Advancing Reproducible Science from Physical Samples

IGSN and the iSamples Research Coordination Network

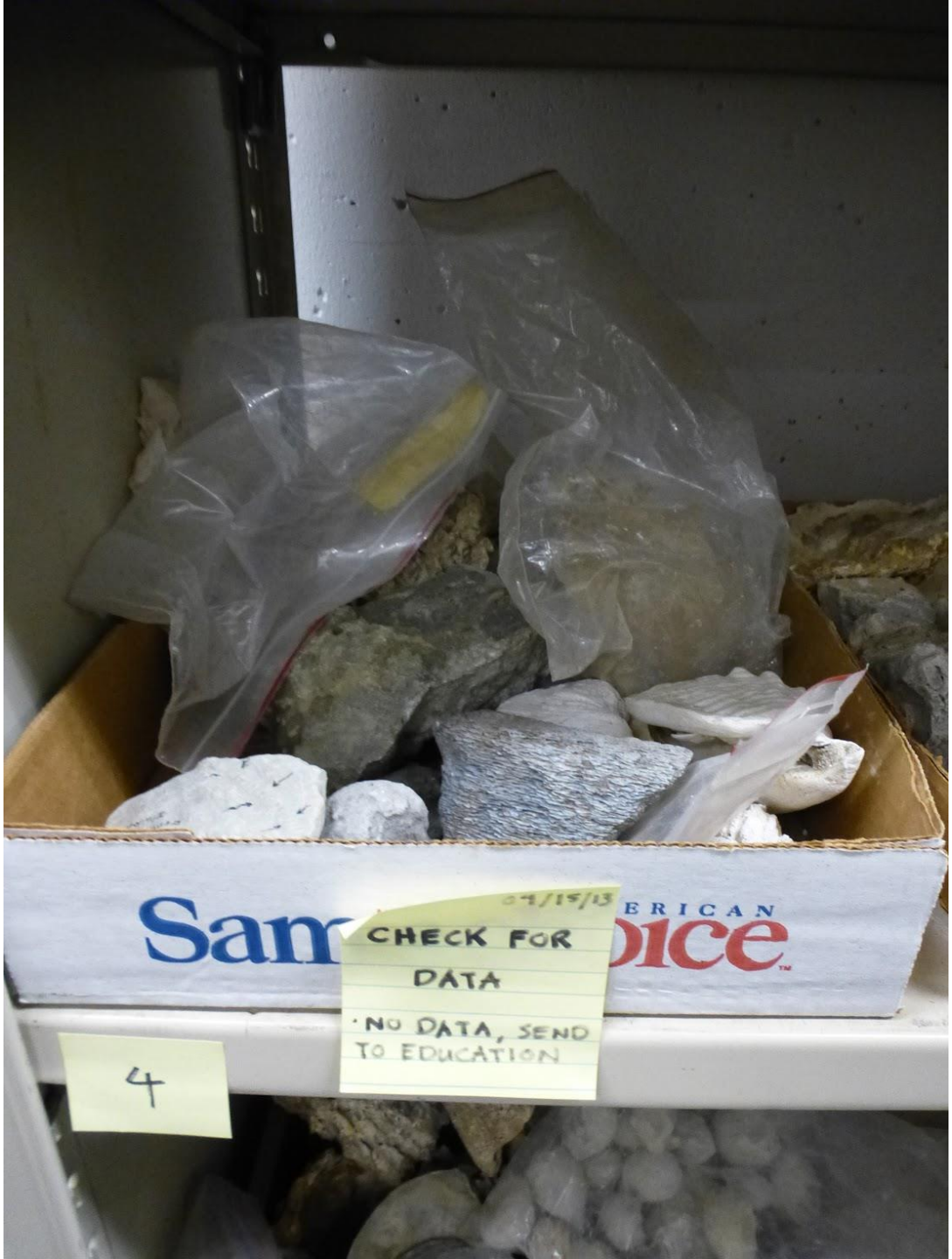
Kerstin Lehnert

Megan Carter

IEDA (Interdisciplinary Earth Data Alliance), www.iedadata.org

IGSN e.V., www.igsn.org

Lamont-Doherty Earth Observatory, Columbia University


04/15/13
CHECK FOR
DATA
· NO DATA, SEND
TO EDUCATION
4

nature geoscience

Search

Journal home > Focus > Transparency in science

Journal content

- Journal home
- Advance online publication
- Current issue
- Archive
- Insights

Focus

Transparency in science

Editorial
Commentary

AGU Blogosphere
American Geophysical Union
A community o

HOME BLOGS ABOUT

GeoSpace
Earth and space science

HOME ABOUT CONTACT US

Home - Data - Open data: Creating a culture of transparency and reproducibility in science

3 MARCH 2016

Open data: Creating a culture of transparency and reproducibility in science

Posted by nbompey

G+1 0 Like 6 Tweet Share 4

By Rebecca Fowler

Science AAAS

Home News Journals Topics Career

Science Science Advances Science Immunology Science Robotics Scie

SHARE

EDITORIAL

Reproducibility

Marcia McNutt

Marcia McNutt is Editor-in-Chief of Science.

Science 17 Jan 2014:
Vol. 343, Issue 6168, pp. 229
DOI: 10.1126/science.1250475

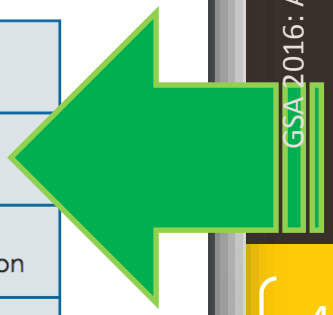
TRANSPARENCY AND OPENNESS PROMOTION (TOP) GUIDELINES

THE GUIDELINES

Transparency and independent replication are core values of science. However, scientists must publish, which is more likely with positive and tidy results, even at the expense of transparent, reproducible research. The TOP Guidelines were created by a diverse group of science publishers, editors, and funders in order to **align scientific values with its rewards.**

8 MODULAR STANDARDS

Citation Standards Describes citation of data	**Data Transparency** Describes availability and sharing of data
Analytical Methods Transparency Describes analytical code accessibility	**Research Materials Transparency** Describes research materials accessibility
Design and Analysis Transparency Sets standards for research design disclosures	**Preregistration of Studies** Specification of study details before data collection
Preregistration of Analysis Plans Specification of analytical details before data collection	**Replication** Encourages publication of replication studies



Source: https://cos.io/top/

TRANSPARENCY AND OPENNESS PROMOTION (TOP) GUIDELINES

ACROSS 3 TIERS

1 DISCLOSURE: the final research output must disclose if the work satisfies the standard

2 REQUIREMENT: the final research output must satisfy the standard

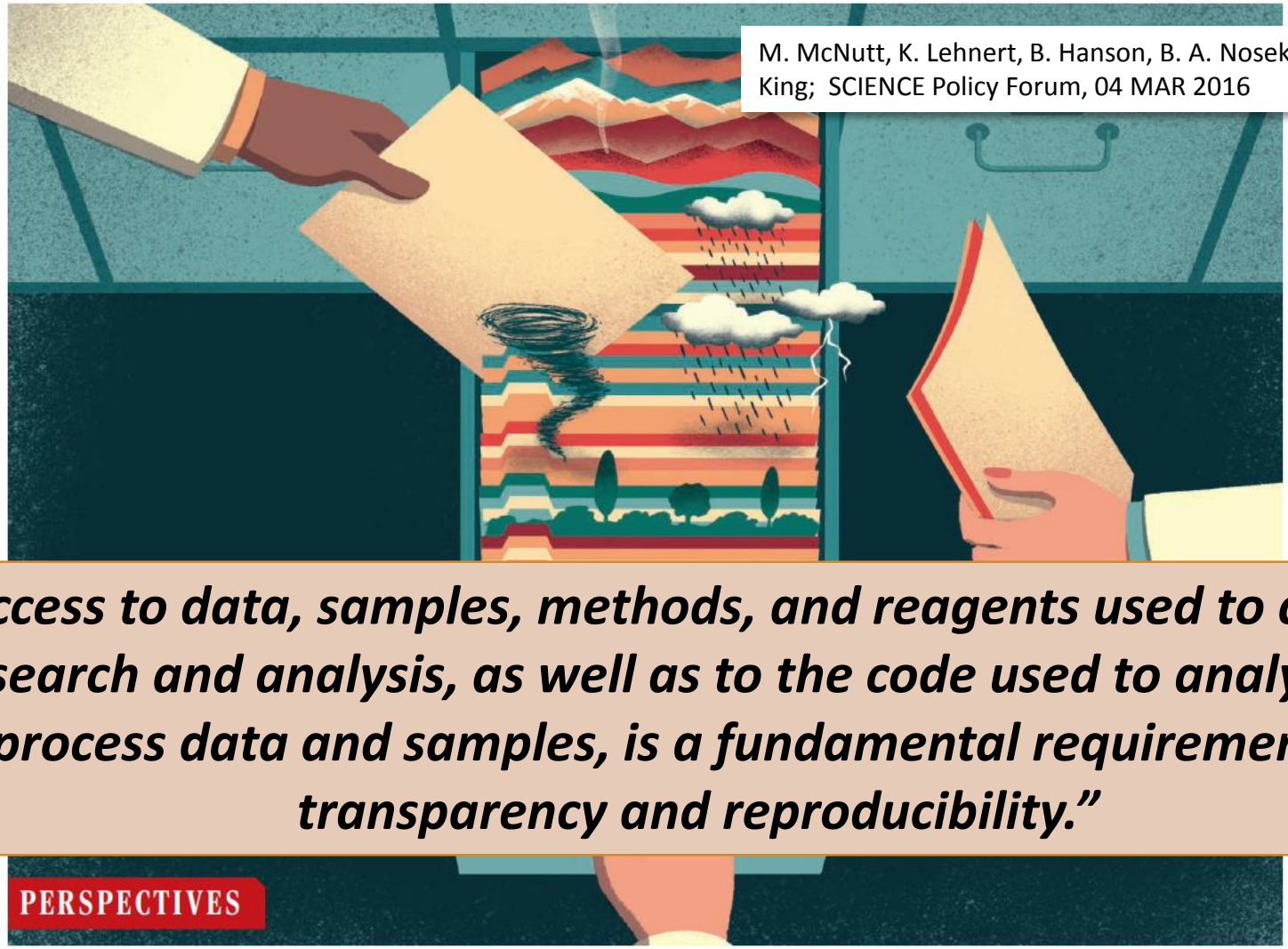
3 VERIFICATION: third party must verify that the standard is being met

8 MODULAR STANDARDS

Citation Standards Describes citation of data	**Data Transparency** Describes availability and sharing of data
Analytical Methods Transparency Describes analytical code accessibility	**Research Materials Transparency** Describes research materials accessibility
Design and Analysis Transparency Sets standards for research design disclosures	**Preregistration of Studies** Specification of study details before data collection
Preregistration of Analysis Plans Specification of analytical details before data collection	**Replication** Encourages publication of replication studies

Source: https://cos.io/top/

M. McNutt, K. Lehnert, B. Hanson, B. A. Nosek, A. M. Ellison, J. L. King; SCIENCE Policy Forum, 04 MAR 2016



"Access to data, samples, methods, and reagents used to conduct research and analysis, as well as to the code used to analyze and process data and samples, is a fundamental requirement for transparency and reproducibility."

on March 6, 2016

PERSPECTIVES

RESEARCH INTEGRITY

Liberating field science samples and data

Promote reproducibility by moving beyond "available upon request"

M. McNutt, K. Lehnert, B. Hanson, B. A. Nosek, A. M. Ellison, J. L. King; SCIENCE Policy Forum, 04 MAR 2016

on March 6, 2016

PERSPECTIVES

RESEARCH INTEGRITY

Liberating field science samples and data

Promote reproducibility by moving beyond "available upon request"

"Despite many efforts, there remains widespread disagreement regarding data and sample availability and metadata, as well as uneven sample deposition across the field sciences."

iSamples: Rationale

- Improve discovery, access, sharing, analysis, and curation of physical samples and the data generated by their study as needed by the science community
- Address requirements for reproducibility of sample-based data across domains.
  - **Best practices** for sample documentation, identification, and citation.
  - **Culture** that recognizes sample collection and curation as a scholarly contribution (citation & credit).
  - **Policies** regarding preservation and curation of physical objects and need physical infrastructure.

iSamples: An EarthCube RCN

- **Stakeholder Alignment Survey**
- **Early Career workshops**
- **Webinar series**
- **Two major community workshops** (UT Austin 2015, UNC/RENCI 2016)
- **Joint activities & events with related groups & projects** (ESIP, IGSN e.V., COPDESS, EarthCube RCNs & BBs)

The Internet of Samples aims to break down research barriers

Published: February 17, 2016

rage, but a group of scientists that met recently at (RENCI) wants to add the Internet of Samples

core samples of the Earth, or those collected by submersibles sent to the depths of the ocean.

Not only are these samples difficult to obtain, but once gathered, they

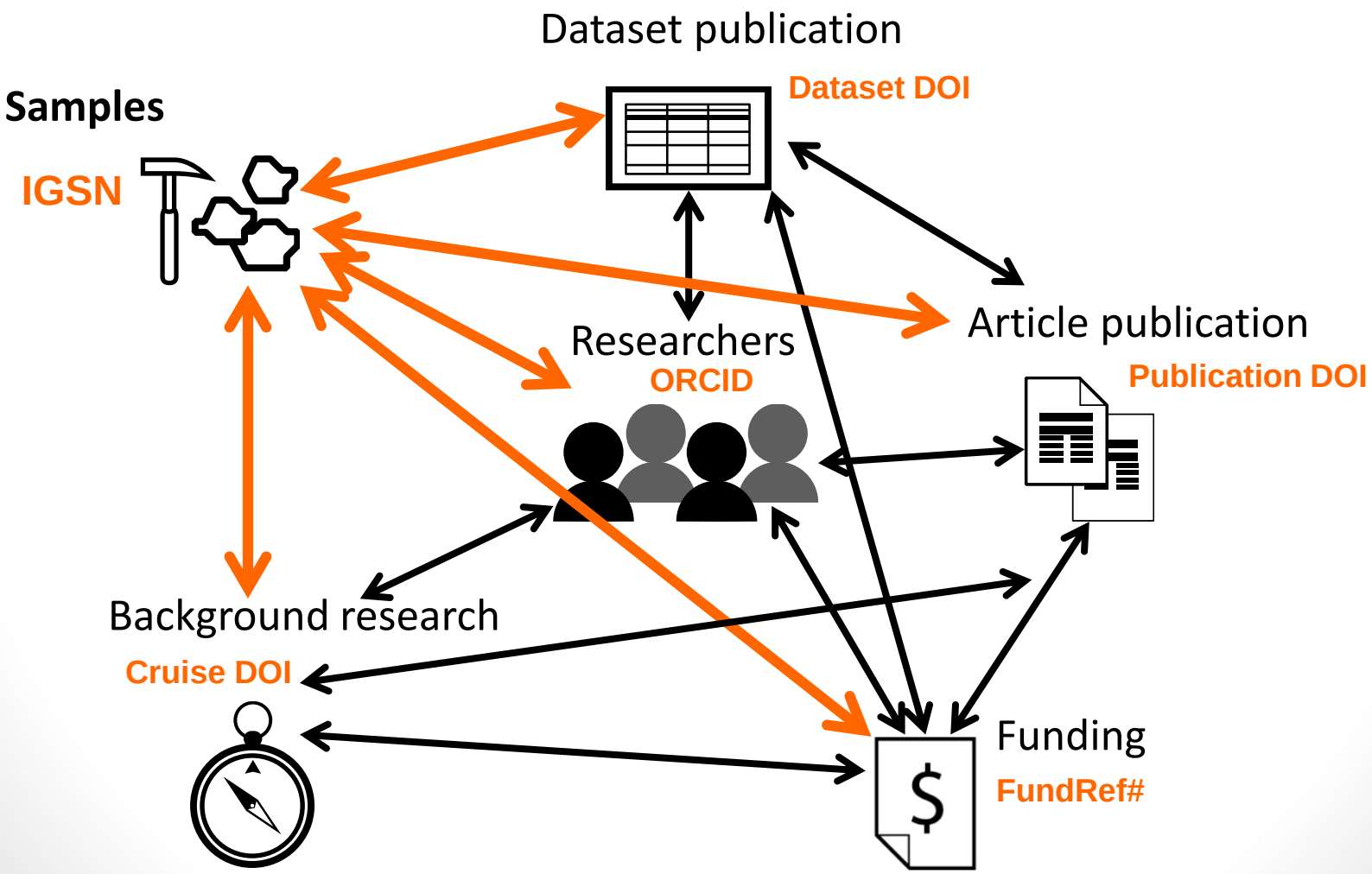
A multidisciplinary group of scientists and researchers brainstorm potential elements of the iSamples action plan for the coming year.

http://renci.org/news/the-internet-of-samples-aims-to-break-down-research-barriers/

iSamples Working Groups

- WG1: **Identifiers & Metadata**
  - Develop recommendations for best practices regarding use of unique sample identifiers, metadata, schema, and semantics
- WG2: **Use Cases**
  - collect user stories that articulate practices in the life cycle of samples for different users
  - identify curatorial points of intervention
  - develop policies and recommendations
- WG3: **Engagement & Training**
  - Enhance communication of the benefits and leading practices of documenting and sharing physical samples
  - develop and disseminate Sample Management Training Modules
- WG4: **Workflow Applications**
  - Requirements for tools to link local practices to standardized protocols

Use of Persistent Identifiers



Use of Persistent Identifiers

NOV
09

PIDapalooza

by ORCID, Crossref, California Digital Library, DataCite

Open identifiers deserve their own festival

Why build an open identifier infrastructure? So that anyone can use it to create cool tools and services for the research community.

IGSN: *International Geo Sample Number*

- Globally unique and persistent identifier for physical samples in the Earth Sciences
  - guaranteed to be unique via a centralized control mechanism operated by IGSN e.V.
  - resolves to virtual sample representations (sample metadata profiles) managed at federated IGSN Allocating Agents.

The implementation of IGSN in the Australian mineral exploration context

globally unique label for environmental samples that are taken out of the organisation IGSN e.V. CSIRO, Geoscience Australia and Curtin University, physical samples and to link to the derived datasets.

istration services

International IGSN registry

Registration interface

IGSN

Australian IGSN Portal – National Catalogue²

Physical Samples and Persistent Identifiers : The Implementation of the International Geo Sample Number (IGSN) Registration Service in CSIRO, Australia

Anusuriya Devaraju, Jens Klump, Victor Tey and Ryan Fraser
CSIRO Mineral Resources, PO Box 1130, Bentley, Western Australia, 6102, Australia.

MINERAL RESOURCES
www.csiro.au

Physical samples may not be easily discoverable due to their missing unique identification and the lack of standardization in

Persistent

- The Internatio unique label for
- IGSN is govern e.V.), who opera
- Clients, e.g., through allocat behalf of the m

IGSN Impl

- CSIRO became agents in Aust University to i
- To implement **metadata m harvester and**

GFZ Helmholtz Centre Potsdam
Poster-EGU2016-12405

IGSN ssdp icdp

GFZ German Research Centre for Geosciences

Utilization of the International GeoSample Number in the ICDP Drilling Project COSC

Ronald Conze (1), Henning Lorenz (2), Damian Ulbricht (1), Thomas Gorgas (1), Kirsten Elger (1)
ronald.conze@gfz-potsdam.de, henning.lorenz@geo.uu.se, damian.ulbricht@gfz-potsdam.de, thomas.gorgas@gfz-potsdam.de, kirsten.elger@gfz-potsdam.de
(1) GFZ German Research Centre for Geosciences, Potsdam, Germany, (2) Uppsala University, Department of Earth Sciences, Sweden

The **International GeoSample Number (IGSN)** is a unique identifier (UID) for physical samples that provides discovery functionality via the internet. The general aim of the IGSN concept is to

- provide easy access to and improve the utilization of sample material worldwide,
- enable the unique identification of each sample, its origin and provenance,
- facilitate citation of acquired samples throughout the literature.

The expedition COSC (Collisional Orogeny in the Scandinavian Caledonides) prompted for the **first time in ICDP's** (International C Program) **history to assign and register IGSNs du campaign**. ICDP drilling expeditions are commonly us System DIS for the inventory of recovered sample ma were assigned to every drill hole, core run, core secti the core material.

The original IGSN specification has been extended to achieve the required uniqueness of IGSNs with our offline-procedure

On-site **assignment of IGSNs** to drill holes, core runs, core sections, and samples using the COSC DIS

Transfer from the field to the core storage (ExpeditionDIS to CurationDIS)

Off-site **assignment of IGSNs** to samples taken during the sampling party in a core storage

PROJECT

Automatic g data from t System DIS

IGSN regi agent for I

Main Benefits:

- Each project has its own name space of open size
- Offline generation of unique IGSNs is possible even on simultaneously running projects
- A wide range of metadata are associated with each IGSN
- Parent-child relations between IGSNs define the Sample Family hierarchy ranging from Drill Hole>Core Run>Core Section>Sample
- Easy citation using the IGSN handle directly or a DOI (Digital-Object Identifier) of a corresponding paper or report
- IGSNs are already supported by several prominent publishers

Full **metadata retrieval** via the IGSN landing page on the Web

IGSN citation in publications

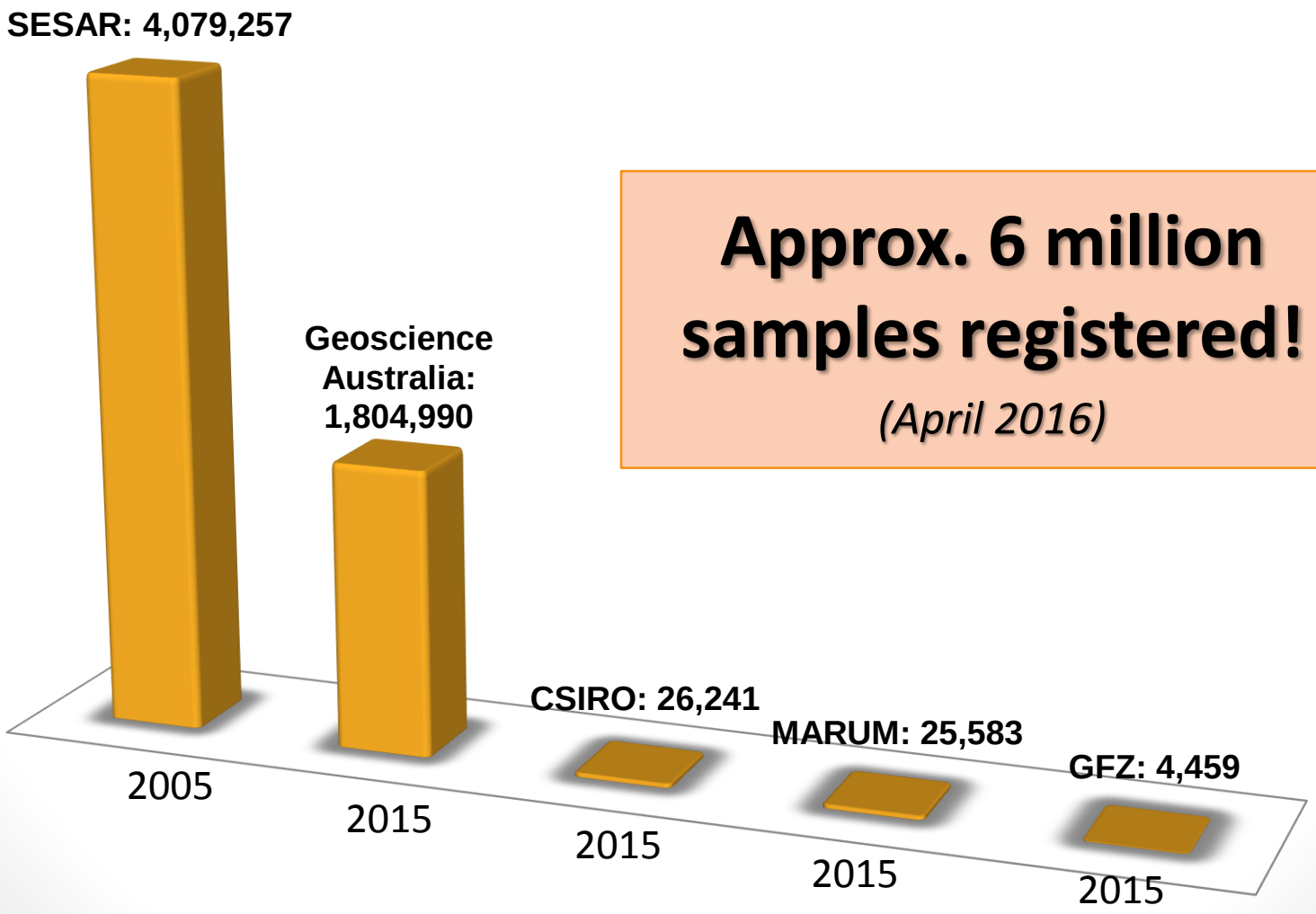
END USER/ SCIENTIST

References:
(1) Conze, R. (2016): Drilling Information System (DIS) and Core Scanner. Journal of large-scale research facilities, 2 A63.

Acknowledgements:
The development of the ICDP Drilling Information System was funded by th for Geosciences Potsdam, the German Research Foundation and ICDP. Expe financed by European Consortium for Ocean Research Drilling (ECORD) and software development was done by smartcube GmbH, Berlin, Germany (ht

Transfer from the field to the core storage (ExpeditionDIS to CurationDIS)

IGSN Status



Approx. 6 million samples registered!

(April 2016)

IGSN Adoption: Publishers

COPDESS Statement of Commitment

- Include in research papers concise statements indicating where data reside and clarifying availability.
- Promote and implement links to data sets in publications and corresponding links to journals in data facilities via persistent identifiers. Data sets should ideally be referenced using registered DOI's.
- Promote use of other relevant community permanent identifiers for samples (IGSN), researchers (ORCID), and funders and grants (FundRef).
- Develop workflows within the repositories that support the peer review process (for example, embargo periods with secure access) and within the editorial management systems that will ease transfer of data to repositories.

IGSN Adoption: Publishers

AGU is promoting use of new open community identifiers, such as International Geo Sample Numbers (IGSNs) for field samples like this coral. Today, AGU and seven other publishers have committed to including ORCIDs, a researcher identifier, in all published papers starting in 2016. Credit: ©Lamont-Doherty Core Repository

By Brooks Hanson 7 January 2016

Openly shared identification codes, such as digital object identifiers (DOIs) for journal articles, have greatly eased accessibility of online scientific papers. Now other open community identifiers for funders, institutions, field samples, and researchers are garnering support, demonstrating similar benefits, and being adopted by many publishers.

These digital identifiers hold great usefulness for securing integrity in science, providing efficiency, enabling scholarly communication and discovery, and aiding researchers in their work. Consequently, American

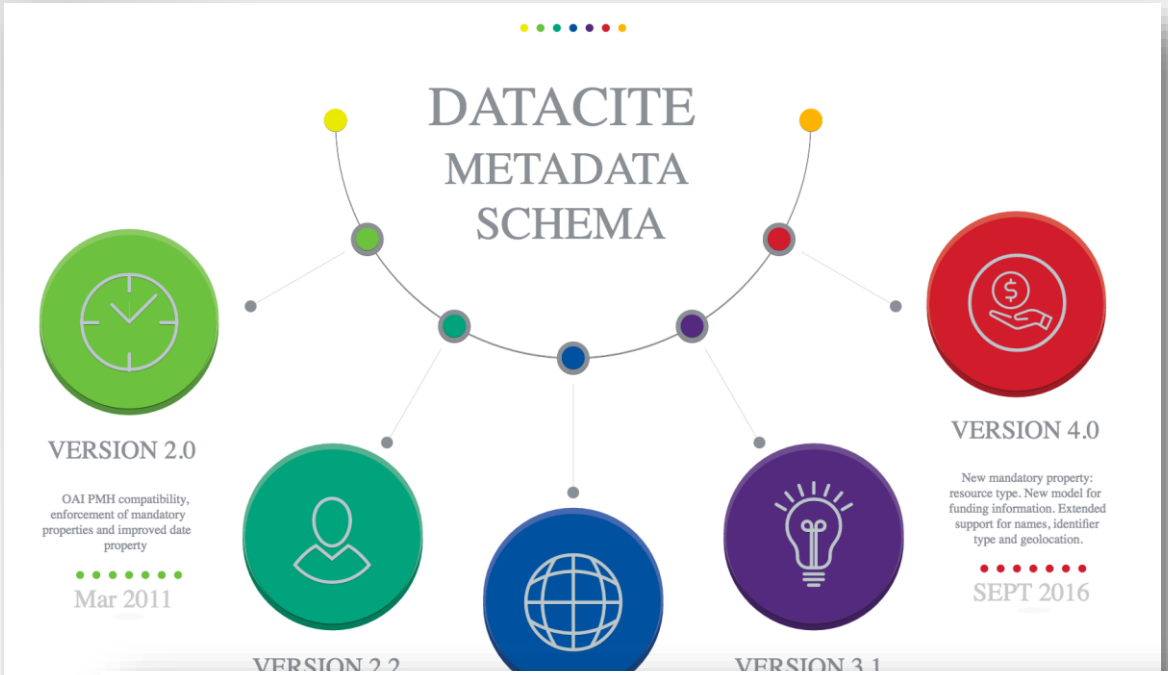
These digital identifiers hold great usefulness for securing integrity in science, providing efficiency, enabling scholarly communication and discovery, and aiding researchers in their work.

"… AGU Publications also strongly encourages use of other identifiers in our journal papers. International Geo Sample Numbers (IGSNs) uniquely identify items, such as a rock sample, a piece of coral, or a vial of water taken from the natural environment, and provide important, consistent information about these samples. Registering samples and including the IGSN in papers helps secure provenance information but most importantly connects common samples across multiple studies in the literature. IGSNs also will help you keep track of your samples. These identifiers can be reserved before a field season or assigned afterward."

Hanson, B. (2016), AGU opens its journals to author identifiers, Eos, 97, doi:10.1029/2016EO043183.

Published on 7 January 2016.

IGSN in DataCite



The DataCite Metadata Schema 4.0 introduces these changes:

- resourceTypeGeneral becomes a mandatory field (from optional)
- The new property FundingReference is added, with subproperties funderName, funderIdentifier, awardNumber, awardURI and awardTitle. This property deprecates contributorType 'funder'
- creatorName and contributorName have two new optional properties: familyName and givenName
- 'IGSN' has been added as a relatedIdentifierType option
- There is a new subproperty for GeoLocation 'geoLocationPolygon'
- geoLocationPoint and geoLocationBox have been updated

More IGSN News

- Consensus on standard IGSN metadata schema for describing samples (**"Birth Certificate"**) at workshop in Fall 2015
  - essential properties to describe a sample's origin & classification
  - published online at *schema.igsn.org*
- Prototype of **Central IGSN Metadata Catalog** developed
  - harvests "birth certificate" metadata records from Allocating Agents via the Open Archives Initiative Protocol for Metadata Harvest (OAI-PMH)
  - publishes metadata as a Linked Open Data graph for reuse by Semantic Web clients
- GBIF interest, discussion started at International Data Week in September 2016, workshop planned for Spring 2017
- More collaborations emerging: EZID, ORCID

Much Left To Do …

- Develop & implement policies for access and sharing of samples as part of reproducible research
- Create infrastructure for sample preservation and access
  - Capacity?
  - Business models?
  - Governance?
  - Certification of repositories?
- Software tools that support researchers and curators to comply with best practices
- Further coordination & alignment between Geo & Bio

Plans

- iSamples – IGSN workshop at AGU Fall Meeting
  - Sunday, Dec 11, 2016, 1pm
- iSamples workshop on 'Access & Preservation of Samples'
  - Planned for Spring 2017
- IGSN – GBIF workshop planned for Spring 2017 in Australia