

CEIC Meeting (Ann Arbor)

Notes by JHD

18/19 March 2017

1 Preliminaries

JHD Thanked AMS/MR in general and Ed Dunne (ED) in particular for hospitality. Thanks *in absentia* were also due to the University of Michigan's Mathematics Department, and in particular Doreen Fussman and Molly Long, in helping the local organiser ensure accommodation for the meeting participants.

ED Glad to welcome you. CEIC is important for the Mathematical Community.

Overlay Journals JHD asked whether MathSciNet would cover these. Tim Gowers (*Discrete Analysis*) had asked MR: ED had replied need

- ISSN (not actually obligatory, but no ISSN reduces utility)
- Unique Persistent Identifiers (preferably DOIs)
- ED consulted copyright lawyers on the sort of licence required (to allow sending to reviewers etc.) and *Discrete Analysis* is in line with this.
- Note that MR's mandate is to cover the refereed literature in mathematical sciences. Note that a journal can be part-covered¹: ED's standard example is PNAS.
- It would be good to explain the above.

JHD/ED

Apologies VS, RV (delayed), AO (electronically), MS (electronically).

Minutes Various corrections and shortenings were made. The following were identified as matters arising: IMUNet; Accessibility; Copyright/Hodges paper update; ICM 2018 Panels; ICSU/WDS liaison; DOIs and website.

* ID noted that GT had corrected her summary of the Skype conversation with him at the 2016 CEIC meeting on the state of electronic backups and archives for the IMU Secretariat. She apologises for her unintentional misrepresentation.

¹The MR technical term is "pre-scanned".

2 Accessibility

MS joined the meeting: accessibility (largely for the visually impaired) was discussed. MS had circulated a paper for CEIC 2016. PI noted that there should be a survey. MathJax is working on developing work out of Google. This is currently not yet a deployed product. TB noted that MathJax is very slow (depending on both network and rendering capability). *ID asked what alternative senses (in particular, high-resolution touch pads that use the sense of touch to read) were being used, but the answer was general ignorance. There is also the question of input.* MS noted that “convoluted” L^AT_EX may well not be solution, and MathML was better in that respect. the “legibility” to software of existing PDFs was very variable. TB reported that his interlocutor worked best with L^AT_EX, failing that “having it read to him”. There is always the accuracy challenge. PI reported that a 15-years-ago AMS panel of VI mathematicians was deeply divided. MS reported that the situation had largely changed.

Call for information, via IMUNet and an early newspaper article at the ICM, asking for it to be passed on. MS provided a very helpful starter document. This should be passed by OC, Neil Soiffer and Volker Sorge. Google Docs ‘anyone can comment’ would be one way of collecting input, but ID pointed out that it’s often blocked/discouraged. Hence a document on the IMU website, and feedback by a [@mathunion.org](mailto:mathunion.org) would be a minimum. Of course, any mechanism should itself be accessible.

GT

Apparently Duxbury will now consume the L^AT_EX produced by infty reader. The description of “simple L^AT_EX” seems to be operational only, though. How accessible is the Mathematics in Google Docs? MS suggested asking Volker Sorge.

JHD/VS

3 Website, DOIs etc.

JHD is currently CEIC’s liaison with the website project. Currently he has heard nothing more.

ID reported that EC had accepted that CEIC was the home of IMU’s expertise in this area, and CEIC should have a member involved. Things are proceeding slowly. **JHD** should check with HH.

JHD

ICM Proceedings A recent ICM is generally published by a commercial publisher (2014: Kyung Moon SA; 2010 via World Scientific; 2006 EMS). It appears that 2010 has article-level DOIs, 2014 and 2006 appear not to have. Rules for uniqueness of DOIs means that IMU could not assign DOIs to objects that already have them, e.g. 2010 articles.

CrossRef TC reported a CrossRef update: more content types; was “need ISSN/ISBN”, now applies to standards and “posted content”. membership

fees are \$275 for organisations whose “total revenue from publishing” is less than \$1M, and \$0.25 per DOI². DataCite have also been used in this area.

IMU would need to keep a record (database) of DOIs versus URLs, which should probably be persistent URLs. Note that, if an object is available in multiple forms (PDF and DjVu for ICM proceedings for example), it is normal for the DOI to point to a “landing page”: example quoted was <http://dx.doi.org/10.1108/LHT-11-2016-0126>, which resolves to <http://www.emeraldinsight.com/doi/abs/10.1108/LHT-11-2016-0126>. Various reasons are size (don’t have to wait for a large irrelevant piece to download) and copyright reasons (you want to advertise their existence).

4 GDML

4.1 Discussion with SMW

JHD noted that last year’s CEIC had passed a resolution, welcoming the progress to date and looking for further progress.

Outreach happened as proposed.

Telecons 30 since last year. Website is still the stub, and IMU GDML pages unchanged.

Charter Approved, but note that this is a “guiding document” not a legal one.

IMKT was proposed to the Alfred P. Sloan Foundation, and was funded in mid-December 2016, after many questions from Sloan. SMW noted that various “letters of support” were added to the proposal at the end, from a wide range: software (NAG, MapleSoft), bibliography (EuDML, zbMath), institutes (Fields, Waterloo). Canada (Federal Law) was chosen as the legal home, and the charitable status. Not mandatory, but recommended that we have 25% Canadian residents on the Board. Also a Scientific Advisory Board. It is “members” who elect the board of directors, and the question is who they are. Must have voting members, and can have non-voting ones. The Directors lay out the criteria, and a renewal procedure is recommended. Voting members will be Board, SAB, and others invited. Non-voting will be individuals, and possibly societies. Lawyer is drafting by-laws currently. Then populate with a skeleton crew, get charitable status, then look at subsidiaries elsewhere for local tax reasons. Sloan wants a suitable board (experience, diversity etc.). Current aim is for a “legally valid” Board to be in place by end of April 2017. Credible Scientific Board, but with vacancies, by end June 2017. Reports to Sloan should also serve as reports to IMU’s EC.

²This was the figure TC reported. The structure is actually more complicated: see <https://www.crossref.org/fees/>.

arXiv CEIC noted (https://arxiv.org/help/stats/2016_by_area/index) that Mathematics, as arXiv defines it, is the largest single category (26.4%) within arXiv, and most arXiv subjects are “mathematics heavy”.

Structure Needs staff just to administer, and therefore an annual income of around \$300K. Therefore need due-paying members, not just charitable grants.

Projects The following are ongoing.

- SFC (special functions concordance). Several people have approved this project. Mathematica (+Bruce Miller + Edgardo from Maple-Soft + others) are involved. Concordance of special functions: note software house support.
- F[ormal]-abstracts³. Envisaged by Tom Hales. Discussed with SMW. TH: various formal systems have been used to prove various results, e.g. Feit-Thomson, Kepler conjecture, and many applications in safety-critical software. A problem is that many papers have errors: what do we then mean by semantic capture.
- Harmonisation in the formal area. Willingness expressed in February 2016. To be pursued in Big Proof at INI in July 2017.
- *n*gram work (bibliographic work on abstracts etc., to involve zb-Math). Still hasn't really started.

ID EC meets April 1/2: what can I tell them?

Planned Outreach Big Proof (INI) and CICM 2017, also JMM 2018 (as in 2016).

Others Corpus collection.

4.2 Discussion with arXiv

Oya Rieger from Cornell joined us by Skype. arXiv 25 years old, 15 years at Cornell, and 10 years of OR's involvement. Maths is 26% of arXiv (and far from saturated, CEIC thinks). Downloads growing as well,

Showed public sustainability wiki. <http://confluence.cornell.edu/display/culpublic/arXiv+Sustainability+Initiative>. Usage figures showed no usage figures from China. arXiv believes its about 1/3, but cannot track it. There's a special interface for China, an arXiv user group etc. 7% of users in user survey quoted China as main place of work, though. “arXiv is a distributed scientific activity”. There's a member advisory board (UK and DE have strong consortium members) and a scientific advisory board.

Life is changing: open access policies, predatory journals, changing user patterns (OR quoted CS). 150 moderators keeping track of 500 submissions/day.

³The idea is to provide a collection of sample formalizations of asserted results, not go for full proofs which would. be much bigger

Biggest requirement was better search facilities and author name disambiguation (surprisingly high awareness of ORCID). Worry about link rot if only links rather than content are added.

4 FTE oversee the moderation process. Several automated scripts filter. Moderators check existence of abstract, bibliography, legibility of graphs etc, sanity of classification. Users like overlap, correct classification, query transparency of moderations. Divided views:

- think boldly and further advance open access;
- emphasis on the importance of sticking to the main mission.

But “don’t mess it up” was frequent, majority view was that it shouldn’t be a social media platform. Adding a rating system: views split; annotations, also split.

Q–TC Was the caution against features a caution against them at all, or against having them in the core?

A People knew it was a shoestring budget. Also people want arXiv to be a neutral space. People are concerned about quality of comments.

Code is 25 years old in places. We’d like to use a community-based modular open source solution with APIs. We would like to make it possible for partners to develop overlay services.

Q–JHD Overlay journals. Working with MR already. Work with you as well?

A Certainly a good idea. Many other disciplines are interested in this. These come and go: there can be a great individual without necessarily a strong governance model. We should be moving towards a layered world. arXiv-NG is meant to encourage an open architecture: APIs etc.

Example 1 (Public access mandates) *Europe, but also US. tax-payer dollars should foster openness of data etc. Our current metadata system is very rigid, so can’t support funder mandates etc.*

OR is also on Project Euclid.

Q–PI You said “predatory journals”.

A It’s an issue, especially in IS/CS. Publishers trying to deposit on behalf of authors.

Q–PI Overlap analysis software: where’s the problem.

A The software is Ginsparg’s. We implemented it before policies really went in. Strong support expressed for this. Progress on policies is slow since SD left.

Q–TC Semantic understanding? What kind of relationships would arXiv want with people doing this work?

A Our current search is very old-fashioned. First thing to look at in a new architecture. Another question is arXiv as Open Data. Amazon cloud will support download of all papers. Some people in Cornell’s IS are in advanced search projects. People are interested in ‘publications as data’: who is reading etc.

Q–TC But as this becomes a two-way street, do we need a more formal relationship.

A We need to continue arXiv as a low-budget model. Therefore we will focus on core services + collaborations. We will need guidelines for collaborations. One request a month for collaboration, and arXiv is not a sandbox. Experiments that get turned off when the grant ends are very frustrating.

Q–ID IMU is a midwife for IMKT. Explains history.

A Great idea, but we really need a Scientific Director first. Probably shouldn’t talk to Greg Kuperberg until this is sorted.

Q–PI MathSciDoc?

A We have seen many things come and go.

4.3 Discussion

Postprocess Examples like Kohlase, who has limited permission to do his research.

Overlay CEIC/MR/arXiv to work on a “cookbook for overlay”. **JHD**

Conclusion arXiv does *not* see itself as a competitor to IMKT, and should probably be technically able, and probably politically capable, to collaborate with IMKT.

Q What do we see as suitable collaborations: F-abstracts is one candidate?

SMW joined Sunday morning.

Semantics GDML would like to be more semantic. So one option is a program reading the PDF, another is more semantic macros. JHD’s example is “ $G_i \leq G$ when $i \leq n$ ”, where the first \leq should be `\subgroup` rather than `\le`. PI: corresponds to DLMF’s enhanced `\LATEX`: currently only used by cognoscenti.

Could there be a “semantically accessible” label in arXiv: probably not if it takes any effort by arXiv. If there’s an `amssemmath` macro, could we install it.

TB What about a mirror of the Cornell interface, such that GDML could experiment with this?

RV The Stacks project But it will only happen when people take this up — “every new PhD student . . .”.

TB arXiv say that they require the \LaTeX if that was used, and that they add `hyperref`, so they were instrumental in pushing this.

All Do we volunteer to be a protective shield for arXiv — experimental requests should be directed to IMKT?

Mathscidoc PI drew attention to this: <http://archive.ymsc.tsinghua.edu.cn/>. They have *Acta Mathematica* and *Arkiv för Matematik*. ID felt that it would be politically wise to approach Mathscidoc as well as arXiv.

4.4 Resolution

CEIC received the report of the GDML group. CEIC had been encouraged to contact arXiv by the Simons Foundation, and had had a presentation from, and discussion with arXiv. CEIC were supportive of the progress made by the WG, would like progress to continue, and would like to see a timetable for the establishment and population of the Boards of IMKT. The WG was encouraged to prepare a plan for cooperation with arXiv. It was suggested that similar contacts should be explored with Mathscidoc.

CEIC believes that the WG should be responsible for arranging publicity at ICM 2018, on the lines of JMM 2016.

5 Archives

We received an excellent report from the archivist.

E-mail transfer CEIC noted that a document “what I have to do and what happens with my emails to make them archived?” was in discussion with HH. CEIC would like to be involved, both from technical aspects and because ID had recently done this.

6 Website etc.

Twitter CEIC had noted in 2015 that the Twitter feed from ICM2014 (i.e. #ICM2014) was not being archived along with the rest of ICM 2014’s information. They can still be seen retrospectively via <https://twitter.com/ICM2014>. Should this be in some way archived before it disappears? TC has taken a copy from the “Twitter archive”. Should something more be done for 2018? TC noted that the live (< 7 days) Twitter provides

more information.

JHD/ICM2018

Update Wordpress blog server, analytics server, mailing systems unchanged.
28/3/2017 β version of Phase 1, then 3/4/2017 to go live for Phase 1
(main IMU site). Afterwards there'd be information for commissions etc.
on building their subsites. Phase 2 would be May/June.

DOIs Weierstrass Institute is a DOI provider. GT to provide details.

7 Actions

1. Update (Appendix?) to Best Practices for Journals to cover overlay journals. JHD to check with Ed Dunne (MR) Oya Rieger (arXiv) and zbMath (Olaf Teschke as starting point). TC asked the question about DOIs: suppose the same item was in two overlay journals, would it get two DOIs? [Thierry noted <http://www.numdam.org/search/C%C3%A9cile%20Huneau-a>.] TC noted that *Nature* is running self-overlay journals. There are requirements on copyright for Crossref DOIs: <https://www.crossref.org/member-obligations/>. **JHD+**
2. Recommendations on Copyright. To be restarted by TB, TC and Henry Cohen. It would be good to get involvement either from Project Euclid (Dave Ruddy?), or from Duke (ID to find). Draft also to be run past MS. The new version to cross-refer to other recommendations, both IMU recommendations, and others elsewhere that have been written since 2001. **TC+**
3. Accessibility. See section 2.
4. Website
 - (a) Preserving old URLs — JHD to check the β website, and make sure the URLs in appendix ?? (and others from other commissions) are preserved.
 - (b) DOIs — GT to find out Weierstrass contact, TC to follow up. If IMU can't use WMI URLs, then CEIC needs to go back to EC with the old proposal to make IMU a DOI agency. **GT/TC**
5. GDML: see section 4.4. **PI**
6. JHD to liaise with Christiane Rousseau to talk. **JHD/CR**

8 Closing

The next meeting in Spring 2018 should be in Berlin. Exact date to be arranged.

JHD/SM

JHD thanked ED and PDFI for the local arrangements.

9 Onshuus

AO joined after RV/TB/ID left. JHD asked AO to test the new IMU website when it goes into β .

AO

AO raised the following points.

- arXiv is extremely valuable for existing researchers in developing countries.
- For people moving into research, e.g. new research students, there is a shortage of good advanced textbooks: will GDML help in this area? PI to take this point back to GDML.

PI

Wikibooks was also mentioned⁴ as a possible place that ought to be encouraged or a location for the relevant results like those posted by AIMath. TB subsequently commented:

Maybe you're not aware of this resource, which is however interesting (both for what it serves, and the way it is organized—there has been a Notices article on this some time ago) <https://aimath.org/textbooks/approved-textbooks/> (kind of refereed overlay over the virtual corpus of all online open access textbooks, or gateway to a selection made by an editorial committee with transparent evaluation criteria).

- The question of judging: often a developing country doesn't have the community to do peer review. The problem arises when comparing between academic areas. So what should be done systematically? Even getting to a final stage, when peer review might be employed, needs some kind of statistical test. Scopus quantile ratings were mentioned as being helpful. This improved the funding of mathematics. How about other factors, such as number of articles per author.

JHD/CEIC

⁴JHD: there seems to be very little here though: https://en.wikibooks.org/wiki/Subject:University_level_mathematics_books only lists on complete book, on Linear Algebra.