

## Facultair Datamanagementbeleid BMS

### Achtergrond

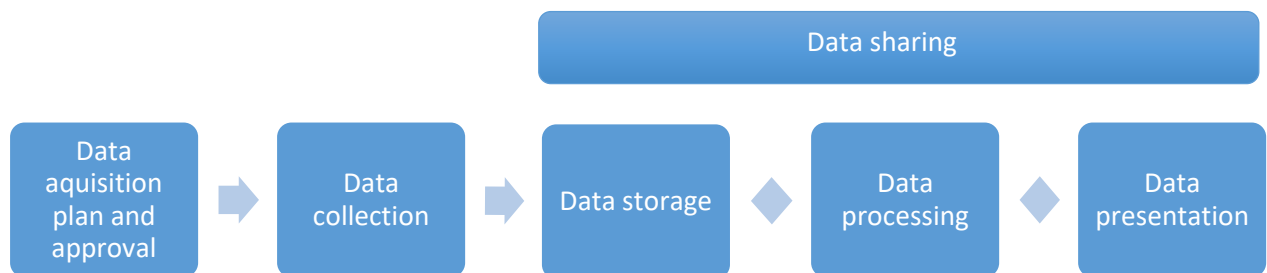
In juli 2015 is het UT onderzoeksdatabeleid vastgesteld. In dit document worden minimale criteria aangegeven met betrekking tot datamanagement en -opslag. Aangezien de omgang met onderzoeksdata per academische discipline uiteenloopt, is het aan elke faculteit om, in aanvulling op het UT beleid, een eigen facultair databeleid op te stellen. Hierin is vastgelegd hoe wetenschappers binnen de faculteit dienen om te gaan met onderzoeksdata.

### Proces

Voor de faculteit BMS heeft de decaan hiertoe begin 2016 een commissie in het leven geroepen. Deze commissie, bestaande uit betrokkenen vanuit faculteit, onderzoeksinstituut IGS en LISA, is gevraagd de huidige initiatieven en voorzieningen (infrastructuur) op het gebied van BMS/IGS databeleid in kaart te brengen (1). Daarnaast is de commissie gevraagd, waar nodig, adviezen uit te brengen ten aanzien van de acties die nodig zijn om de komende jaren optimaal gebruik te kunnen maken van de beschikbare onderzoeksdata en hoe een juiste opslag, beheer en veiligheid gegarandeerd kunnen worden (2).

De commissie heeft zich, op basis van bovenstaande, ten doel gesteld om te komen tot de ontwikkeling van een integraal databeleid, met als doel (nieuwe) medewerkers duidelijk te maken wat, van begin tot het einde, te nemen stappen en randvoorwaarden zijn, in de omgang met onderzoeksdata. Naast de ontwikkeling van een beleid, vraagt dit ook om een plan van aanpak om dit bij de medewerkers "tussen de oren te krijgen" (4).

Een weergave van het datamanagement proces is te vinden in figuur 1.



Figuur 1. Het data management-proces

Daarbij dient opgemerkt te worden dat soms een andere volgorde van stappen wordt gevolgd of dat stappen worden overgeslagen (bij bijv. extern verkregen bestanden). In het algemeen zullen echter de bovenstaande processtappen worden doorlopen.

De afgelopen periode is door de faculteit vooral geïnvesteerd in tools voor datacollectie (Tech4People lab), de aanschaf van servers voor dataopslag en verwerking en personele ondersteuning hiervan. Ook is een begin gemaakt met de software voor data processing en datavisualisatie. Aandachtspunt (ook in dit document) ligt bij data-infrastructuur en een overkoepelend beleid.

### 1. *Uitgangspunten van het UT Databeleid*

Het databeleid dat wordt opgesteld door de afzonderlijke faculteiten moet, volgend uit het centrale UT beleid, tenminste bestaan uit de volgende onderdelen:

#### **Bepalingen aangaande de verantwoordelijkheid voor het databeleid van de groep/faculteit/instituut**

##### *Voorgestelde BMS invulling:*

Binnen de faculteit BMS is de decaan eindverantwoordelijk voor het **vaststellen** en (laten) **implementeren** van **databeleid**, de **communicatie** daarover naar de vakgroepen en medewerkers en het organiseren van accurate **basisvoorzieningen**. De decaan moet zorgdragen voor goede communicatie en delegatie van deze taken richting betrokken vakgroepvoorzitters. Ook is de decaan verantwoordelijk voor het aanbieden van de juiste basisvoorzieningen binnen de faculteit om uitvoering aan het databeleid te kunnen geven. Van betrokken vakgroepvoorzitters, PI's, promotoren, en begeleiders wordt een actieve en bewuste, verantwoordelijke houding gevraagd. Zij hebben een rol bij het verder communiceren en onder de aandacht brengen van de juiste uitvoering van het databeleid richting de individuele onderzoeker. Ook is het belangrijk dat zij zich bewust zijn van hun voorbeeldfunctie en dat zij (vermoedens van) onethisch of incorrect handelen melden bij het bureau Faculteitsdecaan.

De eindverantwoordelijkheid voor de **juiste omgang** met data ligt uiteindelijk altijd bij de individuele onderzoeker/1e auteur van een project. Een onderzoeker dient zich voor aanvang van een project te vergewissen van de juiste expertise en capaciteit om het onderzoek in lijn met de geldende richtlijnen en de gedragscode van de VSNU te kunnen uitvoeren. Dit is één van de onderdelen van een professionele houding in de wetenschap. Daarnaast geldt voor junior onderzoekers en (PhD) studenten dat de 1e begeleider van het project (met PhD) een **medeverantwoordelijkheid** heeft voor correct datamanagement, inclusief een controle van de datapackages. De faculteit monitort op afstand en voert af en toe datapackage audits uit.

**Bepalingen aangaande de duurzame opslag van data, in overeenstemming met wettelijke regelingen, contractuele voorschriften van derde partijen, financiers, etc.;**

*Voorgestelde BMS invulling:*

- Ieder onderzoek binnen de faculteit (van zowel studenten als medewerkers), waarbij sprake is van (in)directe betrokkenheid van proefpersonen, moet worden voorgelegd aan en goedgekeurd door de ethische commissie.
- Direct na verkrijging van de data wordt deze - indien nodig - geanonimiseerd (ontkoppelt van persoonsinformatie). Dit geldt voor zowel zelf verzamelde als extern verkregen data. Geanonimiseerd betekent dat de data niet terug te herleiden is naar een individu. Daarbij moeten de koppelbestanden (die een link leggen tussen persoon en hun data, die onderzoekers gebruiken) versleuteld worden en offline bewaard.
- De omgeving waarin de data en de sleutel/koppelbestanden bewaard worden moet goed beveiligd zijn.

**Bepalingen aangaande de opslag van de data voor zowel korte als lange termijn en de manier waarop de data omschreven worden;**

**Criteria ten aanzien van de selectie van onderzoeksdata die duurzaam worden opgeslagen; dit kan per onderzoek verschillend zijn. Algemene criteria hierbij zijn: reproduceerbaarheid, verifieerbaarheid en herbruikbaarheid**

*Voorgestelde BMS invulling:*

- Opslag van data mag nooit alleen op persoonlijke, en/of lokale media plaatsvinden.
- Bij ontvangst van de (ruwe) data, wordt deze in ieder geval opgeslagen op de centrale en beveiligde BMS server. De file(s) worden bewaard in een map waarin ook de toestemming van de ethische commissie is opgeslagen, zodat altijd helder is om welk onderzoeksproject het gaat.
- Na afloop van het onderzoek wordt de data opgeslagen in een trusted repository (Dans) of blijvend op één van de beveiligde servers van de faculteit. Dit betreft ten minste de ruwe dataset.
- De richtlijn voor het bewaren van de ruwe onderzoeksdata is minimaal 10 jaar<sup>1</sup>.
- De data wordt zo gearchiveerd dat ze – ook op lange termijn – toegankelijk blijven. Dit vergt goede metadata via bijvoorbeeld een *persistent identifier* (doi), zodat de data niet verloren gaat en makkelijk terug te vinden is. Daarnaast betreft het documentatie om de data begrijpelijk te houden, maar ook voorzieningen voor het converteren van bestanden naar een nieuw formaat, voordat het oude formaat verouderd en mogelijk onleesbaar wordt.
- Voor archivering kan de data eventueel extern opgeslagen worden, wanneer deze niet meer actief gebruikt hoeft te worden; er zijn archieven of repositories beschikbaar die een duurzaamheidsmissie hebben met de, in het vorige punt genoemde, voorziening van converteren (gecertificeerd als zogenoemd *trusted digital repository*).

---

<sup>1</sup> Voor zover dit niet schuurt met het Wbp art. 10 verbod om data niet langer te bewaren dan noodzakelijk is.

**Voorwaarden waaronder onderzoeksdata beschikbaar worden gesteld aan derden.**

*Voorgestelde BMS invulling:*

- In het algemeen geldt hierbij: open als het kan, gesloten als het moet. Randvoorwaarden zijn wettelijke eisen (anonimiseren) of contractuele eisen, van bijvoorbeeld financiers of opdrachtgevers.
- Als de data gedeeld wordt, dan is daar voorafgaande toestemming voor nodig van deelnemers middels de informed consent procedure.
- De data zullen waar mogelijk gedeeld worden voor wetenschappelijke doeleinden (wetenschappelijke partners).
- Delen van data voor valorisatiedoelstellingen met eindgebruikers/praktijkpartners is ook mogelijk (maatschappelijk nut). Bij voorkeur wordt die data niet overgedragen aan deze partners en wordt er een voorziening getroffen om deze partijen binnen de UT de data te laten inzien/her-analyseren.
- Onderzoekers dienen uiterst terughoudend te zijn met het delen van data voor commerciële doeleinden.

**Aanvullende kaders voor een juiste omgang met data zijn:**

- De [gedragscodes VSNU](#). Uitgangspunt is dat alle universiteiten en de daar werkzame wetenschapsbeoefenaren al het nodige doen om zich de inhoud van deze code eigen te maken. Zij dragen er bovendien zorg voor dat de gedragscode wordt bediscussieerd in de academische gemeenschap. Hierdoor wordt bewustwording gestimuleerd van wat wordt verstaan onder goed wetenschappelijk onderwijs en onderzoek. De code verplicht de wetenschapsbeoefenaren niet alleen tot naleving, maar impliceert ook een verantwoordelijkheid in het bevorderen van de naleving ervan in de academische omgeving en misstanden te signaleren en rapporteren. Bij de bestuursorganen van de universiteit rust een verplichting om de naleving van de code te bevorderen en te handhaven. De universiteit heeft een openbaar en dwingend reglement dat de onafhankelijke behandeling van klachten over schendingen van de wetenschappelijke integriteit reguleert, inclusief een commissie [wetenschappelijke integriteit](#).
- Het advies zoals geformuleerd door de DSW<sup>2</sup>-commissie wetenschappelijke integriteit, dataopslag en replicerbaarheid (d.d. 06/01/2016). Dit advies is gebaseerd op een inventarisatie van dataprotocolen van de sociale faculteiten van 9 Nederlandse universiteiten. Over een aantal uitgangspunten bestaat consensus. De DSW data commissie adviseert deze te hanteren als minimale standaarden. Het is belangrijk om daar als BMS ook bij aan te sluiten vanwege een eenduidige standaard binnen de discipline, maar ook omdat het samenwerking tussen de onderzoekers tussen de verschillende universiteiten en de overgang van een onderzoeker van de ene naar de andere universiteit vergemakkelijkt. Deze minimale standaarden zijn verwerkt in dit advies.

---

<sup>2</sup> Discipline-overleg Sociale Wetenschappen waarin de decanen van faculteiten sociale wetenschappen zitting hebben.

- Daarnaast is sinds 1 januari 2016 de [Meldplicht Datalekken](#) van kracht geworden en zijn medewerkers van de Universiteit Twente verplicht melding te maken van elk (vermoedelijk) datalek. Deze meldplicht is het gevolg van een wijziging in de Wet bescherming persoonsgegevens (zie Wbp/GDPR<sup>3</sup>).
- Per 1 oktober 2016 worden alle BMS onderzoekers en studenten verplicht voor aanvang van al het mensgebonden onderzoek een aanvraag in te dienen bij de Commissie Ethiek van de faculteit.
- Tot slot is er in 2015 UT breed [informatiebeveiligingsbeleid](#) vastgesteld. De 4 basisregels zijn:

**1. De UT houdt zich, als publiekrechtelijke organisatie, aan de wet.** Volgens velen is dit een open deur. Ook als ondernemende universiteit gaat de UT niet mee in de redenering dat de keuze om je als organisatie aan de wet te houden een kosten-baten analyse hoort te zijn. Anderzijds is de universiteit natuurlijk ook geen politieagent.

**2. Informatie over studenten en medewerkers wordt zo zorgvuldig als mogelijk behandeld.** Aankomende en huidige studenten moeten er op kunnen vertrouwen dat er zo zorgvuldig als mogelijk met hun informatie wordt omgegaan. Deze informatie is voor een groot deel studie-gerelateerd. Zorgvuldig met privacy omgaan is één van de uitdagingen waar we als universiteit voor staan.

**3. Aandacht besteden aan informatiebeveiliging binnen alle processen en activiteiten hoort bij de proactieve houding van de UT-medewerker.** Informatiebeveiliging heeft veel aspecten en raakt aan bijna alle processen en activiteiten. Risico nemen hoort bij de ondernemende houding van de UT. Onderdeel hiervan is vooraf de mogelijke gevolgen te onderzoeken en maatregelen te nemen die onaanvaardbare risico's beperken.

**4. Het ondernemende en creatieve karakter van de UT wordt niet gefrustreerd door het informatiebeveiligingsbeleid.** Noodzakelijke beveiligingsmaatregelen moeten natuurlijk genomen worden, ook als individuen dit minder waarderen, maar dan wel na een afweging. Proportionaliteit is hierbij gewenst. Ingrijpende of beperkende maatregelen die niet in verhouding staan tot het feitelijk verminderen van risico's worden niet genomen.

Daarnaast wordt in het document onderscheid gemaakt naar drie typen data: standaard, gevoelig en kritiek.

## *2. Bestaande infrastructuur op gebied van onderzoeksdata (verzameling, opslag en ondersteuning)*

Welke zaken worden momenteel reeds georganiseerd op het gebied van onderzoeksdata (facultair, UT-breed en landelijk)? Welke ondersteuning wordt daarbij geboden?

### **Binnen BMS/IGS bestaan verschillende voorzieningen:**

- **IGS Datalab:** data-infrastructuur waarbij onderzoekers geholpen worden om data te creëren, op te slaan en te onderhouden volgens de voorwaarden die aan een professionele en verantwoordelijke omgang met data worden gesteld. Daarnaast kan het Datalab onderzoekers voorzien van methodologische adviezen ten

---

<sup>3</sup> Wbp = Wet bescherming persoonsgegevens (thans vigerend), GDPR = General Data Protection Regulation (vigerend vanaf 25 mei 2018)

aanzien van het gebruik van data. Vanaf de opstart van het Datalab, is voor het online opstellen van surveys gebruik gemaakt van een open source tool (Lime Survey). Vanuit IGS werd voorzien in ondersteuning bij het administratieve beheersgedeelte en methodologische advisering. In de loop van 2016 bleek het technisch beheer niet meer op de juiste wijze te kunnen worden ondersteund en was de software niet meer downward compatible, waardoor problemen met updates ontstonden. Daarnaast was er ook vanuit LISA een beveiligingsprobleem gesignaleerd met betrekking tot de externe opslag van de data en mogelijk zelfs van toepassing zijnde US wetgeving. Op dit moment worden op basis van een lijst met wensen en vereisten twee alternatieven verkend: een CBS tool en een applicatie van een nieuwe lokale startup: Easyonsurvey. Bijkomend voordeel van de 2e optie is dat het via hun software mogelijk lijkt om via een aantal stappen data te anonimiseren en dat er veel toepassingen zijn voor aanvullingen in lijn met Tech4People projecten (apps etc). Uiteindelijk zullen deze 2 nieuwe mogelijkheden vergeleken worden met de tools die momenteel gebruikt worden, omdat deze aanbieders - in lijn met datamanagement ontwikkelingen - ook hun *policies* en opties lijken aan te passen.

- **Tech4Peoplelab:** dit lab biedt mogelijkheden tot dataverzameling met behulp van een breed scala aan nieuwe technologische middelen. Binnen het Tech4Peoplelab hebben de afgelopen periode verschillende ontwikkelingen plaatsgevonden: zo is er een fileservers ingericht (met algemene informatie, handleidingen, ondersteunende informatie en projectmappen). Daarnaast is er een programmatuurgedeelte op de fileservers ingericht waarin alle programma's opgenomen zijn die nodig zijn voor het gebruik van de nieuwe tech tools voor onderzoek en dit onderzoek te kunnen analyseren. Naast de fileservers is er ook een FTP-server ingericht die op een (be)veilig(d)e wijze file transfers mogelijk maken wanneer er projecten met derde partijen worden uitgevoerd (zoals bijvoorbeeld met het MST). Komend najaar worden onder alle onderzoekers binnen de faculteit BMS beveiligde (Samsung TD) SSD sticks verstrekt en worden onderzoekers verzocht geen gebruik meer te maken van losse en niet beveiligde sticks. Hiermee willen we zorgen voor grotere bewustwording en toenemend gebruik van veilige opslag en vervoer van (UT) onderzoeksdata.

Een uitgebreid overzicht van wat Tech4People te bieden heeft is te vinden op de website <https://bmslab.utwente.nl/>. Ook aan de opslag van de data is gedacht. Hiervoor is een tweetal nieuwe servers aangeschaft, die beheerd worden vanuit LISA (voormalig ICT en B&A).

- **CBS Remote access terminal:** geeft onderzoekers toegang tot CBS-microdata, geanonimiseerde data op persoons- of bedrijfsniveau.

**Op de UT bestaan de volgende voorzieningen/diensten:**

- **LISA** (UT-breed) biedt onderzoekers hulp bij het duurzaam archiveren en opslaan van onderzoeksdata. LISA ondersteunt daarbij in de gang richting de landelijke universitaire voorzieningen die geboden worden door het 4TU Centre for Research Data en DANS.

- **Health Datalab** (initiatief vanuit MIRA/GZW), ook binnen CTIT soortgelijk initiatief.

**In landelijk universitair verband bestaan de volgende voorzieningen:**

- **4TU Datacenter** (lange-termijn archief voor wetenschappelijke onderzoeksdata met permanente toegang tot onderzoeksdata, en middelen voor het hergebruik hiervan. 4TU Datacenter geeft onderzoekers advies en ondersteuning met betrekking tot databeheer. (DOI toegevoegd, voldoet aan eisen van instellingen en onderzoeksfinanciers).
- **DANS** (instituut dat datasets duurzaam archiveert en waar reeds vergaarde data kunnen worden hergebruikt. Daarnaast adviseert DANS over datamanagement en certificering van digitale archieven).
- **Dataverse.nl (DVN)** (samenwerkingsverband tussen DANS, 4TU Datacenter, Nederlandse universiteiten en hogescholen, gaat om korte termijn archiveren en delen van niet-statische onderzoeksdata met projectteams of onderzoeksafdelingen) biedt onderzoekers de mogelijkheid om data op te slaan, te delen en online te registreren, tijdens hun onderzoek en voor een periode van tien jaar daaropvolgend. Het wordt gezamenlijk aangeboden door deelnemende instellingen).
- **SURF** (ICT samenwerkingsorganisatie van het onderwijs en onderzoek in Nederland) biedt onderzoekers dataopslag, de mogelijkheid om data te delen en ondersteuning bij het omgaan met data.

## Appendix<sup>5</sup> Data Management Plan - Guidelines, Policies, Paragraphs

There are many different Data Management Plan guidelines and templates. Exact data management requirements and conditions differ per funding agency, journal or institution. An overview of the most important guidelines is given below.

### Funder requirements<sup>6</sup>:

#### [DANS](#)

Extensive guidelines for writing a [Data Management Plan for scientific research](#), including a DMP-checklist, are provided by DANS. DANS (Data Archiving and Networked Services) is an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands Organisation for Scientific Research (NWO).

#### [4TU.Datacentrum](#)

The 4TU.Datacentrum focusses on technical studies and offers a template to help setup a Data Management Plan. The template can be found [here](#). Please also check the University of Twente template which can be opened or downloaded [here](#) at the right side of the screen.

#### [Horizon 2020](#)

Horizon 2020 is a European Commission programme that finances and stimulates European research and innovation projects. Applicants and recipients of subsidies can find the Data Management Policy in the document [Guidelines on Data Management in Horizon 2020](#). A handy tool to write a Data Management Plan that meets the Horizon 2020 policy, you can use the tool [DMP Online](#).

#### [ZonMW](#)

ZonMW, the Netherlands Organisation for Health Research and Development, promotes health research and innovation. ZonMW stimulates re-use of existing datasets and accessibility of new data sets. Checklists for writing a Data Management Plan or a Data Paragraph (in Dutch) can be found in the report [“Data: Digitale Diamanten” \(2013\)](#).

#### [KNAW](#)

The policy of KNAW, the Royal Netherlands Academy of Arts and Sciences, focuses on [open access and digital preservation of research data](#). Research proposals should include a “data paragraph” containing information on data management.

#### [NWO](#)

NWO, the Netherlands Organisation for Scientific Research, funds research programmes and manages the national knowledge infrastructure. NWO stimulates permanent archiving and re-use of research data. To receive funding, researchers are required to sign a data contract with DANS. More information on NWO regulation on granting can be found [here](#) and information on the datamanagement protocol NWO [here](#).

### Publisher requirements<sup>7</sup>:

More and more publishers have guidelines on research data management. The Radboud University provides an extensive overview on the different requirements on their website on Research Information Services. In addition to the overview of publisher requirements – adopted below – they have published an Excel sheet with an extensive list of journal requirements. This Excel file can be found [here](#).

#### [Elsevier](#)

Elsevier supports the principle that research data are made freely available to all researchers, but does not oblige authors to make these data available. Authors are, however, strongly encouraged to deposit their datasets in relevant data repositories or to make them available through other channels. If a suitable data repository cannot be

---

<sup>5</sup> met dank aan Sytske Wiegersma voor het opstellen van dit overzicht.

<sup>6</sup> Adopted from <https://www.utwente.nl/igs/datalab/datamanagement/guidelinesdmp/>

<sup>7</sup> Adopted from <http://www.ru.nl/research-information-services/institutional-policy/research-managers/>



found, Elsevier enables its authors to store additional information relevant to the article as a supplementary file. Authors retain their copyright with respect to this additional information.

#### [PLOS](#)

All data and relevant metadata have to be deposited in a relevant and publicly accessible data repository, unless the data already form part of the submitted article. The *data availability* statement should specify the names of the repositories, DOI's and/or access numbers for the relevant data sets. Smaller data sets may be uploaded as supporting-information files. These should be submitted in a format that allows the data to be extracted easily. For instance, where data in tables are concerned, a spreadsheet is preferable to a PDF file.

Whenever ethical or legal considerations prevent the author from handing over the data to a repository or submitting them together with an article, he or she may indicate that the data will be made available upon request to all interested researchers. Particular considerations in a given case, such as possible implications for patent proceedings or possible future research, are not looked upon as valid exemptions to the rule. PLOS requires that authors meet the standards specific to their discipline when preparing and registering their data. Repositories should preferably adhere to accepted standards, such as the criteria formulated by the *Centre for Research Libraries* or the *Data Seal of Approval*. Licenses should be no more restrictive than CC-BY. Information obtained through research on human trial subjects should be handled in such a way that the privacy of the individuals is fully protected.

#### [Nature Publishing Group](#)

When publishing in Nature journals, authors are upon request obliged to put all their material, data, code, and all related protocols immediately at the disposal of their readers. Possible restrictions on availability must be reported to the editors and indicated in the article itself.

As of May 2013, papers submitted in any field within life sciences must be accompanied by detailed information regarding the experimental and analytical design of the research. This is done by filling out a checklist. Supporting information must be available to editors and peer reviewers. Data should be deposited in a publicly accessible repository. *Scientific Data*, an open access sister journal to *Nature*, maintains a [list](#) of approved and recommended repositories, sorted by discipline.

A less desirable alternative would be to make data sets available as supplementary-information files. These become freely accessible upon publication through [www.nature.com](http://www.nature.com). If it appears technically impossible to provide a dataset, authors must provide a URL or another unique identifier.

Authors are encouraged to publish a data descriptor in *Scientific Data*. Data on cell lines should be deposited at a repository that has a certificate of authenticity. Consult [www.nature.com](http://www.nature.com) for links to additional sources to check the identification of cell lines. Computer code should be made available upon request to editors and reviewers. It is recommended that experimental protocols are shared through [Protocol Exchange](#). Separate guidelines exist for clinical trials.

#### [Science](#)

All of the information necessary to understand, evaluate and elaborate upon the conclusions presented in the manuscript should be available to the readers of *Science*. Datasets must be deposited in approved repositories and accession numbers must be mentioned in the article. Adherence to the MIBBI guidelines (*Minimum Information for Biological and Biomedical Investigations*) is encouraged. The same holds true for all computer codes that have been used to create or analyse the data.

Upon publication, all reasonable requests to provide information and materials must be met, and any exception or limitation must be reported to the editors. Fossils or other rare specimens have to be deposited at a public museum or repository and remain available for inspection.

Ellen Giebels 26-10-2016

Large datasets for which no approved repository can be found may be deposited as supplementary materials at *Science* or, if that is not possible, on an institutional website provided a copy of the data is deposited at *Science* in order to guarantee their availability to readers.