Title:

Data management, sharing, and reuse in experimental geomorphology: challenges, strategies, and scientific opportunities

Authors:

Leslie Hsu¹

Raleigh L. Martin²

Brandon McElroy³

Kimberly Litwin-Miller⁴

Wonsuck Kim⁴

Affiliations:

¹ Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10964, USA

² Department of Atmospheric and Oceanic Sciences, University of California – Los Angeles, Los Angeles, CA 90095, USA

³ Department of Geology & Geophysics, University of Wyoming, Laramie, WY 82071, USA

⁴ Department of Geological Sciences, Jackson School of Geosciences, University of Texas, Austin, TX 78712, USA

Corresponding Author: Leslie Hsu

Phone: +1 (845) 365-8484

Email: hsu.leslie@gmail.com

Address: Geoinformatics, Lamont-Doherty Earth Observatory, 61 Route 9W, Palisades, NY 10964, USA

Abstract

The field of experimental geomorphology is in a data-rich era with rapid expansion of high-resolution, digital data sets. Millions of dollars have been invested in building and renovating flume laboratories that run experiments at increasing sophistication, scale, and resolution. However, this overflowing body of laboratory data is not easily analyzed, stored, or accessed. This lack of organization comes at substantial cost to the Earth-surface science community (i.e., geomorphologists, sedimentary geologists, and engineers) through missed opportunities to address scientific challenges. Here, we present an overview of the current state of data management practices and challenges within the context of the experimental lifecycle. Based on this assessment, we identify four areas of greatest need in order to achieve higher rates of data sharing and reuse: metadata guidelines, workflow documentation, data storage, and incentives and training. We suggest specific guidance for addressing these needs and summarize outstanding community debates regarding interoperability and human- versus machine-readability. A proposed metadata list includes basic data set information and disciplinary information for experimental geomorphology, including metadata for evaluating data quality and readiness for reuse. Data publication is presented as a framework for improving data sharing, and we discuss community-specific considerations for review of experimental data sets.

Keywords: experimental geomorphology; data management; data publication; data set review

1. Introduction

Until recently, experimental geomorphology regularly advanced with measurements made from point-based instruments, analog cameras, and videotape recorders (e.g., Southard, 2000). Data from experiments once filled shelves with binders and cassettes of data, but currently, new experiments may generate terabytes of high-resolution digital photos, videos, three-dimensional topographic scans, and megahertz to gigahertz data streams. For example, in 1966 Guy et al. published seminal experiments on alluvial bedforms that included 33 pages of data in an appendix completely synthesizing their experimental conditions and outcomes. The entirety of their manuscript including the data tables fit in a PDF document using ~7 Mb of disk space. In comparison, the five experiments conducted at the St. Anthony Falls Laboratory currently available through the National Center for Earth-surface Dynamics (NCED) data repository (<http://repository.nced.umn.edu> accessed 26 May 2014) range in size from 20 to 600 Gb, 4-5 orders of magnitude larger (e.g., Singh et al., 2013). These and similarly rich data sets allow for novel forms of analysis and exploration of new research themes: for example, autogenic processes and statistical analyses that require dense data collection to extract central tendencies from very dynamic systems (e.g., Kim and Jerolmack, 2008; Straub et al., 2008; Jerolmack and Paola, 2010). Such technological advancements in experimental geomorphology present major opportunities for Earth scientists but major challenges for experimental practitioners.

Novel experimental methods provide the opportunity to address fundamental questions in Earth-surface evolution. Among these, three grand challenges appear ripe for solution: i) generalized relation for scaling experimental systems to natural systems, ii) integration of experimental data in training sets and validation tools for models and theories of surface evolution, and iii) explicit connection of surface evolution to generation of subsurface strata with particular concern for separating autogenic processes and allogenic forcings. Progress on these grand challenges has potential for being derived quickly through comparisons of rich experimental data with field and model data.

Realizing the opportunities encapsulated within the grand challenges requires our community to confront a broad suite of technical issues associated with data management. Arguably the most pressing issue is the insufficient accessibility of existing data sets. Most experimental data fall in the category of dark data (Heidorn, 2008). Dark data are not carefully indexed and stored, so data become nearly invisible to scientists and other potential users. Therefore, data are more likely to remain underutilized and eventually lost. Exposure of dark data is hampered by insufficient availability of long-term storage and archival resources and by a lack of guidelines for investigators to document the contents of their data sets. As a result, experimental data sets are rarely used beyond the lifecycle of the project from which they were generated, reducing the efficiency of scientific research at a time when research budgets are increasingly constrained.

Reproducibility and quality control are two additional issues that must be confronted in order to address the grand challenges. Reproducibility is hampered by generally insufficient documentation of repeatable procedures and workflows and by the lack of funding to run replicate experiments. Quality control is in turn inhibited by this lack of reproducibility, as replicate tests could help to constrain the uncertainty associated with experimental methods. Together, poor reproducibility and quality control restrict researchers' ability to perform the intercomparisons between experiment, model, and field data central to addressing the grand challenges described above.

Some of the difficulties described above, such as lack of availability of data storage facilities, are technical in nature. However, many issues in experimental data management are cultural, and their solutions will entail systemic changes at the community level toward open communication and proactive data sharing. Only when the community fundamentally embraces the importance of data sharing and preservation will an investment be made in data storage facilities and training of early career researchers in data sharing and curation. These issues are not insurmountable — experimental geomorphology researchers express a desire to discuss and develop community standards on reporting of experimental

methods, documentation of datasets, and guidelines for data publication and citation (Martin, 2013).

Experimental geomorphologists are not alone in facing the challenge of data sharing. Like geomorphology, polar research is fundamentally interdisciplinary, spanning land, ocean, atmosphere, and ecologic systems sensitive to global climate change (Qiu, 2011). Efforts such as the International Polar Year of 2007-2009 have sought to establish free and open data access to accelerate this cross-disciplinary research progress (*Nature* editorial board, 2009), but the practical realization of data sharing has been limited by technical, financial, and cultural barriers (Carlson, 2011). Responses within polar research to these challenges have included formation of journals for data publication, data storage facilities, and data formatting standards (Carlson, 2011). Collaborative research centers, such as the University Corporation for Atmospheric Research (UCAR) and the University NAVSTAR Consortium (NAVSTAR), have provided centralized support for such efforts like the VIVO project for web data discovery (Hosansky, 2014).

Best practices for data management and sharing consider the entire data lifecycle in light of the specific challenges of experimental geomorphology. Many flumes and laboratory apparatuses for the study of geomorphic processes are unique in their capabilities. With research flumes customized to answer specific questions, standardized setup is neither possible nor desirable. Therefore, if we are to compare multiple experiments at different facilities in a meaningful way, community-scale strategies for data management across the data lifecycle must consider all aspects from planning to publication and reuse. This is vital because synthesis, comparison, and reuse of previously collected data feed directly into the large-scale intercomparisons underpinning solutions to scientific grand challenges.

In this paper, we review data management practices and challenges within the context of the experimental lifecycle; we then outline major challenges to achieving optimal data sharing and reuse. As one way to advance data sharing and reuse, we propose dataset description guidelines that seek to balance the

competing needs of conformity to broadly accepted scientific dataset documentation standards and of the unique characteristics of geomorphic experimental data. Finally, we consider a peer-review process on geomorphic experimental data papers, which has been proposed as a way to encourage and improve data sharing. We propose these metadata guidelines and data publication ideas with the ultimate hope that improved practices in community-scale data management will expedite scientific discovery in experimental geomorphology.

2. The parallel data and experimental lifecycles

The data management lifecycle provides a structure for considering the many operations that are performed on data throughout their conception, generation, storage, and sharing (e.g., Ball, 2012; Michener and Jones, 2012). The data management lifecycle runs parallel to the different phases of the typical experimental geomorphology study (Fig. 1). Although no two sets of experiments are exactly the same, we attempt here to describe common attributes of data management in the experimental geomorphology lifecycle. We present eight distinct phases of the experimental lifecycle, noting that experimental investigations often vary from this generalization.

2.1. Phase 1 – conception

Based on data from field observations, numerical models, or previous experiments, an investigator will typically begin with an idea or hypothesis about a geomorphic problem. For example, noting the commonality of thick vegetation on the banks of natural meandering rivers, the investigator may be struck with the idea of determining how the presence and density of vegetation affects the occurrence and dynamics of meandering river behavior (e.g., Gran and Paola, 2001; Tal and Paola, 2007; Braudrick et al., 2009). Based on this initial idea, the investigator conceives an experimental approach to testing his/her hypothesis, possibly with an accompanying strategy for managing the experimental data to be generated. Experiments necessarily are simplifications of complex natural systems, so the researchers may also document how the intended experiments scale with and relate to natural systems (e.g., how to grow plants

in experiments that mimic natural vegetation cohesion). Approaches here include classic dynamic scaling through traditional engineering dimensionless numbers or use of natural similarity as discussed in Paola et al. (2009).

2.2 Phase 2 - Discovery (Preliminary investigation)

Before carrying out a full-fledged experimental campaign, the investigator may seek out existing datasets or perform preliminary experiments as a proof of concept. Existing datasets may be found in tables and figures in published papers, personal communication, datasets posted on informal or formal websites, web searches, or previous datasets obtained by the investigator. Preliminary experimental data may be obtained to fill gaps in existing datasets, to understand a new experimental apparatus, to find the region of parameter space best suited to address the scientific problem, or to justify a project to funding agencies. The tentative nature of these early investigations often results in practices for recording and storing data and metadata that differ significantly from those later adopted in the main project phase. Nonetheless, analyses from existing or preliminary datasets may ultimately be published. This data discovery phase of projects therefore emphasizes the importance of data reuse. Preliminary investigations often guide the development of data management plans required by many funding agencies to accompany grant proposals. A thorough data management plan details all expected data products, including size, format, and long-term archival plan (Michener and Jones, 2012).

2.3 Phase 3 - Project start-up (Establishment of workflows)

With the inception of a full-fledged experimental campaign, experimental practices (i.e., *workflows*) are planned and established, and these workflows are allocated to personnel (investigators, students, post-docs, and technicians). The experience of personnel and the quality of communication about workflows can strongly influence the process of generating, annotating, and preserving data. Investigators might begin generating workflows by seeking examples in the published literature or from online resources. They might then personally communicate with other investigators to learn about experimental setups,

contact equipment makers to understand measurement devices, or visit other facilities to gain direct knowledge.

2.4 Phase 4 - Data collection and file creation

In carrying out a set of experiments, the bulk of *raw* data—that is, direct instrumental records from cameras, sediment traps, topographic scanners, velocimeters, and other sensors—are generated. Frequency, spatial resolution, and coverage of data collection are decided in this stage largely based on the original research aims. Depending on types of data collected, these records may have direct physical meaning, but additional processing is commonly required before interpretations can be made. The details of data collection and file creation are not necessarily straightforward to infer from descriptions of data in published manuscripts. Procedures and outputs can be modified during experiments in order to improve data collection, reinforcing the need to maintain documentation on data collection practices throughout experimental workflows. Usually, investigators organize data by experimental run and use a naming convention for different experimental parameters and output data products. Initial storage of datasets is most often achieved with local hard disks or other media at the site of investigation.

2.5 Phase 5 - Data processing and analysis

Once raw datasets have been generated, processing steps extract physically meaningful information. Initial processing is commonly automated (i.e., with a script) for uniform application to all data objects of the same type. Automation might incorporate pre-packaged software to convert proprietary data from a commercially manufactured measurement device (e.g., acoustic Doppler velocimeter) to a generic format (e.g., delimited ASCII text). Alternatively, processing might involve investigator-generated codes customized to unique experimental circumstances for converting values lacking intrinsic meaning (e.g., dye intensity or transducer voltage response) to calibrated values with physical meaning (e.g., flow depth or water pressure). In this case, multiple derived data types might be generated from the initial raw data requiring documentation to reproduce each step of the workflow. User-generated scripts may be

generated by the same person(s) who collected the data, or this workload may be transferred to another member of the lab or to a collaborator at another institution. Quality and documentation of scripts will vary widely depending on who created them, and this will affect the potential for reuse of these scripts by others.

2.6 Phase 6 - Preparation for data sharing

While some investigators share data prior to completion of data analysis and publication of results, it is more common for investigators to begin to prepare data for sharing at the time of publication submission (at the behest of the journal or funding agency) or after final publication (normally because of requests from readers of a journal article). Preparations for data sharing typically consist of organizing data sets into coherent folder structures with accompanying explanatory files and metadata descriptions aiming to provide adequate documentation for other investigators. Some investigators submit data to online databases, and preparations for data sharing in those cases also include attempts to conform to standardized guidelines of the particular database.

2.7 Phase 7 - Depositing data

There are many reasons for depositing data into publicly available repositories, including funding agency requirements, journal requirements, and institutional policy. Different options for data repositories include personal servers, university databases, or consortium resources among multiple universities. Investigators often supplement external data deposits with storage on local media, archiving raw and processed data locally but providing only a select subset of processed data publicly. The most important reasons for this are limitations of server storage and a perceived lack of value in raw data products. The current trend is toward depositing data in disciplinary repositories, recognizing the need for disciplinary engagement (e.g., Lynch, 2008). Several data-focused journals stipulate that data papers reference a dataset in an approved repository and have a persistent identifier (e.g., *Geoscience Data Journal*, *Nature Scientific Data*). Many repositories assign Digital Object Identifiers (DOIs) to serve as a persistent,

unique, citable identifier to a dataset with long-term archiving.

2.8 Phase 8 - Discovery and renewing the cycle

Once published or made available to the public, experimental data are free to be discovered by others to stimulate new research directions, provide a basis for comparative projects, or to validate the reproduction of experimental results. Some experimental analyses depend explicitly on meta-analyses of many experimental datasets from multiple facilities. In these cases, investigators compile data from published tables or figures, from web servers, or through direct communication with investigators of existing experiments. In other cases, datasets originally obtained for one reason will later be found to have some new useful application, stimulating the original investigator or others to reexamine the dataset for further analysis. Such data reuse is particularly encouraged in multi-investigator consortiums such as the National Center for Earth-surface Dynamics (NCED), which, through the close proximity and contact of investigators, has shown many examples of reused datasets from laboratory experiments. For example, the eXperimental EarthScope (XES) basin experiment performed in 2002 was originally designed to examine the control of base-level cycles in stratigraphic development of shorelines (Kim et al., 2006). Later, investigators noticed autogenic signatures in surface processes and morphologic patterns within the experiment (Kim and Jerolmack, 2008), outcomes not intended or expected in the initial experimental design.

3. Current needs and challenges

The above review of the lifecycle for geomorphology experiments emphasizes the many steps at which data are produced and used, and the interconnection of projects through data. Our experimental geomorphology community faces many challenges in efficiently and adequately managing, storing, and sharing large volumes of data now being generated. Below, we identify four major challenges to increasing the sharing and reuse of experimental data: i) lack of metadata guidelines, ii) insufficient workflow documentation and communication, iii) inadequate data storage resources, and iv) lack of

incentives and training. These challenges are directly tied to phases of the experimental lifecycle described above.

3.1 Lack of metadata standards for discoverability and sharing

Lack of metadata standards hinder efforts to deposit data (phase 7 above), while a dataset's discoverability strongly affects its subsequent citation, reuse, and impact (phase 8 above). Recent years have seen the development of many new options for improving data discoverability beyond traditional supplements to peer-reviewed journal publications (e.g., Hsu, 2013; Piwowar, 2013), offering new modes of citation, attribution, and sharing. To be discoverable by web crawlers, scrapers, and indexers, data require certain specific standardized information about the data: metadata. Investigators and automated tools can quickly and efficiently make use of data records when proper metadata are present.

When community consensus is reached on metadata standards, databases can distribute data efficiently, as other Earth science disciplines have already demonstrated. Examples of resources in fields aligned with geomorphology include the National Centers for Environmental Prediction (NCEP) reanalysis meteorological data, Incorporated Research Institutions for Seismology (IRIS) seismic data, Integrated Earth Data Applications (IEDA) EarthChem geochemistry data (Goldstein et al., 2014), and environmental sensor data in the Observations Data Model (Horsburgh et al., 2008). Unfortunately, we know of no metadata standards developed for earth surface experimental datasets, nor have catalogs been created to document datasets in one location. Lack of metadata standards and resources inhibits discovery, sharing, and reuse of experimental datasets.

3.2 Insufficient workflow documentation and communication for experimental repeatability

Conception, discovery, startup, data collection, and data analysis (phases 1-5 of the experimental lifecycle described above) often depend on repeating or slightly modifying procedures, data collection methods, and data analysis techniques from recent experiments. However, the value of methods, techniques, and

results developed for specific studies is limited when these are not repeatable. Currently, experimental workflows tend to be shared through word-of-mouth or within research groups in an ad hoc way. Formal documentation of workflows in journal articles and conference presentations is sparse and inconsistent, rarely capturing the specifics of experimental procedures, vendors for measurement devices, or development of new techniques. Most detailed information is captured only in unpublished and unavailable lab books, procurement records, planning documents, and computer codes. Information about certain devices is further limited by proprietary concerns. Inaccessibility of this detailed information severely limits or slows down the development of new projects and workflows, and it also inhibits potential gains for field geologists and modelers who would use experimental products.

In contrast to experimental geomorphology, some other Earth science fields have established workflows for common tasks, especially when achievement of research outcomes depends on large collaborations. For example, the Hydrologic Information System (HIS), established by the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI), has developed tools such as HydroDesktop for search, download, visualization, and analysis of hydrologic datasets (Ames et al., 2012). Establishing such workflows in experimental geomorphology would enhance repeatability by eliminating the need for investigators to reinvent widely used scripting algorithms for analyzing common data types or for sourcing equipment and materials needed to physically model specific geomorphic settings.

3.3 Inadequate data storage resources

Maintaining data on a publicly available storage platform (phase 7 of the experimental lifecycle described above) is increasingly becoming a requirement for funders and publishers (e.g., Holdren, 2013). Funding agencies want the results of supported research to be available to the public (phase 8 of the experimental lifecycle), and publishers want to link to the data underlying scientific interpretations presented in articles. Unfortunately, the volume of experimental geomorphology data from high frequency sensors and imagery is outpacing growth of long-term storage space. Furthermore, large, high resolution, high frame-rate

videos and topographic scans are generally difficult to move or preview. Storage is not free and is required beyond the funding period of a single project. For a single experimental study, data in raw, intermediate, and final format can easily reach several terabytes, requiring a project to purchase additional storage space either in the laboratory or on institutional servers. However, long-term storage does not simply mean buying multiple external hard drives—data must be curated over the long term to maintain public access, provide backups, and update with changing technology. Identifying existing long-term storage resources or investing in new ones would overcome a major obstacle on the road to data sharing and reuse.

3.4 Lack of incentives and training

Even when all of the technical challenges are met, cultural and institutional challenges may still impede the open sharing of data (e.g., Bolukbasi et al., 2013). Creation of metadata, workflow documentation, and placement of data into approved repositories (experimental lifecycle phases 6 and 7) takes time that is rarely allotted by investigators in planning experimental schedules. Recognition of data sharing efforts is not currently part of the promotion and tenure track (e.g., Cutcher-Gershenfeld et al., 2012). Thus, even if an investigator has been convinced of the community benefits of data sharing, significant time spent on these tasks may not be in the self-interest of a researcher. In addition, even if investigators desire to share their data, many feel that they lack the skills to do so effectively (Diekema et al., 2014).

Because experimental technologies have expanded rapidly over the last decade, today's established researchers likely did not face high-volume datasets during their training. Therefore, they may lack the expertise or awareness to promote improved workflow or data management practices among their students. Training materials for data management and curation are more often disseminated through university libraries, funded projects, a combination of the two, or foundations, as opposed to departmental courses (for examples of training materials, see e.g., University of Minnesota, 'Managing your data', <https://www.lib.umn.edu/datamanagement>, accessed 7 December 2014; JISC Managing Research Data,

<http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx>, accessed 7 December 2014; MANTRA Research Data Management Training, <http://datalib.edina.ac.uk/mantra/>, accessed 7 December 2014; ESIP, Data Management for Scientists, <http://commons.esipfed.org/datamanagementshortcourse>, accessed 7 December 2014). Both students and senior scientists may not be aware that these resources exist. Many of these training materials are new, and progress is needed for raising awareness and for affirming the value of the time spent on these activities.

4. Addressing the challenges: Data publication as a framework

Resources exist now to address these challenges, but these are not widely utilized or even known among investigators. We consider the concept of data publication as a framework for simultaneously addressing the challenge areas of metadata profiles, workflow documentation, data storage, and cultural incentives and training. Data publication is a rapidly evolving concept and incipient practice. Newly formed data journals, disciplinary repositories, and commercial ventures are vying for scientific datasets, but the scientific community has yet to fully embrace any particular resource or model, with the definition of data publication itself still being debated (Parsons and Fox, 2013; Kratz and Strasser, 2014).

When publishing and documenting data, a debate about the proper balance between human-readability and machine-readability often ensues. For the purpose of understanding exactly how to reproduce an experiment, practicing researchers tend to prefer a prose description of the metadata necessary to evaluate and reproduce a dataset. Alternatively, for the purpose of building data catalogs and search interfaces, machine-readable, encoded, and parsable text is necessary. The following sections show options on both ends of the spectrum, though the best-case scenario for discoverability and reuse is for both formats of documentation to exist.

4.1 Proposed metadata guidelines for experimental geomorphology

Drawing from existing efforts, we propose below metadata guidelines that balance unique Earth-surface

community needs with existing resources and standards. Efforts for standardized metadata documentation are most common among large organizations or investigator consortia tied together by the mutual need for developing large shareable interdisciplinary datasets, such as HYDRALAB (Frostick et al., 2011), the Critical Zone Observatories (Zaslavsky et al., 2011), and the National Center for Earth-surface Dynamics (Singh et al., 2013). We considered interoperability with existing data models and systems in the Earth-surface sciences, such as Observations Data Model (Horsburgh et al., 2008) and the SEAD Virtual Archive (Plale et al., 2013).

The proposed metadata template contains required basic information and recommended discipline-specific information. The basic information is derived from Dublin Core (Weibel, 1997) and DataCite (Starr and Gastl, 2011) standards for describing data resources, ensuring that the metadata will be compatible with standards among other science disciplines. The discipline-specific information is proposed based on the needs of experimental geomorphology, and it was designed to capture information that is required for reproducing experiments. These metadata are also required to create meaningful geomorphic analyses from raw experimental data. The optional experimental geomorphology information is developed with guidance from the HYDRALAB documentation (Frostick et al., 2011). Unlike the required basic metadata guidelines, which have been designed to follow specific standards for data interoperability, the recommended but optional experimental geomorphology information is more open-ended and possibly subject to change as research interests develop.

Not all of the metadata fields will be applicable to every experiment, but we recommend that every relevant metadata field be included with each published dataset. We also suggest using vocabularies established in the Earth science community (e.g., CSDMS standard names described in Peckham, 2014). Three examples of the proposed profile template are included in the appendix. Examples of how this type of information may be encoded in machine-readable format are available in the DataCite schema documents (Starr et al., 2014).

388

389 **Proposed metadata profile for experimental geomorphology: required basic information**

- 390 1. **Creator:** Investigator(s) involved in producing the data, or authors of the data publication. We
 391 suggest including ORCID (Haak et al., 2012) or another persistent identifier for investigators, and
 392 including the affiliation(s) of the investigator(s).
- 393 2. **Title:** A name or title by which a dataset is known.
- 394 3. **Publication Year:** Year when the data were or will be made publicly available.
- 395 4. **Unique Identifier:** Dataset DOI or other unique identifier, if it exists.
- 396 5. **Contact:** Individual to whom correspondence should be addressed, with contact information.
- 397 6. **Subject:** Subject, keyword, classification code, or key phrase describing the resource.
- 398 7. **Language:** Primary language of the resource.
- 399 8. **Version:** A version number for the resource, including a DOI history if revised.
- 400 9. **Rights:** Any rights information for the resource (e.g., Public Domain (CC0), Creative Commons
 401 Attribution (CC BY), Limited Copyright: for more information, see
 402 <http://wiki.creativecommons.org/FAQ>; accessed 26 December 2014)
- 403 10. **Description:** Additional information that does not fit the other required categories. Can be used for
 404 abstract describing synopsis of scope and creation of dataset.
- 405 11. **Related Resources**
- 406 a. **Link to the data:** URL or information on how to obtain the dataset.
- 407 b. **Associated Product(s):** DOIs, URLs, or descriptions of papers and datasets related to the
 408 resource.

409

410

411 **Proposed metadata profile for experimental geomorphology: recommended discipline-specific**
 412 **information**

413

- 414 12.
- Experimental Campaign:**
- Identifier to relate individual dataset to a larger set of experiments.

13. **Laboratory Information:**

- a. **Lab facility:** Unabbreviated name of the laboratory facility.
- b. **Location:** Complete laboratory address.
- c. **Facility contact:** Name of a persistent contact at the facility (not necessarily the data collector).

14. **Workflow Information:**

- a. **Degree of processing:** 'Raw,' 'Processed,' or 'Derived'.
- b. **Quality control:** Brief description of quality control process.
- c. **Data type:** Parameter(s) measured, see CSDMS standard names (Peckham, 2014).
- d. **Keywords:** Keywords that serve as tags or search terms.

15. **Experimental Apparatus Information:**

- a. **Design:** Description of apparatus design.
- b. **Physical dimensions:** Size of the apparatus used in experiments.

16. **Experiment Information:**

- a. **Purpose:** Original purpose for which experiments were conducted.
- b. **Dates collected:** Date(s) data collected.
- c. **Dates created:** Date digital data file(s) created.
- d. **Sponsors:** Names of entities that contributed resources to experiments, including award number information.
- e. **List:** Table of experimental runs in the dataset with identifier and description.
- f. **Initial conditions:** A description of the physical initial conditions.
- g. **Boundary conditions:** A description of the physical boundary conditions during experimental runs.
- h. **Laboratory effects:** Description of laboratory-specific effects that could have influenced the experimental outcome.

17. **Instrument(s):**

a. Manufacturer and vendor: Name of instrument manufacturer and vendor (if different)

b. Model: Instrument model and vintage.

c. Settings: Settings used during data collection.

d. Procedure: Implementation of instrument for data collection.

18. *Processing:*

a. Procedure: Description of general processing steps.

b. Software: Description or links to software or codes used to process data.

c. Parameters: Software or code parameters used to process data.

19. *Other freeform notes*

4.2 *Improving workflow documentation*

Workflow documentation should provide clear instructions for users to recreate the original experiments.

Although workflows may evolve through the progress of an experimental investigation, it is best to begin the documentation procedure before and during the study rather than after the study is over. The

Sediment Experimentalist Network (SEN) (Hsu et al., 2013) is developing an online platform, the SEN

Wiki (<http://sedexp.net>), for documenting experimental workflows and links to related datasets in a data catalog. In contrast to the strict guidelines for metadata profiles, experimental methods documentation

can be relatively open-ended, reflecting the diversity of experimental types and methodologies. The SEN

Wiki offers a space for users to create individual pages for different experiments, methods, and

instruments, with open space for text, tables, and figures. Such a free structure offers a low barrier to

entry for users encouraging rapid and organic growth of workflow documentation. Additional features

may be added later to the SEN Wiki based on user feedback and needs. An example of a workflow

document for an experimental geomorphology experiment is included in the supplemental materials.

Videos can be a much more descriptive way of describing workflows than written text. The *Journal of*

Visualized Experiments (JOVE) publishes videos and accompanying articles to document experimental

workflows for a wide range of topics, including geoscience. The visual format helps to overcome uncertainty that leads to poor reproducibility and wasted time in application of existing experimental techniques. If a manuscript is accepted, JOVE sends a video team to do a one-day filming of the protocol and produces a 10- to 15- minute narrated video of the procedure. An example is ‘A protocol for conducting rainfall simulation to study soil runoff’ (Kibet et al., 2014). The JOVE articles are indexed like other peer-reviewed articles in relevant indexing sites.

Custom software and scripts for data collection, processing, and analysis are important parts of the experimental workflow for efficient repeatability. Investigator-written scripts are a way to preserve scientific workflows and analyses, if adequately documented for reuse. Several tools exist for managing software documentation, versioning, development, and dissemination; but these are more commonly used by computer scientists and software developers and less known within the geomorphology community. Two solutions that can be implemented are Subversion (SVN) (e.g., <http://subversion.apache.org>; <http://tortoisetsvn.net>) and Git (e.g., <http://www.github.com>). Both of these systems support software versioning and revision of files such as source code (in any language) or documentation. Tutorials on Git for scientists and other scientific software best practices can be found at the Software Carpentry site, <http://software-carpentry.org> (Wilson et al., 2013). Projects like GeoSoft help scientists to add important information like inputs, outputs, dependencies, licenses, and keywords to their code (<http://www.geosoft-earthcube.org/>, accessed 23 December 2014). Another increasingly popular tool gaining popularity among scientists is the iPython Notebook, where code, outputs, and text are included together in an easily accessible web-based format (Shen, 2014).

4.3 Suggestions for publicly accessible data storage

Data storage in repositories can be provided by discipline-specific, institutional, or commercial systems. Discipline-specific repositories are considered best for data discovery by one’s peers, quality control of discipline-specific data and metadata, and responsiveness to changing community needs. Institutional

repositories, hosted through many universities, provide data archiving services for members of the institution. Economies of scale make centralized data servers more feasible for large university systems than for small colleges or individual departments. Commercial cloud storage systems may provide the most cost-effective option for large volumes of data, especially for smaller communities lacking the ability to build their own systems. For example, Amazon Glacier provides low cost for backup and data archiving, but it becomes more expensive for frequent data access.

A major impediment to long-term data storage and curation is the lack of a clear framework for cost allocation. In principle, an investigator could include data management costs into the budget of a scientific proposal. However, it is unclear how one would fund long-term data management beyond the typical 3-5 year timeframe of most project grants. This may be less of an issue for investigators who are members of major universities or research consortia with their own data storage systems. Such institutions could cover data server and other long-term costs through overhead or general budgets much as they support facilities or library costs.

Berman and Cerf (2013) suggested four coordinated approaches to the problem of who will pay for public access to research data: i) facilitate private-sector stewardship of public access to research data, ii) use public-sector investment to jumpstart sustainable stewardship solutions in other sectors, iii) create and clarify public-sector stewardship commitments for public access to research data, and iv) encourage research culture change to take advantage of what works in the private sector. Ultimately, a best-choice solution requires a funding model that should be determined with input from researchers, funding agencies, and private sector computing providers.

An immediate solution to publicly accessible data storage could involve adapting existing funding mechanisms for research infrastructure. For instance, agencies support high-cost scientific equipment through a variety of regular proposal request processes. In order to facilitate this mode of funding, we

must change attitudes about dark data. Specifically, the community must agree that it is unacceptable that valuable data are not curated and are lost to further scientific inquiry. Other science domains have made community access to data a priority, e.g., IRIS, NGDC (National Geophysical Data Center), NCDC (National Climatic Data Center). The availability of data from these communities allows scientists and public alike to test Earth science hypotheses without the need to spend dollars to reacquire data with every new proposal. If any of these data were collected and subsequently lost to the community, it would be as if they were never generated. In this way, the physical infrastructure to curate and share existing data is as valuable as the original instruments used to collect the data. We suggest this as one possible framework for acquiring data infrastructure to serve our community's needs.

4.4 Creating incentives and training opportunities

The most effective way to motivate investigators to document and provide access to their datasets is to give them fair and public credit for their work. This is becoming possible through initiatives like data journals, data citation indices (e.g., Web of Science Data Citation Index), and alternative metrics for nontraditional research products (i.e., www.altmetric.com). Another strategy is to train early career scientists on methods and protocols for data management, thereby building good habits and planting seeds for a culture of proactive data sharing and stewardship (e.g., Cutcher-Gershenfeld et al., 2012). Activities of the Sediment Experimentalist Network Education and Data Standards (SEN-ED) subinitiative include workshops, town hall forums, lectures at early career workshops, and other activities to promote this training. The Sediment Experimentalist Network will also leverage the efforts of its association with EarthCube, a geoscience-wide project focused on building digital infrastructure for managing, sharing, and exploring geoscience data with a strong emphasis on community engagement (Richard et al., 2014).

4.5 Considerations for Data Publication and Review in Experimental Geomorphology

Data publication presents one avenue for formalized sharing of datasets. Successful data publication depends on addressing the four major challenge areas of metadata consistency: workflow documentation,

data storage, and cultural willingness. Data papers are published by journals such as *Nature Scientific Data*, *Water Resources Research*, and *Geoscience Data Journal*. These journals make scientific data available, discoverable, citable, and reusable so credit is given to the original data generators each time their data are included in analyses for a typical publication. Data papers are not evaluated on novel scientific findings associated with data sets; rather, the main criteria for evaluation are the quality, reuse potential, and completeness of a data set. Most data journals require that data sets are deposited and publically available in one or more community-recognized, web-based repositories to be considered for publication.

Trustworthiness of published datasets must be assured, and one way to do this is through peer review (Parsons et al., 2010). Peer review of data is distinct from traditional peer review of papers. Two types of peer review of data have been identified—the first type of review checks if the data has adequate documentation, while the second type evaluates the integrity of data for answering the originally intended scientific question (Callaghan et al., 2012). In experimental geomorphology, an example of the first type of quality control would be a check for inclusion of units and instrument calibrations with datasets, while an example of the second type would be a check on the proper scaling of the experiment to the environment or process of interest. Mayernick et al. (2014) noted that data peer review practices and guidelines typically emphasize completeness of the dataset, the level of detail of the description, and the usefulness of the data.

If correctly executed, the review process for data publications could provide a check that investigators are adequately documenting metadata and workflows, providing trustworthiness for data reuse in the community. Citations of data publications could encourage a culture of data sharing by providing recognition for efforts of investigators taking the time to share and document and publish their datasets and workflows. However, data publications alone will not address underlying structural needs for metadata standards, data management training, and data storage resources.

Below, we document some important aspects to be considered in Earth-surface data publications:

repeatability, quality control, uniqueness, and scaling.

Repeatability: A data paper should deliver complete descriptions of workflows and datasets so that experiments can be repeated in other facilities. An issue here is that different facilities often contain custom-designed flumes, specific sediment mixtures, or unique measurement devices. For example, flume wall materials (e.g., glass or steel) might modestly but importantly alter surface roughness and therefore stress partitioning in an experiment. Data papers should address all such concerns to be most useful and successful.

Quality control: It is almost impossible to find two completely identical experiments conducted in two separate facilities. At the very least, data papers should provide confidence that data sets are internally self-consistent. For example, independent measures (e.g., repeat topographic scans, weigh scales) could provide quantification of uncertainty in sediment discharge estimates. Such quality control metrics would inspire confidence in potential data reusers, especially modelers utilizing experimental datasets for parameterization and validation purposes. Currently, there are few if any guidelines for quality control for experimental geomorphology datasets, and a major effort should be made to decide on community-endorsed quality control metrics for evaluation of data papers.

Uniqueness: Most experiments are designed and conducted for examining one or a few research problems. While in-depth explanations of research findings and interpretations are best left to traditional publications, data papers could still include brief statements of research questions to provide context for users to understand, interpret, and reuse datasets. Data authors should further indicate how their data are distinguished from other data sets, thereby demonstrating the potential use of datasets beyond the original research question(s).

Scaling: One major challenge in experimental geomorphology is to provide proper scaling. The data paper should provide enough description (and possibly accompanying measurements) to demonstrate traditional dynamic scaling or to demonstrate natural similarity to real-world geomorphic systems. Guidelines for how experimental data could be extrapolated to field scale should be provided in the data description.

5. Summary and conclusions

Rapid growth in the size and complexity of experimental geomorphology datasets has accompanied the proliferation of high-frequency and high-resolution measurement technologies. Massive experimental datasets provide opportunities to address grand challenges associated with model-field-laboratory intercomparisons and autogenic/allogenic fluctuations in linked surface and stratigraphic processes. However, issues in discoverability, reproducibility, quality control, accessibility, and storage of datasets currently hinder direct attacks on grand challenge questions.

We have summarized the typical experimental lifecycle as it pertains to generation and use of experimental Earth-surface data. Exploration of this data lifecycle reveals four critical issues for data management: (i) lack of established metadata profiles; (ii) insufficient workflow documentation; (iii) inadequate data storage resources; and (iv) absence of data sharing incentives and training. We have proposed solutions to these challenges (summarized in Table 1), many of which are currently being addressed by the Sediment Experimentalist Network initiative. In particular, we provide examples of a dataset description template (allowing machine-readability) and a workflow document (emphasizing human-readability). In addition, we have identified the rise of the data publication as a possible vehicle for achieving many of our goals, though additional outreach and infrastructure efforts must accompany growth of data publications to be effective. Sharing and reuse of data are integral to making progress on grand challenges that can be met most efficiently with coordinated experimental effort, modeling, and

field work.

Acknowledgements

The authors would like to thank Charles Nguyen for conversations that informed this paper. We also appreciate constructive comments from two anonymous reviewers of the original manuscript. Figure 1 was created by G. Ahn. This work was funded by EarthCube Domain Workshop NSF 1252324, and NSF EarthCube RCN Award 1324760.

References

- Ames, D.P., Horsburgh, J.S., Cao, Y., Kadlec, J., Whiteaker, T., Valentine, D., 2012. HydroDesktop: Web services-based software for hydrologic data discovery, download, visualization, and analysis, *Environmental Modelling & Software* 37: 146-156, doi: 10.1016/j.envsoft.2012.03.013.
- Ball, A. 2012. Review of Data Management Lifecycle Models. University of Bath, Bath, UK. <http://opus.bath.ac.uk/28587/1/redm1rep120110ab10.pdf>.
- Berman, F., Cerf, V., 2013. Who will pay for public access to research data? *Science* 341(6146): 616-617. <http://dx.doi.org/10.1126/science.1241625>.
- Bolukbasi, B., et al., 2013. Open Data: Crediting a culture of cooperation, *Science* 342(6162): 1041-1042. <http://dx.doi.org/10.1126/science.342.6162.1041-b>.
- Braudrick, C., Dietrich, W.E., Leverich, G.T., Sklar, L.S., 2009. Experimental evidence for the conditions necessary to maintain meandering in coarse-bedded rivers. *Proceedings of the National Academy of Science* 106, 16936-16941, doi: 10.1073/pnas.0909417106.
- Callaghan, S., Donnegan, S., Pepler, S., Thorley, M., Cunningham, N., Kirsch, P., Ault, L., Bell, P., Bowie, R., Leadbetter, A., Lowry, R., Moncoiffe, G., Harrison, K., Smith-Haddon, B., Westherby, A., Wright, D., 2012. Making data a first class scientific output: data citation and publication by NERC's Environmental Data Centres, *The International Journal of Digital Curation* 7(1), 107-113, doi: 10.2218/ijdc.v7i1.218.
- Carlson, David. 2011. A Lesson in Sharing. *Nature* 469 (7330), 293–293. doi:10.1038/469293a.
- Cutcher-Gershenfeld, J., Diggs, S., Gil, Y., Hazen, B., Kinkade, D., 2012. Early Career Strategic Visioning End User Workshop: Workshop Results and Executive Summary, <http://workspace.earthcube.org/sites/default/files/files/document->

[repository/EarlyCareerStrategicVisioningEndUserWorkshop.pdf](#), accessed 6/1/2014.

Diekema, A.R., Wesolek, A., Walters, C.D., 2014. The NSF/NIH effect: surveying the effect of data management requirements on faculty, sponsored programs, and institutional repositories, *The Journal of Academic Librarianship*, <http://dx.doi.org/10.1016/j.acalib.2014.04.010>.

Frostick, L.E., McLelland, S.J., Mercer, T.G., 2011. Users guide to Physical Modelling and Experimentation: Experience of the HYDRALAB Network. IAHR Design Manual, CRC Press.
Wells, S., Sutherland, J. & Millard, K. (2009). Data management tools for HYDRALAB – a review. HYDRALAB Report NA3-09-02, available from <http://www.hydralab.eu/publications.asp> (page accessed 08/01/10).

Goldstein, Steven L.; Hofmann, Albrecht W.; Lehnert, Kerstin A., 2014. Requirements for the Publication of Geochemical Data. EarthChem Library. <http://dx.doi.org/10.1594/IEDA/100426>.

Gran, K., Paola, C., 2001. Riparian vegetation controls on braided stream dynamics, *Water Resour. Res.* 37(12), 3275-3283, doi: 10.1029/2000WR000203.

Guy, H. P., Simons, D. B., Richardson, E. V., 1966. Sediment transport in alluvial channels: summary of alluvial channel data from flume experiments, 1951-61. US Geological Survey Professional Paper 422-I, 104 pp.

Haak, L.L., Fenner, M., Paglione, L., Pentz, E., Ratner, H., 2012. ORCID: a system to uniquely identify researchers. *Learned Publishing* 25(4), 259-264, doi: 10.1087/20120404.

Heidorn, P.B., 2008. Shedding light on the dark data in the long tail of science. *Library Trends* 57(2), 280-299, doi: 10.1353/lib.0.0036.

Holdren, J.P., 2013. Public access memorandum from the Office of Science and Technology Policy, http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf, accessed 6/1/2014.

Hook, L.A., Santhana Vannan, S.K., Beaty, T.W., Cook, R.B., Wilson, B.E, 2010. Best practices for preparing environmental data sets to share and archive, Oak ridge National Laboratory Distributed Active Archive Center, doi:10.3334/ORNLDAAAC/BestPractices-2010.

Horsburgh, J.S., Tarboton, D.G., Maidment, D.R., Zaslavsky, I., 2008. A relational model for environmental and water resources data. *Water Resources Research* 44(5), W05406, doi:10.1029/2007WR006392.

Hosansky, David. 2014. Making Research Data More Traceable. UCAR Atmos News, September 3. <http://www2.ucar.edu/atmosnews/in-brief/12278/making-research-data-more-traceable>, accessed 12/1/2014.

- Hsu, L. 2013. Four Things You Can Do To Extend the Impact, Reach, and Longevity of Your Research Data. *Bulletin of the Ecological Society of America* 94, 389–393. <http://dx.doi.org/10.1890/0012-9623-94.4.389>.
- Hsu, L., McElroy, B., Martin, R., Kim, W., 2013. Building a Sediment Experimentalist Network (SEN): sharing best practices for experimental methods and data management: *Sedimentary Record* 11(4), 9-12.
- Jerolmack, D. J., & Paola, C., 2010. Shredding of environmental signals by sediment transport. *Geophysical Research Letters* 37(19), doi: 10.1029/2010GL044638.
- Jones, K., 2011. Research360: Managing data across the institutional research lifecycle. Poster presented the 7th International Digital Curation Conference, Bristol, UK, 5–8 Dec. http://www.dcc.ac.uk/webfm_send/589, accessed 6/1/2014.
- Kibet, L. C., Saporito, L. S., Allen, A. L., May, E. B., Kleinman, P. J. A., Hashem, F. M., Bryant, R.B., 2014. A Protocol for Conducting Rainfall Simulation to Study Soil Runoff. *J. Vis. Exp.* 86, e51664, doi:10.3791/51664.
- Kim, W., Jerolmack, D. J., 2008. The pulse of calm fan deltas. *The Journal of Geology* 116(4), 315-330, doi: 10.1086/588830.
- Kim, W., Paola, C., Voller, V. R., Swenson, J. B., 2006. Experimental Measurement of the Relative Importance of Controls on Shoreline Migration: *Journal of Sedimentary Research* 76(2), 270-283, doi: 10.2110/jsr.2006.019.
- Kratz, J., Strasser, C., 2014. Data publication consensus and controversies [v3; ref status: indexed, <http://f1000r.es/4ja>] *F1000Research* 2014 3:94 (doi: 10.12688/f1000research.3979.3).
- Lynch, C. 2008. Big data: How do your data grow? *Nature* 455, 28-29. doi:10.1038/455028a.
- Martin, R., 2013. Town Hall Report, Surface Process Experiments – A Community Discussion, <http://tinyurl.com/sedex-townhall2012>, accessed 6/1/2014.
- Mayernick, M.S., Callaghan, S., Leigh, R., Tedds, J., Worley, S., 2014. Peer Review of Datasets: When, Why, and How, *Bulletin of the American Meteorological Society*, e-View, <http://dx.doi.org/10.1175/BAMS-D-13-00083.1>.
- Michener, W.K., Jones, M.B., 2012. Ecoinformatics: supporting ecology as a data-intensive science, *Trends in Ecology & Evolution* 27(2), 84-93, doi: 10.1016/j.tree.2011.11.016.
- Nature editorial board. 2009. The Way Ahead for Polar Science. *Nature* 457 (7233), 1057–1057. doi:10.1038/4571057a.
- Paola, C., Straub, K., Mohrig, D., Reinhardt, L., 2009. The "unreasonable effectiveness" of stratigraphic

- and geomorphic experiments: *Earth-Science Reviews* 97(1-4), p. 1-43, doi: 10.1016/j.earscirev.2009.05.003.
- Parsons, M.A., Fox, P.A., 2013. Is data publication the right metaphor? *Data Science Journal* 12, WDS32-WDS46, <http://dx.doi.org/10.2481/dsj.WDS-042>.
- Parsons, M.A., Duerr, R., Minster, J.-B., 2010. Data Citation and Peer Review, *Eos Transactions, American Geophysical Union* 91(34), p. 297-298. doi:10.1029/2010EO340001.
- Peckham, S.D., 2014. The CSDMS Standard Names: Cross-domain naming conventions for describing process models, data sets and their associated variables, *International Environmental Modelling and Software Society (IEMSs)*, In: Ames, D.P., Quinn, N.W.T., Rizzoli, A.E. (Eds.), 7th Intl. congress on Env. Modeling and Software, San Diego, California, USA, <http://www.iemss.org/society/index.php/iemss-2014-proceedings>.
- Piowar, H., 2013. Altmetrics: Value all research products, *Nature* 493, 159. doi: 10.1038/493159a.
- Plale, B., McDonald, R.H., Chandrasekar, K., Kouper, I., Konkiel, S., Hedstrom, M., Myers, J., Kumar, P., 2013. SEAD Virtual Archive: Building a federation of institutional repositories for long-term data preservation in sustainability science, *The International Journal of Digital Curation* 8(2), 172-180. doi:10.2218/ijdc.v8i2.281.
- Qiu, Jane. 2011. US Science Academy Report Calls for ‘Bipolar’ Research. *Nature News* 469 (7329), 145–145. doi:10.1038/469145a.
- Richard, S.M., Pearthree, G., Aufdenkampe, A.K., Cutcher-Gershenfeld, J., Daniels, M., Gomez, B., Kinkade, D., Percivall, G., 2014. Community-developed geoscience cyberinfrastructure, *EOS Transactions, American Geophysical Union* 95(20), 165-166, doi: 10.1002/2014EO200001.
- Shen, Helen. 2014. Interactive Notebooks: Sharing the Code. *Nature* 515 (7525), 151–52. doi:10.1038/515151a.
- Singh, A., Czuba, J.A., Foufoula-Georgiou, E., Marr, J.D.G., Hill, C., Johnson, S., Ellis, C., Mullin, J., Orr, C.H., Wilcock, P.R., Hondzo, M., Paola, C., 2013. StreamLab Collaboratory: Experiments, data sets, and research synthesis, *Water Resour. Res.* 49, 1746–1752, doi:10.1002/wrcr.20142.
- Southard, J. B., 2000. An introduction to fluid motions and sediment transport. *Eastern Section SEPM short course*.
- Starr, J., Gastl, A., 2011. isCitedBy: A Metadata Scheme for DataCite, *D-Lib Magazine* 17(1/2) doi:10.1045/january2011-starr.
- Starr, J., Members of the Metadata Working Group, 2014. DataCite Metadata Schema for the Publication and Citation of Research Data, DataCite, doi: 10.5438/0010.

- 794 Straub, K. M., Mohrig, D., McElroy, B., Buttles, J., Pirmez, C., 2008. Interactions between turbidity
795 currents and topography in aggrading sinuous submarine channels: A laboratory study. *Geological*
796 *Society of America Bulletin* 120(3-4), 368-385, doi: 10.1130/B25983.1.
- 797 Tal, M., Paola, C., 2007. Dynamic single-thread channels maintained by the interaction of flow and
798 vegetation, *Geology* 35(4), 347-350, doi: 10.1130/G23260A.1.
- 799
- 800 Weibel, S., 1997. The Dublin Core: A Simple Content Description Model for Electronic Resources.
801 *Bulletin of the American Society for Information Science and Technology* 24(1): 9-11, doi:
802 10.1002/bult.70.
- 803
- 804 Wilson, G., Aruliah, D.A., Brown, T., Chue Hong, N.P., Davis, M., Guy, R.T., Haddock, S.H.D., Huff,
805 K.D., Mitchell, I.M., Plumbley, M.D., Waugh, B., White, E.P., Wilson, P., 2013. Best Practices for
806 Scientific Computing, <http://arxiv.org/pdf/1210.0530v4.pdf>
- 807
- 808 Zaslavsky, I., Whitenack, T., Williams, M., Tarboton, D.G., Schreuders, K., Aufdenkampe, A., 2011. The
809 initial design of data sharing infrastructure for the Critical Zone Observatory, *Proceedings of the*
810 *Environmental Information Management Conference*, Santa Barbara, CA, 28-29 September, EIM'2011,
811 pp. 145-150.
- 812
- 813
- 814

Tables and Figures

Figure 1. (1.5 column width)

Caption: The data management lifecycle (inner circle, following Jones, 2011) is paralleled by the phases in the experimental geomorphology lifecycle (outer circle).

Table 1**Challenges and resources for data sharing in experimental geomorphology**

Challenge	Resources	Reference or example
Metadata guidelines	Suggested metadata profile Hydralab guidelines CSDMS standard names	This study Frostick et al., 2011 Peckham, 2014
Workflow documentation	SEN wiki Jove SVN/GIT	sedexp.net www.jove.com www.github.com
Storage of large datasets	Community repositories e.g., SEAD Institutional repositories Amazon Glacier	sead-data.net aws.amazon.com/glacier
Community incentives and training	Data Publications New metrics of production Early career training	www.altmetric.com

Appendix A. Dataset description for Eurotank Community Experiment 2014

A. Basic information

1. **Creator:** Kimberly Litwin Miller (University of Texas at Austin, <http://orcid.org/0000-0002-4299-9006>) and SEN Workshop Participants
2. **Title:** Community Experiment: Eurotank 2014
3. **Publication year:** 2014
4. **Unique identifier:** n/a
5. **Contact:** Kimberly Litwin Miller, litwinmiller@jsg.utexas.edu
6. **Subject:** Experimental Geomorphology; Earth Surface Processes
7. **Language:** English
8. **Version:** 1
9. **Rights:** CC0
10. **Description:** The experiment formed two parallel and simultaneous deltas with the same input parameters, but different substrates: one with a static substrate and another with a deformable substrate.
11. **Related resources:**
 - a. **Link to the data:** goo.gl/cM1tfE
 - b. **Workflow document:** included as a supplement to this article (Hsu et al., 2015)

B. Discipline-specific information

12. **Experimental campaign:** SEN Community Experiments
13. **Laboratory information**
 - a. **Lab facility:** Eurotank Flume Facility
 - b. **Location:** Eurotank laboratorium, Utrecht University, Princetonlaan 6, 3584 CB Utrecht, The Netherlands
 - c. **Facility contact:** Joris Eggenhuisen, J.T.Eggenhuisen@uu.nl
14. **Data information**
 - a. **Degree of processing:** raw and processed
 - b. **Quality control:** the processed data were reviewed by lab members
 - c. **Data types:** photographs, surface_topography, stratigraphy, shoreline_position
 - d. **Keywords:** delta, sediment flux, shoreline, stratigraphy, mobile substrate
15. **Experimental apparatus information**
 - a. **Design:** The Eurotank Flume Facility is designed for sedimentary experiments with an adjustable basin floor. The Eurotank 2014 Community Experiment occupied only a fraction of the entire basin.
 - b. **Physical dimensions:** entire flume: 10.5 x 6.5 x 1.1 m
16. **Experiment information**
 - a. **Purpose:** Compare simultaneous delta experiments with the same input parameters, but different substrates: one with a static substrate and another with a deformable substrate.
 - b. **Date of collection:** 2014-11-03/2014-11-06
 - c. **Sponsors:** NSF EAR 1324760
17. **Other notes**

- a. Further information is available in the SEN Eurotank Community Experiment 2014 Workflow Document (Supplemental Material) including: Experiment list, initial conditions, boundary conditions, laboratory effects, instruments, and processing

Appendix B. Dataset description STEP basin Community Experiment 2012

A. Basic information

1. **Creator:** Wonsuck Kim (University of Texas at Austin, <http://orcid.org/0000-0002-4709-971X>)
2. **Title:** Community Experiment: Experimental Stratigraphy 2012
3. **Publication year:** 2012
4. **Unique identifier:** n/a
5. **Contact:** Wonsuck Kim, delta@jsg.utexas.edu
6. **Subject:** Experimental Geomorphology; Earth Surface Processes
7. **Language:** English
8. **Version:** 1
9. **Rights:** CC0
10. **Description:** STEP basin experiment with two different wavelengths of sediment discharge fluctuations
11. **Related resources**
 - a. **Link to data:** <http://goo.gl/ubH50S>
 - b. **Related publication:** <http://goo.gl/y6sGD8>, Hsu, L., McElroy, B., Martin, R., and Kim, W., 2013, Building a Sediment Experimentalist Network (SEN): sharing best practices for experimental methods and data management: The Sedimentary Record, v. 11, no. 4, p. 9 – 12.

B. Discipline-specific information

12. **Experimental campaign:** SEN Community Experiments
13. **Laboratory information**
 - a. **Lab facility:** Sediment Transport and Earth-surface Processes (STEP) Basin, Morphodynamics and Quantitative Stratigraphy Lab, UT PRC Bldg
 - b. **Location:** 120JJ Pickle Research Campus, University of Texas, Austin, 10100 Burnet Rd, Bldg 120, Austin, TX 78758
 - c. **Facility contact:** Wonsuck Kim, delta@jsg.utexas.edu
14. **Data information**
 - a. **Degree of processing:** raw and processed
 - b. **Quality control:** the processed data were reviewed by lab members
 - c. **Data types** photographs, surface_topography, stratigraphy, shoreline_position
 - d. **Keywords:** delta, sediment flux, shoreline, stratigraphy
15. **Experimental apparatus information**

- 917 a. **Design:** Sediment Transport and Earth-surface Processes (STEP) Basin, see
918 <http://www.ig.utexas.edu/people/staff/delta/research/experiments.html>
919 b. **Physical dimensions:** entire flume 4 x 5 x 1.5 m
920 16. **Experiment information**
921 a. **Purpose:** Observe shoreline development and resulting stratigraphy of varying sediment
922 and uplift rates
923 b. **Date of collection:** 2012-12-10/2012-12-11
924 c. **Sponsors:** NSF 1252324, EarthCube End-User Workshop: Experimental Stratigraphy
925 d. **List:** Available in <http://goo.gl/ubH50S> in subfolder Raw Data > 20121210 STEP
926 Community Experiment 2012 > Control Programs CE2012 > ExpDesign Model
927 CE2012.xlsx
928
929
930
931
932
933
934
935
936

Appendix C. Dataset description Debris Flow Erosion Experiments

A. Basic information

1. **Creator:** Leslie Hsu (University of California, Berkeley, <http://orcid.org/0000-0002-5353-807X>), William E. Dietrich (University of California, Berkeley), Leonard Sklar (University of California, Berkeley, <http://orcid.org/0000-0001-9626-733X>)
2. **Title:** Debris flow erosion experiments
3. **Publication year:** 2014
4. **Unique identifier:** n/a
5. **Contact:** Leslie Hsu, hsu.leslie@gmail.com
6. **Subject:** Experimental Geomorphology; Earth Surface Processes
7. **Language:** English
8. **Version:** 1
9. **Rights:** CC0
10. **Description:** Experiments in a 4-meter diameter, 80-cm wide vertically rotating flume to study bedrock erosion by debris flows. Force plate, height, bedrock topography, video, and image data.
11. **Related resources**
 - a. **Link to data:** <http://dx.doi.org/10.6084/m9.figshare.988297>; <http://dx.doi.org/10.6084/m9.figshare.978686>; <http://dx.doi.org/10.5967/M0KH0K9H>
 - b. **Related publication:** <http://dx.doi.org/10.1002/2013JF003078>, Hsu, L., Dietrich, W.E., Sklar, L.S. (2014) Mean and fluctuating basal forces generated by granular flows: laboratory observations in a large vertically rotating drum, Journal of Geophysical Research.

B. Discipline-specific information

12. **Experimental campaign:** n/a
13. **Laboratory information**
 - a. **Lab facility:** Richmond Field Station, University of California, Berkeley
 - b. **Location:** 301 S 46th St #478, Richmond, CA 94804
 - c. **Facility contact:** William E. Dietrich, bill@eps.berkeley.edu
14. **Data information**
 - a. **Degree of processing:** raw and processed
 - b. **Quality control:** The data were reviewed by the investigator and portions of the dataset are in a peer reviewed publication.
 - c. **Data types:** basal_force_normal, flow_height_longitudinal, video, photo, bedrock_strength_tensile, channel_base_topography
 - d. **Keywords:** debris flow, granular flow, bedrock erosion
15. **Experimental apparatus information**
 - a. **Design:** Designed by Engineering Laboratory Design, MN, completed in 2005
 - b. **Physical dimensions:** 4 meter diameter, 80 centimeter wide vertically-rotating drum flume

981 **16. Experiment information**

- 982 a. **Purpose:** Observe particle dynamics and bedrock erosion in a suite of granular flows,
983 including water and mud saturated grain flows.
984 b. **Dates collected:** 2005/2008
985 c. **Sponsors:** EAR-0120914
986 d. **Experiment list:** Available in the fileset <http://dx.doi.org/10.6084/m9.figshare.988297>

987 **17. Instruments**

- 988 a. Acuity AR4000 Laser Scanner; Interface Force Model SWP10-5K-b000 Precision Force
989 Transducer
990