

Radio 2017 - Timetable 2017

- November 2016: BMBF Discussion
- January : Protocol
 - 38 MEU (MPG) + 9 MEU (BMBF)
 - 1,5 MEUR Verbundforschung (2017-2020)
 - **ToDo: SKA Big Data Analytics Project**
 - **ToDo: Legal Entity D-SKA e.V. (cf. CTAO GmbH → ERIC)**
 - **ToDo: Strengthen science cases in SKA**
 - Community Building: Radio Community complementary to Gamma Community → RADIO2017
- March:
 - Negotiations with Francesco Bignami about associate membership
 - New interim Chair of the SKAO Board: Prof. Lars Börjesson
 - New contact people in Abt. 7/BMBF

Radio 2017 - Timetable 2017

- June: Big Data Symposium Berlin
 - Memorandum
 - Astro-/particle Science Cloud for next-level of data together with CERN
 - Inherit SKA in-kind contribution to central computing via CERN-SKA cooperation agreement
 - Ca. 10 University data science labs
 - FAZ Woche, Physik-Journal (→GLOW Webpage?)
 - DPG Arbeitskreis „Data-intensive Physics, IT, & AI“ (mid Nov 2017/March 2018)
 - WE-Heraeus Seminar Series (2018) (Streit/Gülzow/Steinmetz)
 - **ToDo: „Ideenskizze“ for SKA IT-Infrastructure in the frame of a European Science-Cloud (cf. KAPPA initiative of KIT, CDS, Jülich)**

Der Anfang einer Ära

Die neuen Forschungsprojekte der Radioastronomen könnten unser Leben völlig auf den Kopf stellen. Die Zauberformel lautet: Big Data.

Von Sibylle Anderl

Grundlagenforschung hat eine seltsame Eigenschaft: Egal wie weltfremd und anwendungsfern sie erscheint, man ist nie davor gefeit, dass sie ganz plötzlich etwas hervorbringt, das die Welt verändert. Als beispielsweise Anfang des vergangenen Jahrhunderts die Quantentheorie langsam und unter großen Schwierigkeiten Form annahm, erschien sie zunächst wie eine verrückte, rein theoretische Spielerei ohne absehbaren praktischen Nutzen. Heute beruhen etwa 30 Prozent des amerikanischen Bruttoinlandsprodukts auf Erfindungen, die erst durch die Quantenmechanik möglich wurden. Das World Wide Web wurde bekanntlich 1989 am Cern, der Europäischen Organisation für Kernforschung, entwickelt, um den Informationsaustausch zwischen Wissenschaftlern zu optimieren. Weniger bekannt ist, dass auch die Wi-Fi-Technologie, im Deutschen meist W-Lan genannt, ihren Ursprung in der Grundlagenforschung hat: Radioastronomen der australischen Commonwealth Scientific and Industrial Research Organisation (CSIRO) nutzten ihr Wissen über Radiowellen und mathematische Transformationstechniken, um eine robuste und schnelle Datenübertragung in Gebäuden zu entwickeln. Für die 1992 und 1996 eingereichten Patente erhält die Forschungsbehörde noch immer jährlich einige hundert Millionen Dollar an Nutzungsgebühren.

Auch heute bereiten Radioastronomen neue Forschungsprojekte vor, deren eigentliche Ziele mit unserem irdischen Alltag erst einmal wenig zu tun haben. Trotzdem könnten sie ein weiteres Mal unser Leben auf den Kopf stellen. Die Zau-

berformel der heutigen Grundlagenforschung heißt „Big Data“, wobei sich das „Big“ nicht nur auf das reine Datenvolumen bezieht, sondern auch auf die Geschwindigkeiten der Erzeugung und Bearbeitung der Daten. Diese Entwicklung ist nicht ganz neu – praktisch auf allen Forschungsfeldern arbeiten Wissenschaftler heute mit riesigen Mengen an Daten. Deren Speicherung und Auswertung stellt sie vor besondere Herausforderungen.

Insbesondere der Large Hadron Collider, der Teilchenbeschleuniger des Cern bei Genf, an dem 2012 das berühmte Higgs-Boson entdeckt wurde, hat seit seinem offiziellen Start 2008 bereits neue Maßstäbe im Umgang mit Daten gesetzt. Allein im vergangenen Jahr wurden hier 50 Petabyte an Forschungsrohdaten generiert. Ein Petabyte sind eine Million Gigabyte, und eine Million Gigabyte entsprechen rund einer halben Milliarde je 500 Seiten umfassender Bücher, fünfzehnmal mehr, als in der Deutschen Bücherei in Frankfurt und Leipzig an Büchern stehen.

Die Daten, die im Experiment selbst von den Detektoren generiert und unmittelbar vorselektiert werden, sind sogar noch erheblich größer als die 50 Petabyte des vergangenen Jahres. Für die Speicherung und Verarbeitung solcher Datenmengen mussten die Wissenschaftler des Cern sich völlig neue Organisationsstrukturen schaffen. Die Bearbeitung nimmt ein über die ganze Welt verteiltes Rechenetz aus mehr als einer halben Million Rechenkernen in Anspruch. Das besitzt eine Speicherkapazität im Exabyte-Bereich. Ein Exabyte sind 1000 Petabyte.

Das größte Radioteleskop der Welt

Nun machen sich die Radioastronomen auf, in Bezug auf die Datenerzeugung mit den Teilchenphysikern des Cern gleichzuziehen. Das von einer Organisation aus zehn verschiedenen Ländern koordinierte Square Kilometre Array (SKA), das sich derzeit noch in der Designphase befindet und dessen Konstruktion für die kommenden Jahre in Australien und Afrika geplant ist, soll mit einer Gesamtlänge von mehr als einem Quadratkilometer das größte Radioteleskop der Welt werden.

Diese Fläche wird sich aus vielen Einzelteleskopen zusammensetzen: Tausende von Teleskopschüsseln sollen mit Millionen von Radioantennen zusammengeschlossen werden. Bei Frequenzen von rund 50 Megahertz bis zu einigen Gigahertz wird das SKA den Himmel faktisch durchmusternd, um etwa die Entwicklung von Galaxien, Sternen und schwarzen Löchern sowie die Natur der Dunklen Energie zu ergründen, aber auch die Suche nach Spuren extraterrestrischen Lebens steht auf der Liste der Ziele.

Im Jahr 2012 wurde entschieden, in der ersten Projektphase bis 2024 zu zunächst knapp 200 Teleskopen mit einem Durchmesser von 15 Metern in der Wüste Südafrikas auf-



zustellen, während die ersten 130 000 bei niedrigen Frequenzen arbeitenden Radioantennen in der australischen Murchisonregion aufgebaut werden. Beide Regionen liegen fern jeglicher Zivilisation. Das ist wichtig, um Störsignale wie die aus der Handy- und Radiokommunikation zu vermeiden. Insbesondere der australische Standort ist dafür gut geeignet.

Er liegt in einem Gebiet von der Größe der Niederlande, in dem weniger als 100 Menschen leben. Gleichzeitig erzeugt diese Einsamkeit aber auch Herausforderungen in Bezug auf die für das Projekt notwendige Infrastruktur. Die Stromversorgung des am Standort bereits existierenden Vorgängerobservatoriums, des Murchison Radio Astronomy Observatory, erfolgt anhand einer containergroßen Lithium-Ionenbatterie – die größte Australiens –, die immerhin 2,6 Megawattstunden Energie speichern kann.

Vor die größten Herausforderungen werden die Radioastronomen aber durch die zu erwartenden Datenmengen gestellt. Allein die Radioantennen des Riesenteleskops werden je Sekunde 2 Petabyte an Daten erzeugen – das entspricht dem Sechsfachen des aktuellen Internetverkehrs. Um die Daten der ersten Projektphase zu speichern, werden voraussichtlich 300 Petabyte Speicher je Jahr benötigt, das ist knapp das doppelte Datenvolumen der heutigen Dateneinsparungen aller Nutzer von Facebook. Für die gesamte Laufdauer des Experiments wird mit einem Volumen gespeicherter Rohdaten im Rahmen von Zetabyte (eine Million Petabyte) gerechnet. Um diese Datenflut zu bewältigen, haben sich die Radioastronomen nun mit den Teilchenphysikern zusammengesetzt. Mitte Juli hat die SKA Organisation eine Ver-



Zwei schwarze Löcher: Einmal werden sie sich noch umrunden, dann verschmelzen sie.

Abbildung aus Quantum Grav. (Bd. 32, Nr. 063002)

einbarung mit dem Cern zur Zusammenarbeit im Bereich der Extremskalen-Datenbearbeitung getroffen. Dabei geht es um die Entwicklung neuer Lösungen im Umgang mit Daten im Exa-Bereich (ein Exabyte bezeichnet eine Milliarde Gigabyte). Schließlich wächst auch die am Large Hadron Collider generierte Menge an Rohdaten mit jedem Experiment um ein Vielfaches. Das bisherige Rechenetzwerk sollte daher eine Art wissenschaftlicher Daten-Cloud werden, sagt Eckhard Elsen, Direktor für Forschung und Datenverarbeitung am Cern: „Das SKA wird für dieses Vorhaben ein idealer Partner sein.“

Neue Wege der Wissenschaft

Auch für die Industrie sind solche wissenschaftlichen Großprojekte interessant. Denn die Entwicklung von Echtzeitanwendungen und -analysen riesiger Datensätze etwa in der Pharmazie, dem Auto- oder dem Flugzeugbau kann durch die Arbeiten in der Grundlagenforschung getrieben und optimiert werden. Ein Vorteil ist, dass diese Daten praktisch keinerlei datenschutzrechtliche Schwierigkeiten mit sich bringen. Beispielsweise prüft SAP schon die Anwendbarkeit seiner Big-Data-Technologien für die radioastronomischen Herausforderungen des SKA. Und T-Systems wird an der Entwicklung der neuen vom Cern genutzten Forschungs-Cloud „Helix Nebula“ beteiligt sein.

Die Wissenschaft aber braucht neue Strategien. Sind doch schon heute viele der von ihr zu klärenden Fragen so komplex, dass sie mehr Daten enthalten, als dass sie ein menschlicher Wissenschaftler zeit seines Lebens auch nur anschauen könnte. Software übernimmt diese Arbeit. Ein Dilemma der Übertragung dieser Aufgabe an geeignete Algorithmen besteht aber grob gesagt darin, passende Kriterien zu finden, um die Menge der Rohdaten schon zu einem frühen Zeitpunkt ihrer Bearbeitung automatisch zu reduzieren, ohne dabei Daten zu verlieren, die möglicherweise interessant sein könnten.

Wenn man die Algorithmen zu sehr auf der Grundlage theoretischer Erwartungen konstruiert, läuft man Gefahr, nur das zu finden, was man schon gesucht hatte. Wenn man andererseits völlig datenbasiert arbeitet und allgemein nach „Mustern“ in den Daten sucht, steht man vor der Aufgabe, von bloßen Korrelationen zu wissenschaftlichen Kausalitäten zu gelangen. Beim Large Hadron Collider wird das Problem dadurch gelöst, dass die verwendeten Kriterien der Selektion der Daten teilweise so allgemein gehalten sind, dass sie auf viele theoretische Szenarien zutreffen.

Das traditionelle Bild von Wissenschaft, demgemäß man theoretische Modelle entwickelt, aus denen dann experimentell überprüfbar Vorhersagen abgeleitet werden, deren Test möglicherweise zu einer Falsifizierung des Modells führt, passt immer weniger zur heutigen „Big-Data-Wissenschaft“. Was das für die resultierenden wissenschaftlichen Ergebnisse, aber auch für die notwendigen Strukturen der Forschung bedeuten wird, ist heute noch gar nicht absehbar. Ebenso bleibt abzuwarten, ob und wenn ja welche technologischen Innovationen und Revolutionen sich im Zuge dieser neuen wissenschaftlichen Ära auch jenseits der klassischen Forschung ergeben.

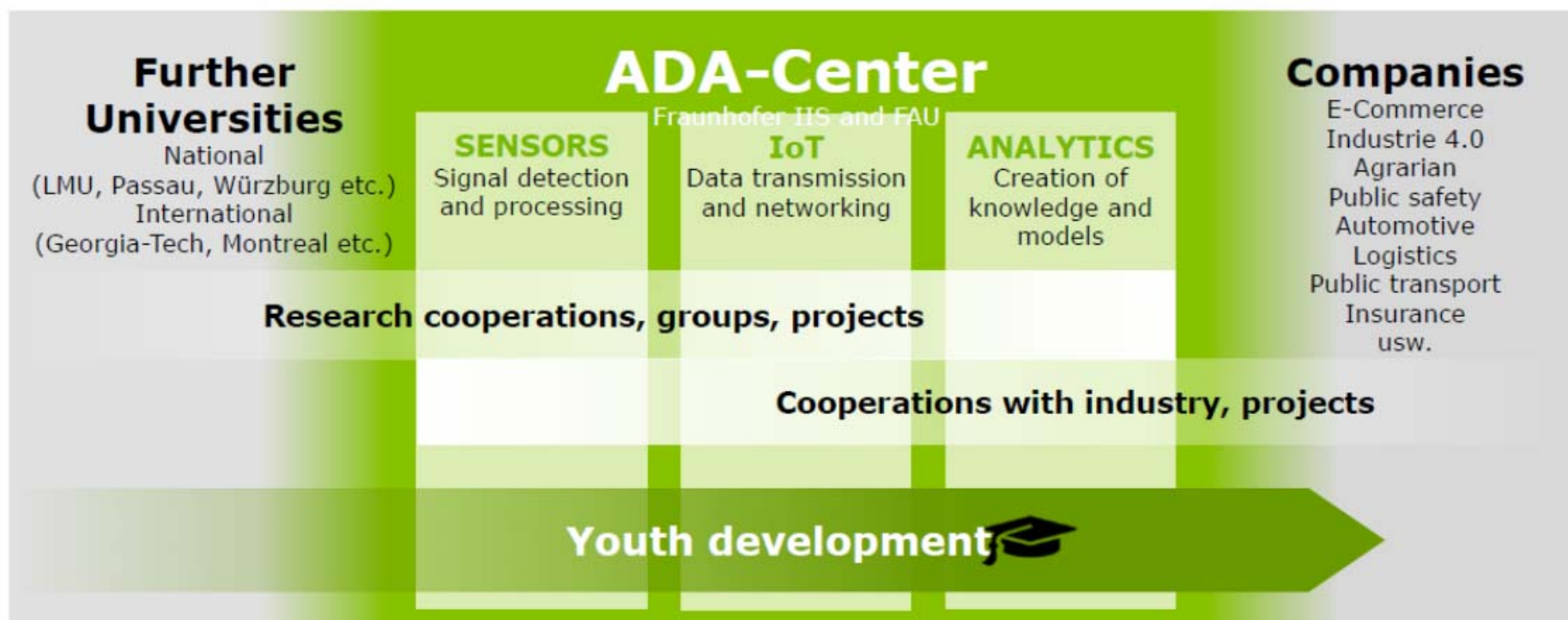
Vorbereitungsarbeiten zum Square Kilometre Array in Australien. Foto SKA Organisation

Future computing models?

- Federation of resources
- Federation across domains
- Only a few large data centre
- Many centres for compute
- Hybrid solutions
- Commercial clouds



ADA-Center Structure

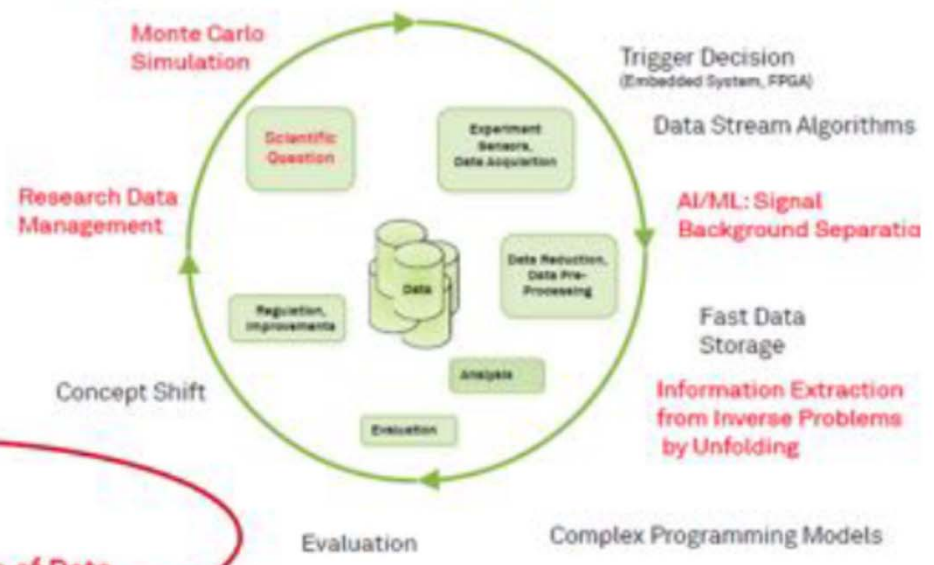


Software, Algorithms, Methods

Computer Science View

- Research Topics:**
- Embedded Systems
 - Machine Learning
 - Robustness
 - Resource Awareness
 - Data Streaming
 - Distributed Computing
 - Data Storage
 - Privacy Awareness

Attractively for CS:
 ... in ONE System
 ... Test with PetaBytes of Data



Radio 2017 - Timetable 2017

- July: MeerKAT/LOFAR Verbundforschung
 - **ToDo MeerKAT MoU → next topic**
- October/November (Nov 6th – Bologna)/December:
 - **ToDo:** Next BMBF Meeting to suggest final plan to implement SKA associate membership (prior to Nov 6th)
 - **ToDo:** Visits to SA (MeerKAT analysis meeting ...)
 - **ToDo:** RADIO2018: Science Meeting (MeerKAT/SKA/LOFAR/VLBI, Multifrequency astronomy, CMB, Data science)