

Efficient management of high level *XMM-Newton* data products

Ivan Zolotukhin (IRAP / Moscow U)

Talk outline

- ✦ Who I am
- ✦ *XMM-Newton* catalog and web app motivation
- ✦ Website: what batteries inside?
- ✦ New development model (citizen science)
- ✦ Recent science cases
- ✦ Conclusions

Who I am

- ✦ 2003–2009: large social networks, PostgreSQL consulting, independent consulting to Top10 Internet companies, POC member of the world's largest IT conference series
- ✦ 2009: PhD in astrophysics, Moscow State U
- ✦ 2010–2012: Paris Observatory
- ✦ 2012–now: IRAP, Toulouse, France

Who I am

- 2003–2009: large social networks, PostgreSQL consulting, independent consulting to Top 10 Internet companies, , POC member of the world's largest IT conference series

IT

-
- 2009: PhD in astrophysics, Moscow State U
 - 2010–2012: Paris Observatory
 - 2012–now: IRAP, Toulouse, France

Research

XMM-Newton catalog

- *XMM-Newton* is X-ray observatory by ESA launched in Dec 1999
- Large FOV: ~70 sources / pointing (very prospective for serendipitous science)
- 3XMM-DR5 (arXiv:1504.07051) is the largest X-ray source catalog ever created: 2.5% of the sky, 560k detections, 400k sources (credit: XMM Science Survey Center)



Motivation for the new catalog app

- ✦ I was main responsible for the 3XMM-DR5 catalog compilation
- ✦ Old XMM-Newton catalog webpages: expensive to take over
- ✦ Wealth of high level science products by the community (XMM SSC) but without convenient access
- ✦ Time cost comparable to reimplementing improved version from scratch
- ✦ Unlimited source of a manpower as an experiment: why not trying it?

ARNIE Index
ARNIE Quick Help
ARNIE Tutorial

Search...

All Databases
All Helpfiles

For comments or help, e-mail:
ledas-help@star.le.ac.uk

Database: 3XMM

XMM Third Serendipitous Source Survey Data Release 4: 3XMM-DR4

[Database HELP](#)

[Database Index](#) | [Basic Search](#) | [Advanced Search](#)

Name Resolver

[HELP](#)

Name:

Search Co-ordinates

[HELP](#)

Co-ords:

RESOLVE NAME

SUBMIT QUERY

Co-ordinate system:

☒ Equatorial ☐ Ecliptic ☐ Galactic

Equinox: ☐ 1950 ☒ 2000

Search Type

[HELP](#)

- ☒ Cone search, radius: arcmin.
☐ Square search, width: arcmin.
☐ Rectangle search, size: x arcmin

Display Columns

[HELP](#)

- ☒ Display default table columns
☐ Display all table columns

Output Options

[HELP](#)

Output coordinates in:

☐ Decimal ☒ Sexagesimal

Output system:

☒ Equatorial ☐ Ecliptic ☐ Galactic

Output epoch:

☒ J2000 ☐ B1950

Output format:

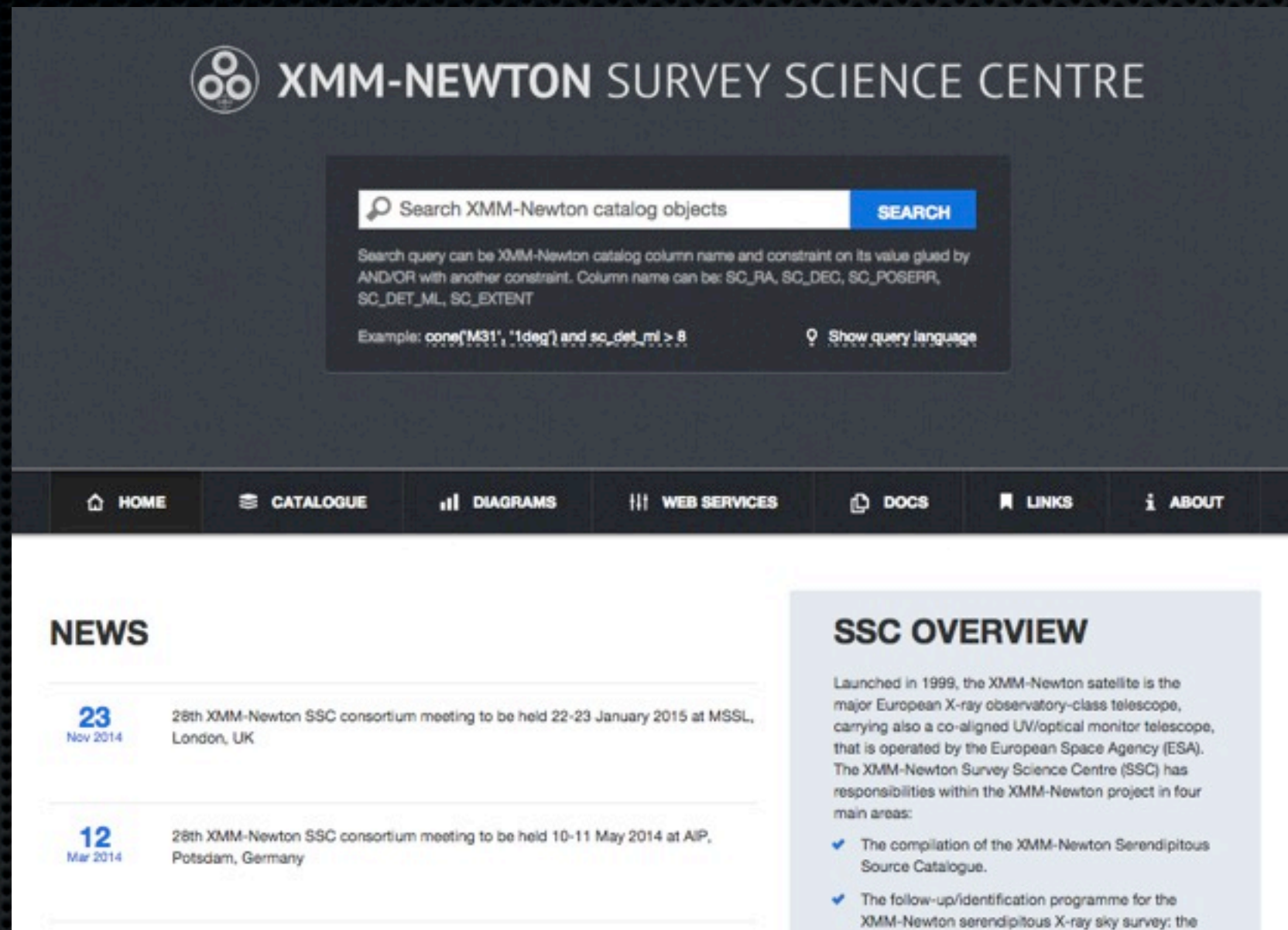
HTML Table

Output number of lines:

SUBMIT QUERY

XMM-Newton catalog

before

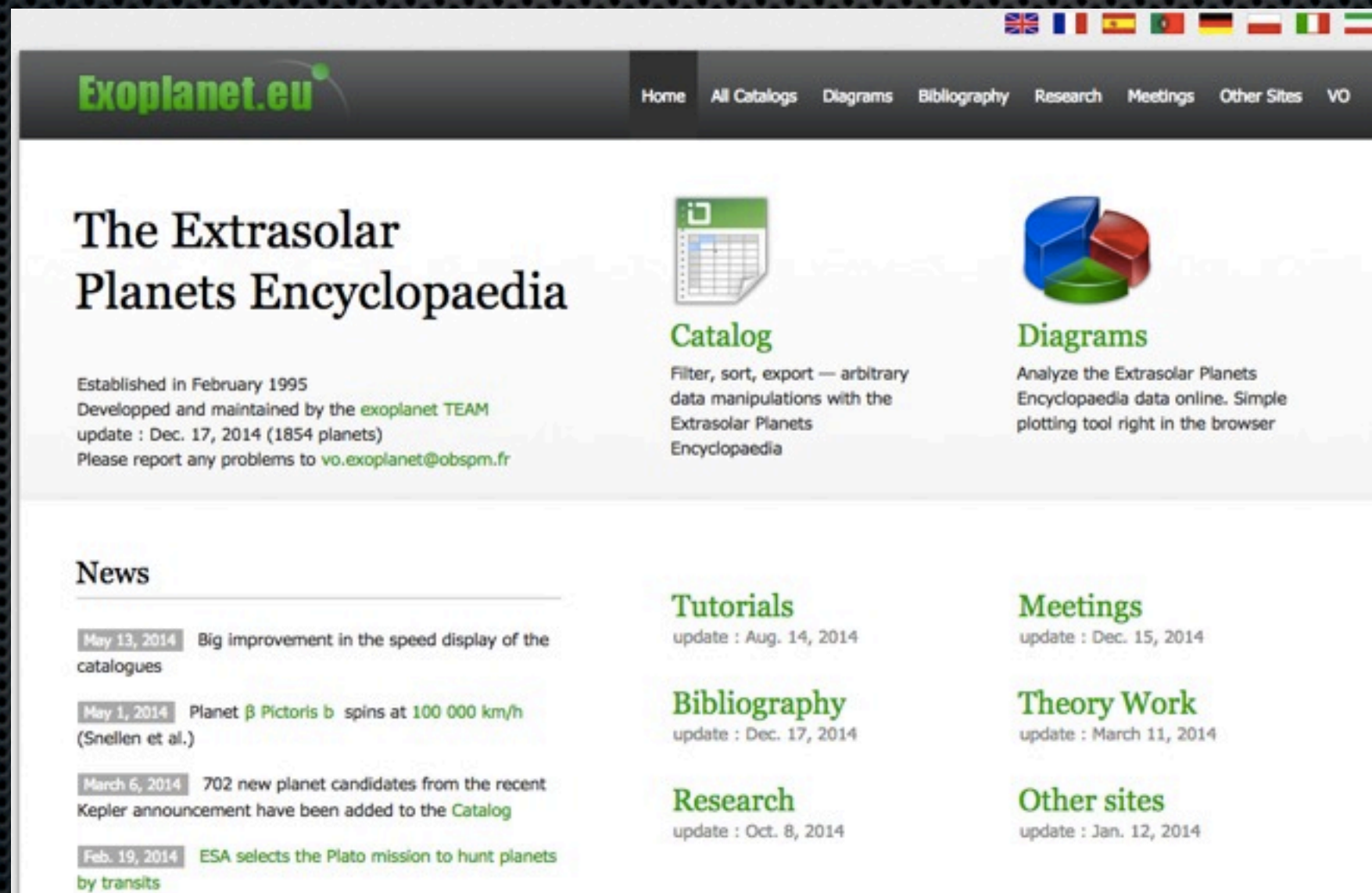


XMM-Newton catalog

after

Legacy

Reincarnation of the world reference exoplanet database, <http://exoplanet.eu>



The screenshot displays the Exoplanet.eu website. At the top, there is a navigation bar with the site logo and a menu including Home, All Catalogs, Diagrams, Bibliography, Research, Meetings, Other Sites, and VO. A row of flags represents various languages. The main content area is divided into several sections. On the left, 'The Extrasolar Planets Encyclopaedia' is highlighted, noting its establishment in 1995 and its maintenance by the exoplanet TEAM, with a recent update in December 2014 listing 1854 planets. To the right of this are two featured tools: 'Catalog', which allows for filtering, sorting, and exporting data, and 'Diagrams', an online plotting tool. Below these are six more sections arranged in two columns: 'News' (with recent updates on catalog speed, planet rotation, Kepler candidates, and the Plato mission), 'Tutorials' (updated August 2014), 'Bibliography' (updated December 2014), 'Research' (updated October 2014), 'Meetings' (updated December 2014), and 'Theory Work' (updated March 2014). The 'Other sites' section is also present, updated in January 2014.

Exoplanet.eu

Home All Catalogs Diagrams Bibliography Research Meetings Other Sites VO

The Extrasolar Planets Encyclopaedia

Established in February 1995
Developped and maintained by the **exoplanet TEAM**
update : Dec. 17, 2014 (1854 planets)
Please report any problems to vo.exoplanet@obspm.fr

Catalog

Filter, sort, export — arbitrary data manipulations with the Extrasolar Planets Encyclopaedia

Diagrams

Analyze the Extrasolar Planets Encyclopaedia data online. Simple plotting tool right in the browser

News

- May 13, 2014 Big improvement in the speed display of the catalogues
- May 1, 2014 Planet β Pictoris b spins at 100 000 km/h (Snellen et al.)
- March 6, 2014 702 new planet candidates from the recent Kepler announcement have been added to the **Catalog**
- Feb. 19, 2014 **ESA selects the Plato mission to hunt planets by transits**

Tutorials

update : Aug. 14, 2014

Bibliography

update : Dec. 17, 2014

Research

update : Oct. 8, 2014

Meetings

update : Dec. 15, 2014

Theory Work

update : March 11, 2014

Other sites

update : Jan. 12, 2014

Technology stack

- ✦ RDMBS: **PostgreSQL**



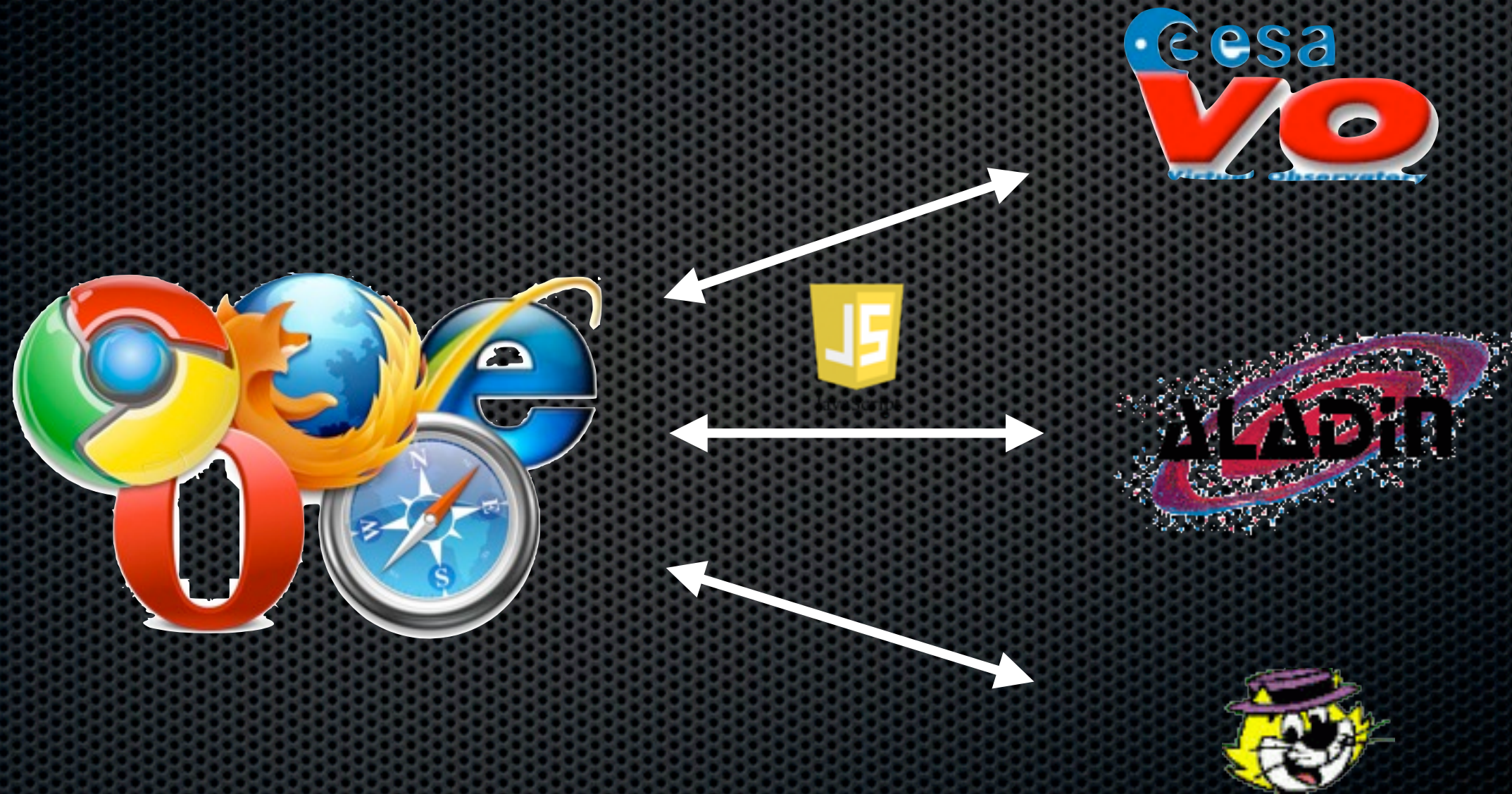
- ✦ Application language: **Python**



- ✦ Web framework: **Django**



SAMP in a browser




AstroTools library: <http://goo.gl/zyF0id>

X-ray spectral fitting

- ✦ Web implementation of a complex thing: source, background, RMF, ARF
- ✦ Wrapper over **Sherpa**
- ✦ Powered by **Xspec**



X-ray spectral fitting

**XMM-NEWTON**
SURVEY SCIENCE CENTRE

SEARCH

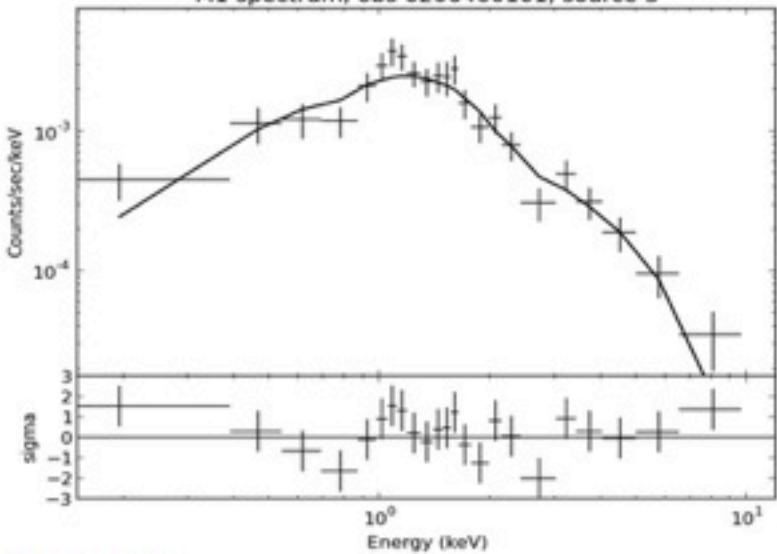
Show query language

[HOME](#) [CATALOG](#) [DIAGRAMS](#) [WEB SERVICES](#) [DOCS](#) [LINKS](#) [ABOUT](#)

[Search results](#) > [Source 200068101010002](#) > [Detection 102064001010003](#) > [Spectrum fitting](#)

[Direct spectrum plot url](#)

M1 spectrum, obs 0206400101, source 3



Counts/sec/keV

Energy (keV)

sigma

Fit:Dataset = 1
Method = levmar
Statistic = chi2specvar
Initial fit statistic = 222.031
Final fit statistic = 22.0104 at function evaluation 46
Data points = 23
Degrees of freedom = 20
Probability [Q-value] = 0.339948
Reduced statistic = 1.10052

Instrument: M1

Energy range min (in keV): 0.2

Energy range max (in keV): 10

Group counts (per bin): 20

Optimization: levmar

Statistics: chi2specvar

Model: phabs_pl

	value	min	max	frozen
nH	0.06684209650	0	100000	<input type="checkbox"/>
PhoIndex	1.27869586947	-2	9	<input checked="" type="checkbox"/>
norm	0.00000346838	0	1e+24	<input type="checkbox"/>

Fit spectrum [Spectral fitting docs](#)

Query language

Boolean expressions instead of endless forms
(think Google) + SIMBAD resolver

Query examples

- `M82` – select sources in 10 arcmin vicinity of M82 center
- `cone('M31', '1deg') AND sc_det_m1 < 100` – select faint X-ray sources not
- `is_ulx = true AND n_detections > 2` – select ULXs which were detected m
- `iauname IN {"3XMM J053406.7+220337", "3XMM J053406.6+220438"}` – select
- `srcid IN {3, 4}` – select specific sources by their source IDs (useful for la

All-time *XMM-Newton* photon database

- ✦ 100 billion photons – that's all photons ever registered by the *XMM-Newton* for 15 years
- ✦ JS9 to access combined images
- ✦ Events extraction and barycentering
- ✦ Bringing (python) codes to the data
- ✦ “*Give me all photons from neutron stars*”

All-time *XMM-Newton* photon database

XMM-NEWTON SURVEY SCIENCE CENTRE [Show query language](#)

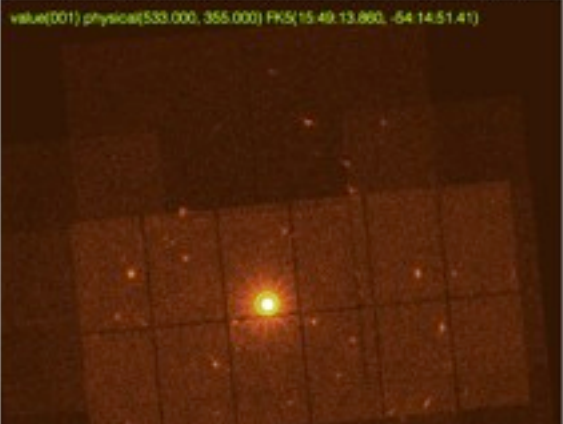
[HOME](#) [CATALOG](#) [DIAGRAMS](#) [WEB SERVICES](#) [DOCS](#) [LINKS](#) [ABOUT](#)

Search results > Source 202039101010002 > Detection 106048801010001 > Events extraction

3XMM J155054.1-541824

Get data File View Zoom Scale Color Region WCS Analysis Help

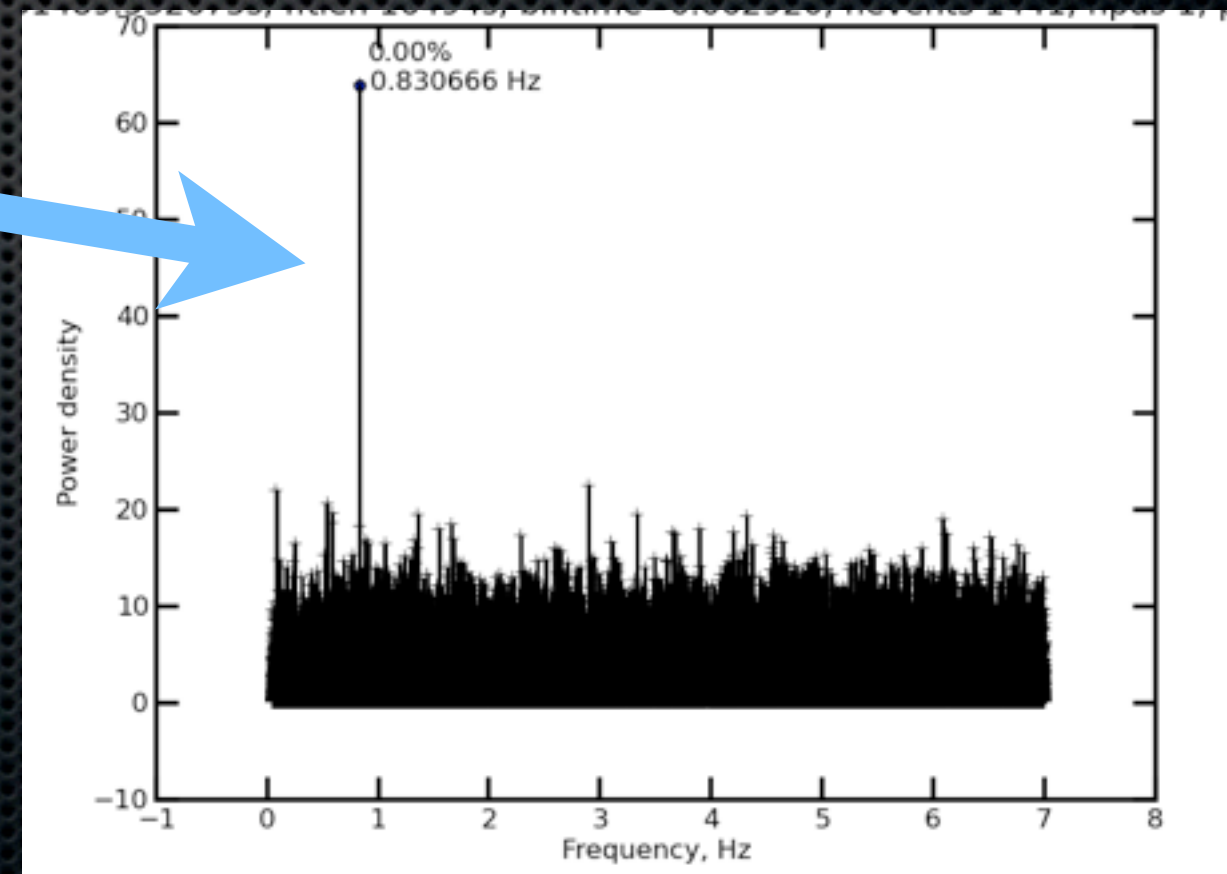
value(001) physical(533.000, 355.000) FK5(15.49:13.860, -54.14:51.41)



Detection image and event list tools
This is a combined XMM-Newton image for observation 0604880101 with source 202039101010002 (source 3XMM J155054.1-541824) marked as a circle region. Green region means its radius is S/N optimized and was produced by the XMM-Newton pipeline. Blue region means the pipeline has not provided an optimized region for this detection (this happens e.g. for faint detections) and the radius is arbitrary set to 40". You can resize and move the region by clicking on it.

Mouse click somewhere on the image, hold and move to change image contrast. Use menu buttons to access image zoom, scaling, palette and other JS9 features. Use Get data menu item to extract (barycentered) photons within specified region from PN instrument event list. This will create an OGIP-compatible FITS file with arrival times and coordinates of photons and also their energies. The file includes good time intervals (GTI) and is suitable for offline analysis with most popular timing tools.

Please note that events extraction and barycentering can take significant time (up-to few minutes), so be patient and do not reload the page. When the extraction is finished, a pop-up window will be opened with the link to FITS file download. Please close all pop-up windows from this website.



More batteries

- ✦ Web sessions (personalization)

- ✦ Name resolver



ANGULARJS
by Google

- ✦ JavaScript diagrams (jQuery, Angular)

- ✦ Aladin Lite by CDS

- ✦ JS9

- ✦ Java WebStart of Aladin and TOPCAT

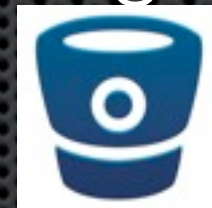


Technical essentials

- ✦ 120k of X-ray spectra
- ✦ 7500 observations (event lists)
- ✦ ~1 TB data volume (7k euro infrastructure)
- ✦ We took existing SSC data products available publicly in the ESA's XMM archive and just organized “scientific” access to them, so that researchers from other domains can do quick-look science on the web

New development model

- ✦ High-level full-time employed IT engineers a.k.a. volunteers



git

- ✦ Coordination through Bitbucket (git)
- ✦ New type of citizen science? Unlimited source of free manpower?

New development model

- ✦ Project duration: ~1.5 yr
- ✦ My time: 5% FTE
- ✦ Volunteers time: up to 3-4 months FTE
- ✦ This team is so far unique, but there are much more citizen science enthusiasts!

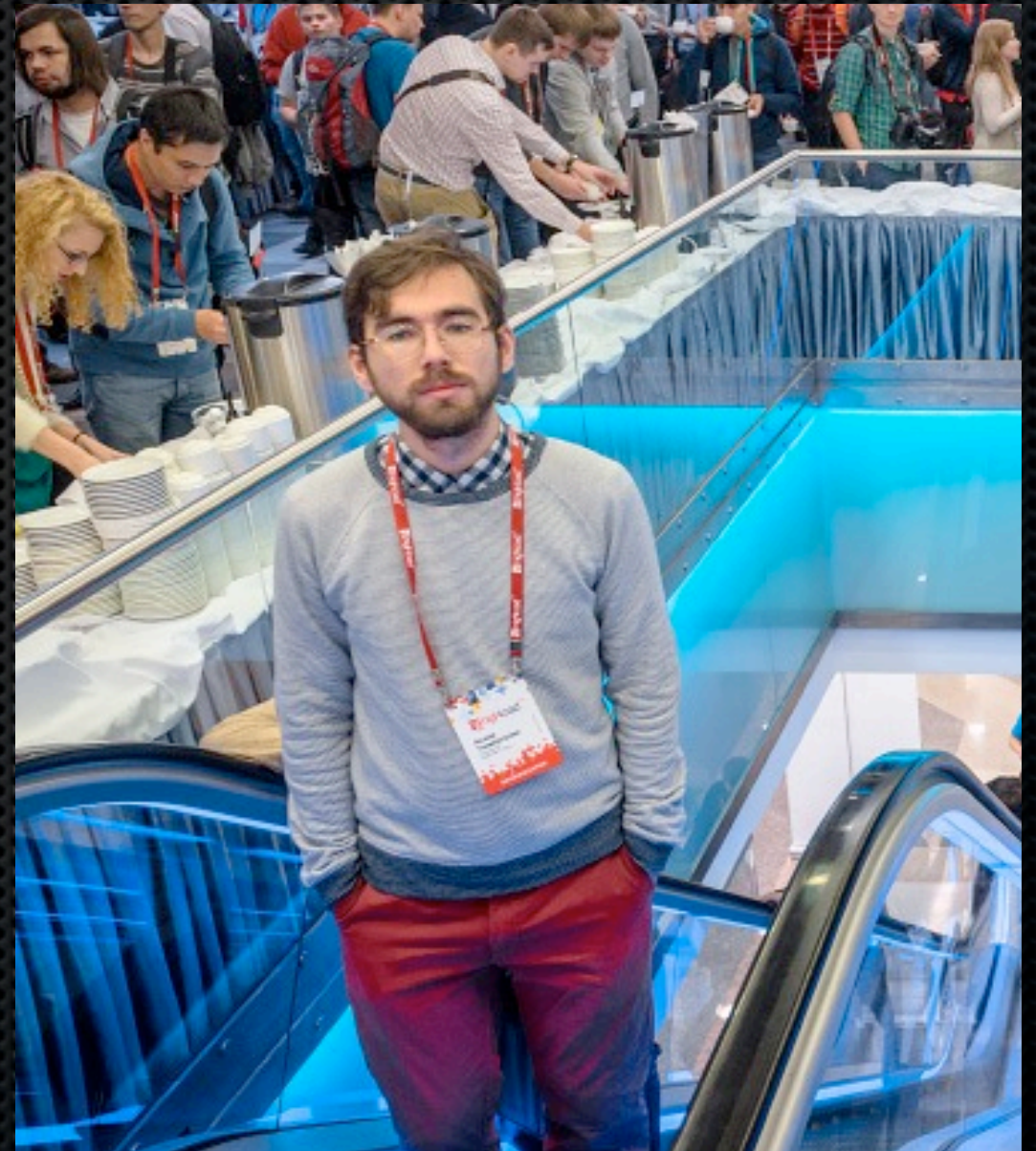
Citizen scientists

Alexey Sergeev,
deputy VP at Mail.RU,
Moscow – *design,*
artwork



Citizen scientists

Askar Timirgazin,
programmer at Ontico, St
Petersburg – *front-end,*
JavaScript



Citizen scientists

Maxim Chernyshov,
programmer at ?,
Vladivostok –
back-end, Python



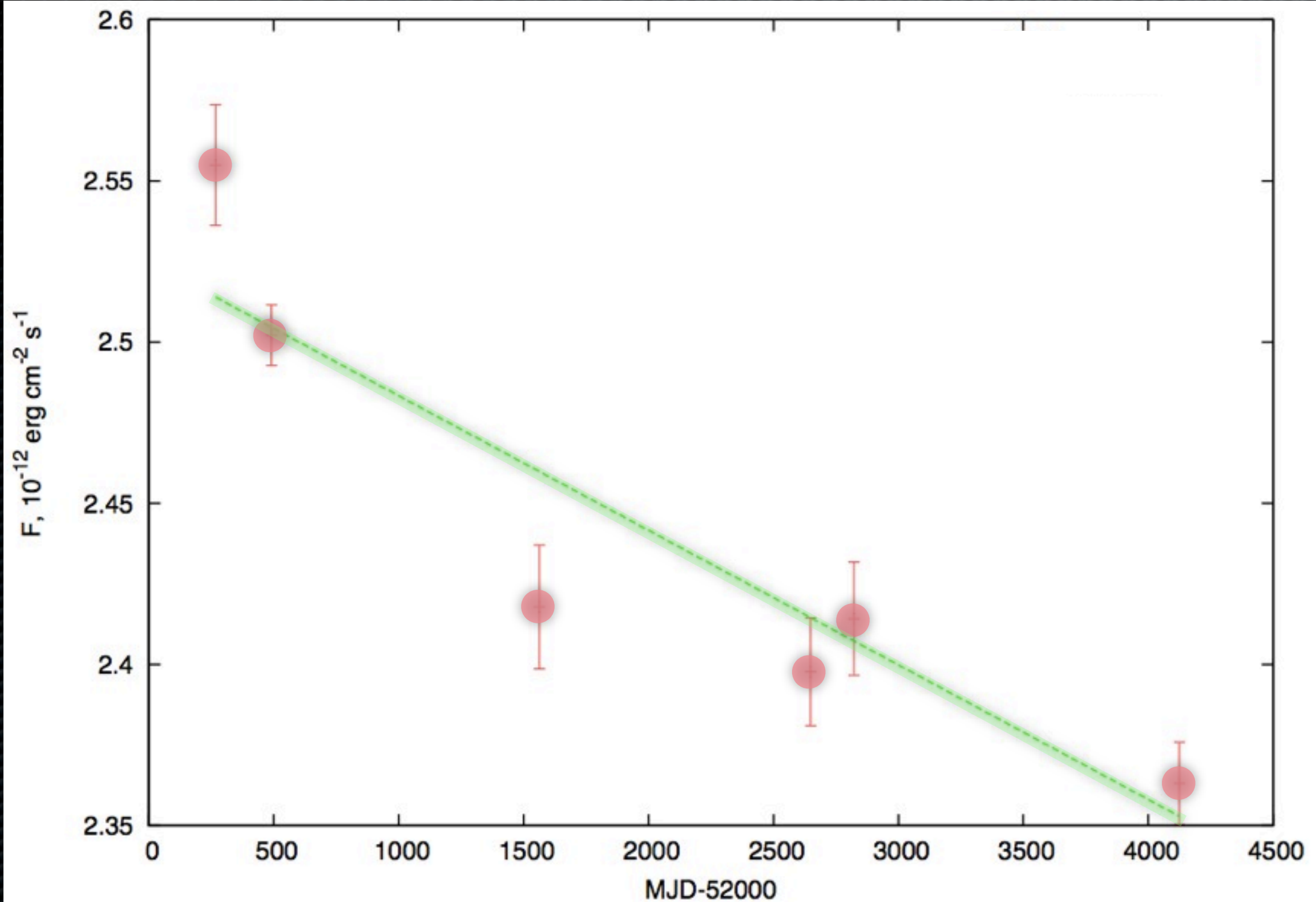
Science cases

(after few months of operation)

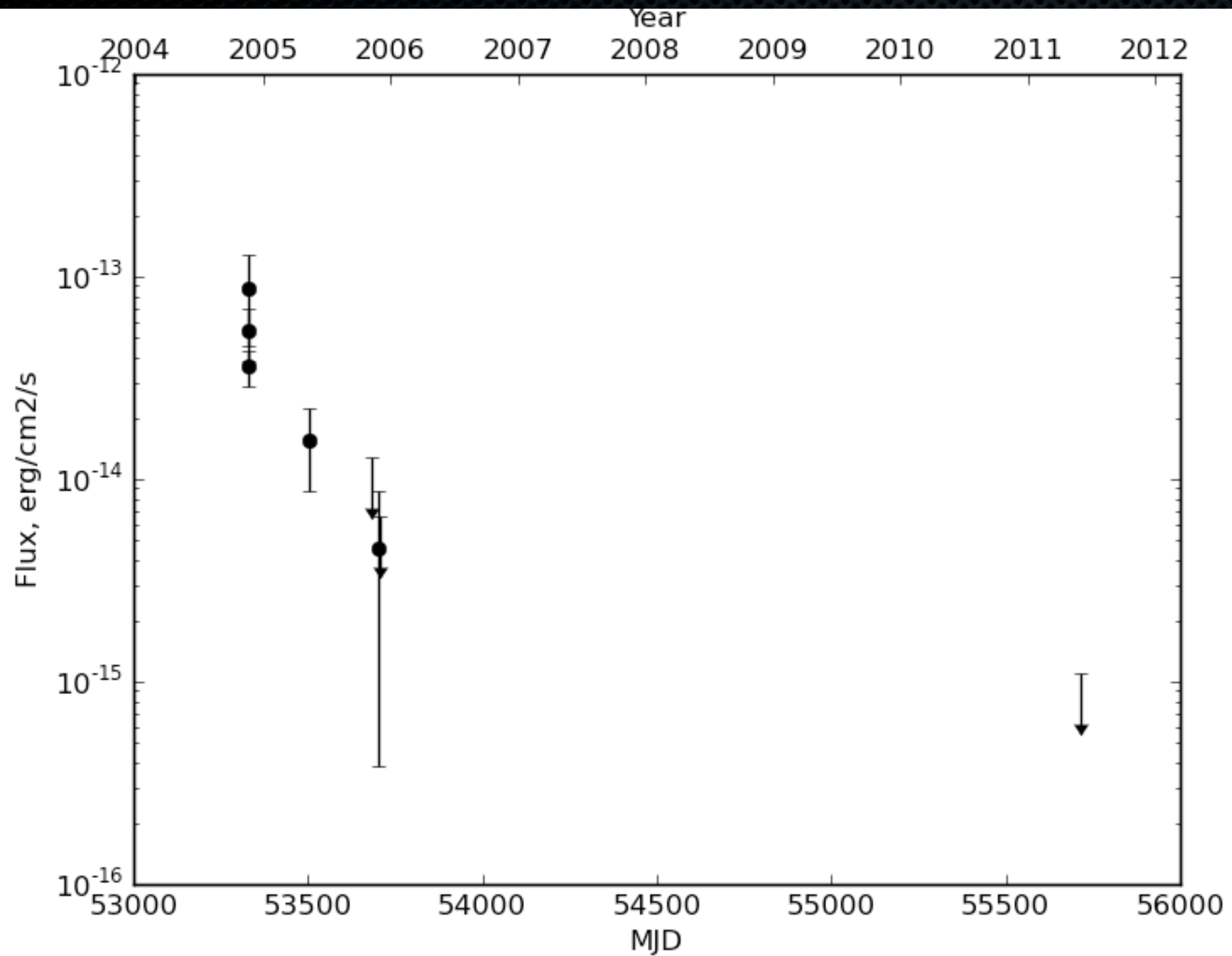
- ✦ Low-luminosity AGN
- ✦ New cataclysmic variables
- ✦ Discovery of the 1st non-recycled pulsar (pulsar factory)
- ✦ Discovery of a cooling neutron star
- ✦ Tidal disruption events
- ✦ Population of hyper-luminous X-ray sources

Ask for the demo, let's discover
something at the coffee break!

<http://xmm-catalog.irap.omp.eu>



2nd known cooling neutron star

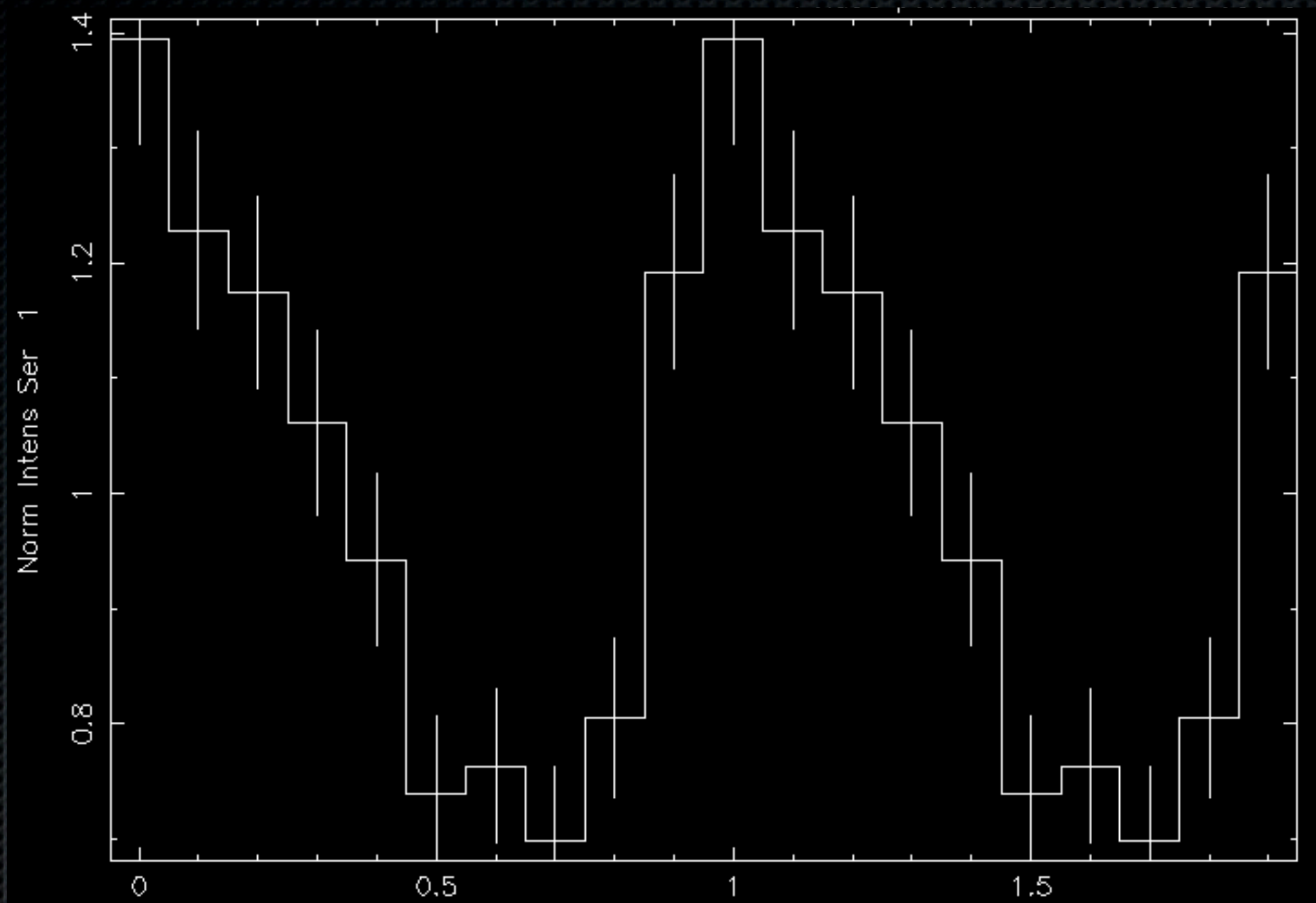


Tidal Disruption Events



1st pulsar in ... Andromeda

1st non-recycled pulsar: too slow rotation for such intensive accretion ($P = 1.2$ sec, $L_X = 1e38$)



1st pulsar in ... Andromeda

1st non-recycled pulsar: too slow rotation for such intensive accretion ($P = 1.2$ sec, $L_X = 1e38$)

Credits

- ✦ N. Webb (PM of the XMM SSC) – hardware
- ✦ M. Bachetti, N. Sartore, I. Chilingarian, M. Pshirkov – research collaboration

Future

- ✦ Wealth of research to be published
- ✦ Tens of new pulsars in these data
- ✦ *eRosita* all-sky X-ray catalog
- ✦ *Chandra?* *Swift?*
- ✦ ...but the author is on the verge of leaving astronomy

Lessons learnt

- ✦ Source catalog = table of contents of the archive (which means pipeline and archive teams should be the same!)
- ✦ Interoperable data and software is a must
- ✦ All-time (initially heterogeneous) observations should be seen as a single event list
- ✦ Bringing code to the data (yet manually) matters
- ✦ Big archives should provide quick-look science tools on the web

Message to take away

- ✦ <http://xmm-catalog.irap.omp.eu> is a new archive model where prominent discoveries can be made online
- ✦ X-ray data (photon lists, source/background spectra, RMFs, ARFs, timeseries) and tools (Sherpa, JS9, Aladin Lite, TOPCAT) form powerful analysis environment and finally help reach research **synergy**
- ✦ Tiny motivated team of IT professionals (a.k.a. **citizen scientists**) can create extraordinary projects