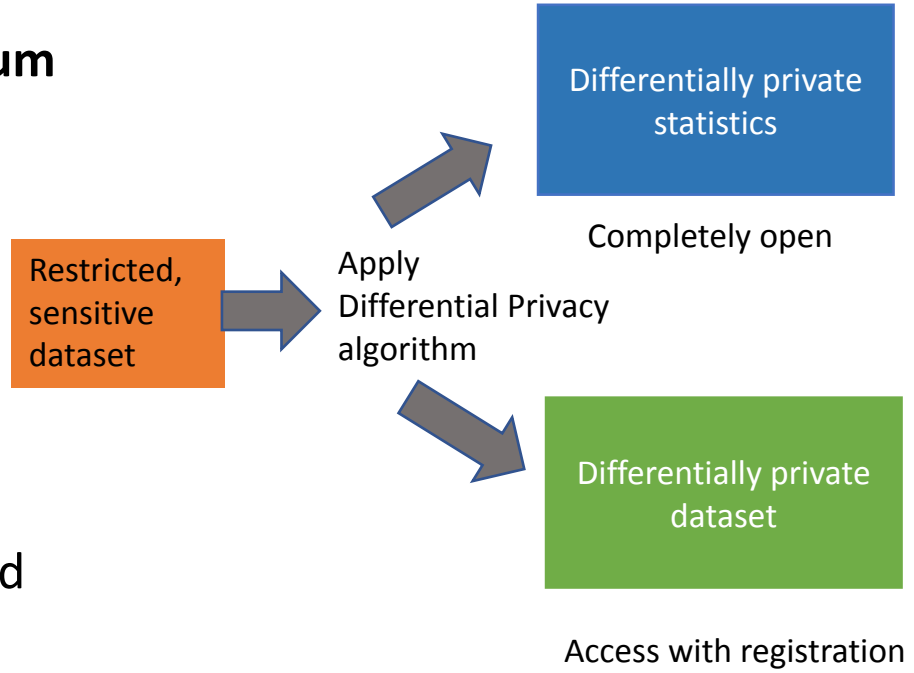
Sensitive Data (stored encrypted)

- **Orange:** signed user agreement
- **Red:** two-factor authentication
- **Crimson:** fully restricted

DataTags + Differential Privacy

Differential Privacy can facilitate opening sensitive data while preserving privacy:

- A differentially private (DP) algorithm introduces a **minimum amount of noise** to released statistics to **mathematically guarantee the privacy of any individual** in a dataset
- Aims to [Dwork, McSherry, Nissim, Smith, '06]:
  - enable statistical analysis of datasets - **utility**
  - while protecting individual level data – **privacy**
- In the last years, DP has moved from theory to practice and starts to be deployed in products

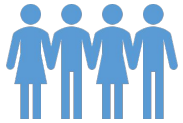


Open Science should support Today’s Science

Increasingly collaborative

Increasingly automated

Responsible towards data



Increasingly relevant to society

Open Government: the care and feeding of democracy (2022)

By Mercè Crosas (Secretary of Open Government, Catalan Government)

Reference:

http://www.gencat.cat/eapc/epum/N17/index.html

Open government: the care and feeding of democracy

Mercè Crosas
Secretary for Open Government, Generalitat de Catalunya (Catalonia)

Abstract

Open government provides us with a path towards a better quality of democracy, with more effective and trusted decision-making. It is a way of governing based on the principles of transparency, participation, and collaboration. After first presenting the historical evolution of this public tendency, the present article focuses on a new vision of open government in Catalonia, in which open government evolves into an open society, involving whole communities and local

EPuM
European Public Mosaic
June 2022
OPEN JOURNAL ON PUBLIC SERVICE
17
OPEN GOVERNMENT

ARTICLES
Mercè Crosas, Catalonia
Álvaro V. Ramírez-Alujas, Chile
Carina Lopes, Catalonia
Bapon Fakhruddin, New Zealand
Eva Frade & Olivier Schulbaum, Catalonia & Balearic Islands

INTERVIEW
Doreen Grove, Scotland

Good practices
New trends
Newsflash

Generalitat de Catalunya
Escola d'Administració Pública de Catalunya

Why open government?

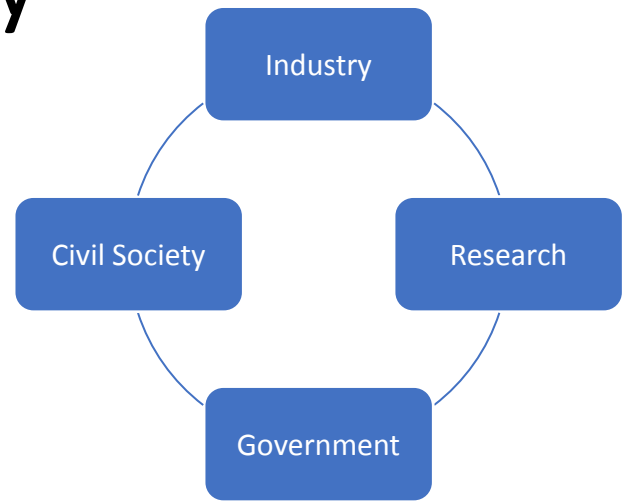
More than a decade after the first broad initiatives on open government were brought forward, open government is now more meaningful and necessary than ever. Time and time again, we witness that democracy cannot be taken for granted, that it needs nurturing, and that public trust in it can easily be lost.

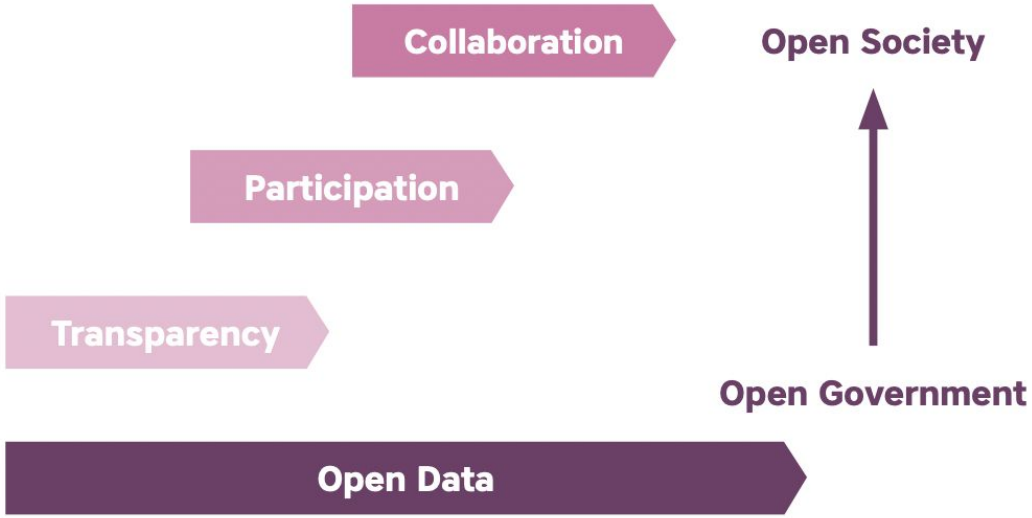
Open government provides us with the means to build a stronger democracy and more effective and efficient government. How? According to the Obama Administration's Commitment to Open Government Status Report in 2010, "Open government is a means, not an end". It is a way of governing based on three principles: **transparency, participation, and collaboration**. *Transparency* provides citizens with the right to know what the government is doing, for what purpose, and how decisions are being made, giving them access to data and arguments in an accessible way to facilitate the acquisition of this knowledge. Building upon *transparency*, participation provides citizens with the right to play a more active role in the definition, progress, and evaluation of public action through deliberative processes and consultation, beyond merely voting in elections, thereby helping the government create policies

Open government is a way of governing based on three principles: transparency, participation, and collaboration

Beyond the scientific community

- **To be relevant to society, open science must:**
  - Involve stakeholders outside science
  - Share data across industry, gov, and research
  - Engage citizens
  - Apply the principles of open government – transparency, participation, collaboration
  - Inform evidence-based policy-making
  - Help address societal problems
  - Towards a more open society





Summary

Increasingly collaborative

Increasingly automated

Responsible towards data

Increasingly relevant to society

The future of open science should be open to:

- Foster multidisciplinary and multiple actors' collaboration and co-creation
- Facilitate interoperation with the latest automated research workflows based on AI, machine-learning and automation tools, following FAIR principles
- Support levels of openness and new privacy-preserving methods to facilitate using sensitive data for research
- Engage citizens, governments and industry to help solve societal problems with the principles of an open society – transparency, participation, collaboration

Thanks!