INTERNATIONAL
SPACE
SCIENCE
INSTITUTE

Hallerstrasse 6 • CH-3012 Bern • Switzerland

ISSI Forum on
"The Impact of Big Data in Astronomy"

4-5 July 2019 at ISSI premises in Bern,
Switzerland

FINAL REPORT

Summary

The ISSI Forum "The Impact of Big Data in Astronomy" took place on July 4 and 5 at ISSI Bern. The rationale behind this topic is the fact that the explosion of astronomical data in volume and complexity is changing the way we do science. Space-based and ground-based observatories like Gaia, Euclid, LSST or SKA, as well as state-of-the-art astrophysical simulations require new tools, means and methodologies for doing research. Among the new techniques range data mining, artificial intelligence, machine learning, pattern recognition and other data-driven methods.

At the ISSI Forum, 26 international experts met in order to discuss these new developments. Underlying questions were, e.g.: How do large surveys and Big Data science change astronomy? What additional training/skills will future astronomers require? What are the challenges for data storage, data curation, data quality, data provenance? Four keynote talks on "Big Data Challenges in Observations" (by Wil O'Mullane), on "Data Challenges in Archiving" (by Françoise Genova and Bruno Merín), on "Big Data Challenges in Simulations" (by Volker Springel) and on "Big Data Challenges – Now and in the Future" (by Alex Szalay) formed the backbone of the Forum. Nine short contributions were presented by participants. The most important and interesting part of the Forum were the very lively and interesting discussions, they gave a very broad and comprehensive view on the benefits and

challenges of “Big Data in Astronomy”. In addition to the demanding technical aspects, also “social” and “sociological” topics were addressed: How can young as well as experienced scientists be taught about these new techniques? Will the future successful scientist have to be an astrophysicist, AND a computer scientist, AND a software engineer? Or can future Big Data science only be done in teams with experts from all these fields? How much knowledge in data science is necessary for an astronomer? Can young scientists knowledgeable in data science share this with “classical” older astronomers and teach them the new tools? What are the career perspectives and paths for young scientists who specialize in “data science”? At the end of the Forum, it was felt that this successful “kick-off” should be followed by suitable further activities, e.g. within IAU or EAS or COSPAR, or maybe with a further event at ISSI.

Introduction

The explosion of digital data in volume and complexity available through internet has driven a revolution in handling large flows of information. This challenge has developed into new opportunities in many domains like health, transport, security, tourism, or e-business, with the blooming of new applications. Computers changed the way to manage data few decades ago and the new challenge is to exploit larger and more diverse amounts. In fact, we refer to Big Data when the volume itself, as well as complexity and heterogeneity, becomes part of the problem, when available techniques are not good enough. Scientific research, and astronomy in particular, should of course benefit of the new data handling tools, like it did with computers and the world wide web before, and deep learning, an active research area in machine learning and pattern recognition, offers excellent opportunities.

Astronomy is indeed a paradigm case for Big Data science. The continuing development of ground and space-based observatories, including large sky surveys, is bringing astronomy to the Big Data era. Gaia or Euclid are examples in space but new ground-based projects, like LSST or SKA, will need the new tools even more. Means and methodologies to do research with these facilities will, no doubt, be needed.

Young astronomers are more and more involved in the use and development of Big Data science for their research, not only mining databases to get answers, but also producing complex simulations or finding new questions by recognizing unexpected patterns in the data, in other words, moving from a model-driven to a data-driven approach. Moreover, the traditional way of asking for observing time to investigate specific targets is also being drastically modified by increasingly complete surveys.

The purpose of this Forum was to convene a number of experts in the use of Big Data science for astronomy with the aim of reflecting on the benefits and challenges of this research tool, now and in the coming years, what is its current status and if we need to worry about it.

Big Data Challenges in Observations

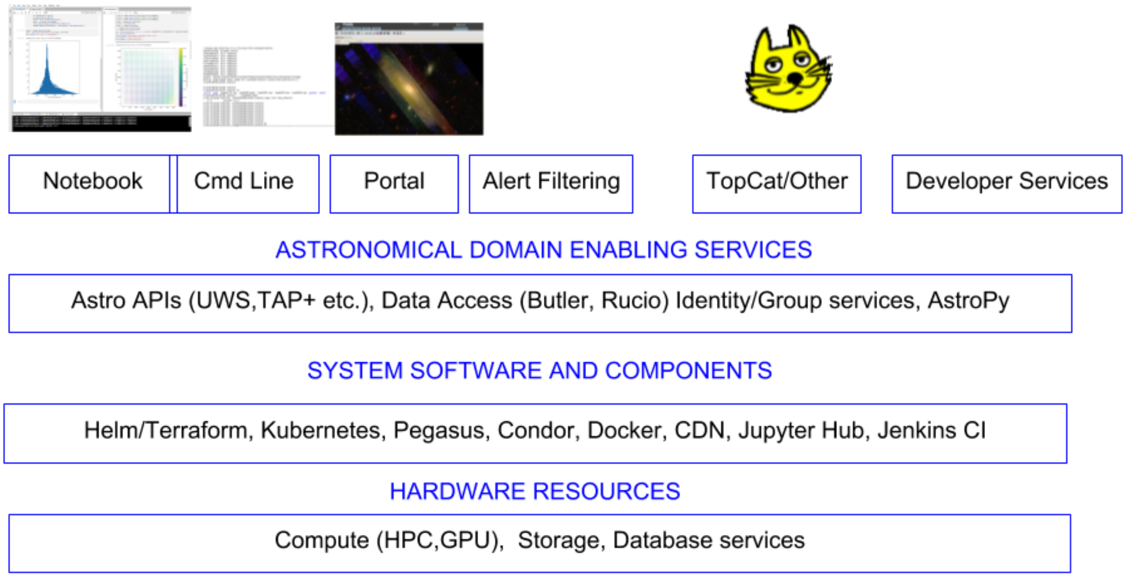
The Gaia mission when it was being conceived was thought of as Big Data. As time progressed the data volume became relatively small, it was and is still considered Big Data in terms of complexity. Gaia is changing astronomy today and beginning to rival HST in terms of number of publications, this is especially so in the areas of Galactic structure and formation but also in smaller area studies or using a few objects - this Big Data is assisting both big and small science. As a technologist/astronomer deeply involved in Gaia one asks "What next?". For many of us LSST is the logical extension of Gaia, underpinned by Gaia's reference frame LSST will extend the survey out several more magnitudes. The combined surveys allow probing of theories across astronomy with LSST really opening up dark matter/energy possibilities.

LSST and Gaia both had simple plumbing problems - large amounts of data have to be shipped around substantial distances - both embraced the network requiring investment in higher speed interconnects at various locations. Euclid will challenge the ESA infrastructure even more. We shall have to continue improving infrastructure.

As we move to deeper observing with LSST new problems arise, not only do the data volumes increase but the complexity also increases. As we stack images to gain access to the 27th magnitude practically all objects are blended, the photometry becomes very difficult to pin down, there are unknown statistical distributions, truncated, censored and missing data and unreliable quantities.

Neither Gaia nor LSST have observing proposals - all data access is via the archive, and a great deal of effort is put into making the archive useful for doing science. Indeed, the trend across the board is an increase in the number of papers based on archival data (now over 50% for the Hubble Space Telescope). The paradigm is shifting - astronomers are actually looking to see if the observation they want has already been made. As our data volumes increase this shift will intensify, data will be looking for astronomers not vice versa. As pointed out later - everyone says they want more data while meaning they want data *more relevant* to their science case. We shall have to work ever harder on making our data access more relevant to the science cases while removing the barriers to access. This would include removing or significantly reducing proprietary periods for data rights. Though we may not all need to become data scientists there is a clear shift in this direction seen by the popularity of data science courses in educational institutions. We must be ready for a more savvy user base with new demands.

As volumes grow and more significant fractions of mission money is put in software our constant reinventing of the wheel will no longer be tolerable. Some missions have understood processing software as essential. About 10% of Gaia budget was for processing software (20% or more if you include the substantial community contributions) and about 20% for LSST - not all missions do this. Astronomy should agree on the components of a Cyber Infrastructure such that we can have one or two implementations of each component instead of one per institution or mission. A potential reference architecture for astronomy using LSST as a basis is presented in dmtn-115 (see footnote), the diagram from which is reproduced below.



Also as data volumes grow the filesystem becomes a real bottleneck - we need to move to industry standard object store such as used by Amazon, Facebook, Google etc.[1] Our traditional approaches to data processing such as shared nothing batch processing may not scale well to new problems where data is more connected. To get the Gaia astrometric solution a significant fraction of the data must be processed as a single data set, on LSST the Forward Global Calibration Model will have a similar requirement. Machine learning on the user end will challenge our current approaches to data access for users. It is clear we need standards here but not clear we can wait for a long standardisation process, we need to accelerate the process somehow.

We need to consider long term support of software. Open sourcing is one way to increase engagement and allow public scrutiny - from a science perspective it matters not who does the science, more eyes means more science. This also implies a better support system for cross disciplinary individuals in astronomy and better education for everyone (especially managers) on handling open source and cross disciplinary issues. This education must extend and continue to inclusion in general - we still have dismal gender balance issues in astronomy.

Big Data Challenges in Archiving

According to the Wikipedia, "*Big data* is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are *too large* or *complex* to be dealt with by traditional data-processing application software." While most current astronomy datasets are typically of the TeraByte scale, it is definitively their complexity and

[1] https://dmtn-115.lsst.io/

heterogeneity that make them candidates for the application of Big Data techniques. More specifically this will become a necessity as catalogues and hence query results grow in size. For example, analysis of the queries against the Gaia Data Release #2 (DR2) catalogue from most papers published during 2018 show result tables with typical sizes of a few hundred MegaBytes. In the course of the next few years, several new surveys will start to produce PetaBytes of data, which will result in TeraByte query results. This will force data centres to support users in applying Big Data techniques.

Possibly, the largest Big Data challenge for archiving might be the development of a set of systems that will allow the AI-supported exploration and exploitation of the ever-growing datasets with user friendly interfaces. These systems should allow users to move from querying for data to having conversations with their AI-supported research assistants to refine and enrich the exploration patterns of the existing data and of the associated literature. Astronomy can take advantage of the quick evolution in industry of these types of applications. Many astronomy datasets are today fulfilling the FAIR principle (Findable, Accessible, Interoperable and Reusable) on scientific data preservation making it easier to apply non-astronomy applications. This important foundational work has been made possible by fruitful collaboration in the context of the International Virtual Observatory Alliance[2] over more than 15 years that has now matured and is ready to move into higher-level user services. In order for this future vision to be realized, astronomy data centres should learn about and deploy AI systems to automate the complex domain-specific tasks of data curation and preservation (by e.g. optimizing global cost functions). Data centres need to move away from current very human-intensive processes by implementing AI-supported data sampling and dimensionality reduction routines in data processing chains. This should be underpinned by developing code-to-the-data platforms that commoditize the computing resources for users, breaking barriers between data silos and embracing contemporary data science.

Preliminary exploration of the impact of Big Data techniques in astronomical archives shows big potential for uncovering new patterns in data in a relatively fast way, potentially linkable to new astrophysical phenomenology, impossible to detect with traditional data analysis workflows, restricted in most cases to small samples of the full data collections. A recent example of this paradigm shift is the discovery with the full Gaia DR2 catalogue of traces of the stellar population from another galaxy that merged with the Milky Way many million years ago.

The second realization from early work with Big Data techniques in astronomy identifies the urgent need to expand and/or re-prioritize work at the data centres to support the higher demand in data preservation and computing infrastructures as well as to train internal staff and the user community into the new techniques. In particular, one technical challenge to overcome is the connection of various data silos with volumes in the TeraByte or Petabyte scales, which are not easily shareable with current network performance.

[2] ivoa.net

Big Data Challenges in Simulations

Numerical simulations on high-performance supercomputers have become a backbone in theoretical astrophysical research. Such simulations provide solutions for complex systems of partial differential equations which cannot be solved analytically, thereby allowing precision tests of physical hypotheses and the conduction of virtual experiments that would otherwise be impossible. The relentless growth of the peak performance and storage capacity of large computers over recent decades has enabled rapid progress in the physical fidelity of modern simulations, causing an overall steady expansion of their share in the astronomical research literature. In addition, the data volume of simulations has exponentially grown, too, and now creates both new problems and opportunities in their own right.

A number of successful public data releases of simulations of cosmic structure formation over the past decade highlight the huge benefit of making simulation data widely available and usable beyond the researchers originally carrying out the calculation. For example, the Millennium Simulation of 2005 released its galaxy catalogues through an SQL-queryable database infrastructure, which was widely picked up by the community. More than 1000 research papers have been written with this data thus far, and it keeps being used even today. More recently, hydrodynamic simulations of galaxy formation like the Illustris, IllustrisTNG, and Eagle simulations have provided even richer, and much larger, datasets to the astronomical community, which are on track to reach comparable or even larger impact. The data volumes concerned here already reach several hundred TB, making it technically very challenging to provide the data in a usable form to the full astronomical community. Currently, this is addressed by allowing downloads of parts of the raw data (in HDF5 format), and through a custom-made API that allows filtering of the data through powerful server-side queries.

The problem of data volume is much more acute for the largest cosmological simulations that are presently carried out. These are simulation of cosmic structure formation with several trillion particles in volumes several Gpc across, and which are needed in support of current and future missions to study dark energy, such as Euclid. The simulation data in this case exceeds several PetaByte. Yet, first attempts to make this data publicly available have been made, like in the recent release of the Outer Rim simulations, which amount to nearly 5 PB in total. The enormous size of this data requires kludges like the restriction to release only heavily downsampled data. This in part defeats a central objective of these types of simulations, namely to provide such exquisite statistical accuracy that even small deviations of dark energy from a cosmological constant become measurable with confidence.

It is a largely unsolved problem how simulation groups should cope with the rapidly growing data volume of HPC-based simulation models, and how this data should be made available to the astronomical community to gain the enormous amplification of its scientific impact this could entail. An expensive astronomical telescope is useless without a scientific instrument put onto it. In numerical astrophysics, one can likewise view a supercomputer as a useless machine without a scientific simulation code that can run on it efficiently. But the analogy ends when one considers the funding structures traditionally associated with instrument development and code development. While the former is typically supported with dedicated positions and often large budgets, the latter is commonly expected to arise as a side product of the ordinary research of theoretical astrophysicists. It appears questionable whether this unsystematic approach is still adequate today, given the high cost of supercomputers on one hand, and the need to exploit them to their full capacity on the other hand. The Big Data explosion seen in simulations makes this problem even more acute, as we are also falling short of the required development effort for parallel analysis tools that are needed to process simulation data at the leading edge.

Big Data Challenges: NOW and in the FUTURE

Big Data comes with various challenges. The five most dramatic challenges appear to be: (i) scalability and computability, (ii) statistical variance vs systematic errors, (iii) novel uses of simulations, (iv) sociological changes, and (v) disruptive technologies.

Scalability

Data volume and computing power double about every year. This means indirectly, that no polynomial algorithm can survive, only N log N algorithms are suitable. In order to implement non-polynomial algorithms, we will need incremental algorithms, where computing is part of the cost function, so that we can decide when to stop computing based on the question: what is the best estimator in a minute, in a day, in a week, in a year?

Systematic Errors

The role of systematic errors needs to be better understood. With billions of data points statistical errors are no longer the dominant ones. Large known (and obvious) systematic errors are dealt through calibration. Hence, systematic errors are small by definition. Often they are detected only years after the data was taken, and often they show up only at the edges of the distribution (magnitude limits). With using massively parallel instrumentation, where every device has slightly different characteristics, these are going to be the hardest to tackle now and in the future.

Computer Simulations

High performance computing has developed into an instrument in its own right. The largest numerical simulations approach PetaBytes. They are becoming very expensive as well. We will need public access to the best and latest astrophysical simulations through interactive numerical laboratories. This creates new challenges and questions, e.g.: How to move the PetaBytes of data, how to look at it, how to interface and analyze them, which computer architectures to use (supercomputers, database servers, Jupyter)? An example of the best practices is the Millennium Simulation database with 600 registered users, so far with 17.3 million queries! The Via Lactea-II simulation shows how to use cosmology simulations as an immersive laboratory for general users: Users can insert test particles (“Bring your own dwarf galaxies!") into the system and follow trajectories in pre-computed simulations, that means users can interact remotely with a PetaByte in 'real time'!

The Challenge of Long-Term Data Preservation

As the first generation of these large-scale projects are getting close to the end of their lives as far as data collection is concerned, as the original instruments are slowly becoming obsolete, a new challenge is emerging: *what happens to the data after the instruments are shut down?* This is a much harder problem, than it may first appear. In order to tackle the problem, we need to establish a common "currency" on which one can make easier comparisons and try to formulate a rational decision making process. Let us try to establish the three different aspects of the business model for these surveys: the price, value and cost of the data.

The Price of Data

The data collection created by a Big Science project represents a major public investment of a few hundred million dollars. This includes the capital investment in the experimental facility, the data infrastructure, the cost of operating the instrument, reducing the data, and building and operating an open and accessible data archive. This is the price of the data, this is how much government funding (in some cases augmented by private foundations and individuals) has helped to create this singularly unique resource. Generally, this process is well understood, and all of the aspects of the projects are well under control while the experiment is running.

The Value of Data

We can also ask how we could estimate the value of the data. It is clearly reflected in how much science it generates. While it is difficult to put a monetary value on the results of scientific research in an algorithmic fashion, we can use another approximate metric. Each scientific paper published in a refereed journal represents a research effort that costs approximately $100k (an estimate, but certainly more than $10k, and less than $1M). This is the amount of research funds spent on paying for staff, students, postdocs, research tools, computer time, to be able to write a credible scientific publication. The number of papers based upon the analysis of a given data set are then measuring how much the members of the research community are willing to spend from their own research funds to work on this particular data set, they vote on the value of the data with their research dollars.

The Cost of Data

This is the third component of the problem. This is measuring how much it costs annually to curate, preserve and keep serving the data to the community in an open and accessible way, after the original instrument has been turned off, and there is no new data added to the archive any more. This is more than archiving, as the data use is through intelligent software interfaces, often based on a large database, combined with a collaborative data analysis platform. This requires a lot more than just copying data on disks. Operating systems change, database systems change, web browsers change, computing hardware changes, and the user's expectation is also increasing with time. A few years ago they were happy to download a few flat files and analyze them at their workstations at their home institution, today they are expecting access to iPython notebooks, and GPUs, but soon they will want to reprocess PetaBytes of data on hundreds of computer nodes interactively. Much of the cost is not so much in saving the bytes, but rather keeping the services alive, and up-to-date. As the cost of storage and even computing cycles keep decreasing every year, the dominant part of the costs are mostly in people.

Comparison of Price, Value and Cost

From our 20 years of experience with the Sloan Digital Sky Survey, the price of the data to date has been about $200M. The project's data has generated to date about 9,000 refereed publications, i.e. attracting about $900M of research over this period. After operating the archive for 20 years, we estimate the cost of maintaining the necessary technological advances into the future is approximately $500K/year. Let us express this annual cost in terms of the price: $500K/$200M = 0.25%/year. We can see that a 5% addition to the project's budget would secure the archive for 20 more years. Or, if the continued operation of the archive results in just 5 refereed papers in a year, it is still a reasonable investment to keep the archive alive.

The same numbers for the Large Synoptic Survey Telescope, the current national flagship project for astronomy, are similar. The price is expected to be around $1.2B by the end of the project, and the cost to be about $6M/year, i.e. $6M/$1.2B = 0.5%, still in the same ballpark.

These costs are quite trivial compared to the price of the data, yet we have no coherent plans or long-term funding mechanisms in place to address this problem. A potential loss of one of these data sets would create an enormous damage to science and endanger the national willingness to continue more future experiments, if we cannot demonstrate that past investments are adequately protected, preserved and cared for.

Issues for the future are:

- *Data Lifecycle*: New data standards emerge; metadata standards change; usage patterns change.
- *Service Lifecycle*: Survey data presented in Smart Services; browsers change (HTML5); operating systems change; databases change; servers become obsolete, disks die; new software technologies (iPython) emerging.
- *Long Term Data Publishing/Preservation Models*: Traditionally we included the data in the printed publications ("we threw it over the wall for someone to catch it"); publishers and journals were there to do so (for a profit); this model is cracking. Formal (FAIR) requirements have emerged: Findable, Accessible, Interoperable, Reproducible. Open scientific data is becoming increasingly mandated by funding agencies. The current for-profit model of scientific publishing is under attack, transition is inevitable. But this is all

about publications, what about the data? A FAIR-approach needs more automation, manual approach cannot keep up. What happens to large, high-value data sets when they are completed?

Ideally, data should be open/free, accessible and self-sustaining! In practice: Pick any two, and the third is determined! Relevant questions are: How can one ensure a steady, long-term support? Who do we trust with all this irreplaceable data? How can we decide what to preserve?

Suggestion: Set up a National Endowment / Data Trust for High-Value Data!

Disruptive Technologies – what happens in 30 years?

Major, inevitable changes are coming. It is enough to look at the music publishing:

LPs => iTunes => Pandora!

In the old days, we bought and took home the physical copy of an LP or CD, this is equivalent to downloading the whole data set to our own computers. With iTunes, we can download a very specific song form a large collection. The data equivalent is to perform a database query that returns a small subset of the data precisely matching our criteria. Finally, Spotify and Pandora stream music to us according to some broad criteria, and we do not know the format or what the next song will be, everything happens in the cloud, we just see the results. This is where we are heading with modern machine learning tools, which will find various potential patterns in the data and present it to us for a human evaluation and visual inspection in a broader context, to turn automated detections into potential discoveries – the future we are looking at.
Today we spend of order a $1B per project to acquire valuable data (typical cost of a big project/telescope). Much of these will not be superseded in the foreseeable future/decades. At the end of the projects the data sets will be handed off to someone. We need an organization(s) which

- has a long track record with a predictable future,
- understands data preservation,
- is trusted by everyone,
- is technically capable,
- can run under a sustainable model, and
- has no single points of failure.

Suggestion: Consortium of University Libraries or Data Centres + embedded domain scientists!

How do we prioritize? The data explosion means: science is becoming data driven. It is becoming “too easy” to collect even more data: e.g. with robotic telescopes, next generation sequencers, complex simulations. How long can this go on? If asked: “Do you

have enough data or would you like to have more?", no scientist ever wanted less data… But: Big Data is synonymous with Dirty Data. How can we collect data that is more relevant? We need to improve ideas on survey design… e.g. use artificial intelligence in large-scale experimental design. For example, observing spectra is 1,000 times more expensive than imaging the same objects. We could use reinforcement learning to continuously refine information from observed targets to improve target selection algorithms through machine learning. Active learning will help pick the best objects to observe, given our existing data from the many available targets, what will give the best improvement in our knowledge?
Big Data is at the forefront of astronomy (and science in general). Major disruptive changes happen rapidly. Astronomy is embracing machine learning. So far we made much of it up as it happened, there was no grand plan. There was/is a tremendous impact of key individuals (John Gray, Tony Tyson, Jim Gunn …). The community at large by now is convinced that this is happening, it is ready for transformation. Funding agencies are much slower to adopt. We are now spending billions on surveys, yet there is no coherent plan for long term data when we need it, we have to improvise again. It was/is an amazing journey so far, transforming people's lives and careers. The journey continues … and is accelerated.

Conclusion

A very rewarding Forum was held by ISSI in Bern, there was a high level of engagement and a lively exchange of ideas. This document has summarized the Forum discussions. The following list represents a distillation of the most salient recommendations:

- The developments and provided services linked to Big Data should be driven by user needs, taking the best advantage of new technological capabilities but not driven by these technological capabilities.
- The usage of AI technologies to automate data curation and preservation should be explored and implemented if relevant.
- Researchers should be provided with data and tools relevant to their science using the new technologies.
- The efforts to release simulation/numerical model data should be pursued, this leads to increased data volumes on top of the instrument generated data.
- Coping with increasing data volume requires investment and improvements in e-infrastructure in particular networks and storage - this will be essential for connecting data silos with volumes in TeraByte and PetaByte scales.
- As, out of necessity, the percentage of budget for software increases we must become more efficient and have less duplicated systems for simulation and processing.
- Long term maintenance of software needs to become a project priority.
- Career paths for astro-cyber-infrastructure need to be laid out properly to show people this is a valid and useful career.

Agenda of the Meeting

Thursday, July 4, 2019

09:00 Welcome
Tilman Spohn/Joachim Wambsganss & Alvaro Giménez

09:10 Introduction of the participants

09:30 Keynote 1: Big Data Challenges in Observations
Wil O'Mullane

10:00 Discussion & Contributed Talks to Session 1:
Contributed Talk 1a: The Square Kilometer Array: Exploring the Universe with an Exascale Telescope **Chiara Ferrari**
Contributed Talk 1b: From MWA, ASKAP to FAST and SKA1-LOW **Andreas Wicenec**
Contributed Talk 1c: Changing the way we do science: some notes on reproducibility **Johan Knapen**

10:45 ***Coffee Break***

11:15 Keynote 2: Big Data Challenges in Archiving
Bruno Merín/Françoise Genova

11:45 Discussion & Contributed Talks to Session 2:
Contributed Talk 2a: ESCAPE – European Science Cluster of Astronomy and Particle Physics ESFRI research infrastructure **Simone Campana**
Contributed Talk 2b: Big data in exoplanets and planetary sciences - Sone very brief thoughts **Ingo Waldmann**

12:30-14:00 Lunch break

14:00 Keynote 3: Big Data Challenges in Simulations
Volker Springel

14:30 Discussion & Contributed Talks to Session 3:
Contributed Talk 3a: Data Science in Astronomy **Michelle Ntampaka**
Contributed Talk 3b: Collaborative models and software in the big data era **Arfon Smith**
Contributed Talk 3c: Data Observatory **Jorge Ibsen**
Contributed Talk 3d: Google Cloud **Ross Thomson**

15:15 ***Coffee Break***

15:45 Individual Contributions & Structured Discussions

19:30 ***Dinner***

Friday, July 5, 2019

9:00	Keynote 4: Big Data Challenges NOW and in the FUTURE **Alex Szalay**
9:30	Discussion & Contributed Talks to Session 4: Contributed Talk 4a: SDSS-IV **Anne-Marie Weijmans** Contributed Talk 4b: The ESO ELT **Michael Sterizig**
10:15	***Coffee break***
10:45	Final Discussions and Conclusion
13:00	**End of Forum**

List of Participants

Name	Affiliation
Conveners	
Genova Françoise	CDS, Strasbourg, France
Giménez Alvaro	ISSI, Bern, Switzerland
Merín Bruno	ESA/ESAC, Madrid, Spain
O'Mullane William	LSST, Tucson, USA
Springel Volker	MPA, Garching, Germany
Wambsganss Joachim	ISSI, Bern, Switzerland
Invitees	
Arviset Christophe	ESA/ESAC, Madrid, Spain
Bonnet Roger-Maurice	ISSI, Bern, Switzerland
Campana Simone	CERN, Geneva, Switzerland
Dabin Christophe	CVNES, Toulouse, France
Ferrari Chiara	OCA, Nice, France
Ibsen Jorge	ESO/ALMA, Santiago, Chile
Knapen Johan	IAC, Tenerife, Spain
Kron Martins Alberto	Lisbon University, Portugal
Lieu Maggie	ESA/ESAC, Villanueva de la Canada, Spain

Miyazaki Satoshi	National Astronomical Observatory, Japan
Ntampaka Michelle	Harvard Smithsonian Center, Cambridge, USA
Polsterer Kai	HITS, Heidelberg, Germany
Smith Arfon	STSCI, Baltimore, USA
Spohn Tilman	ISSI, Bern, Switzerland
Sterzik Michael	ESO, Santiago, Chile
Szalay Alex	JHU, Baltimore, USA
Thomson Ross	Google, USA
Waldman Ingo	UCL, London, UK
Weijmans Anne-Marie	University of St. Andrews, Scotland, UK
Wicenec Andreas	The University of Western Australia, Crawley
Wu Ji	NSSC, CAS, Beijing, China