GakuNin Cloud

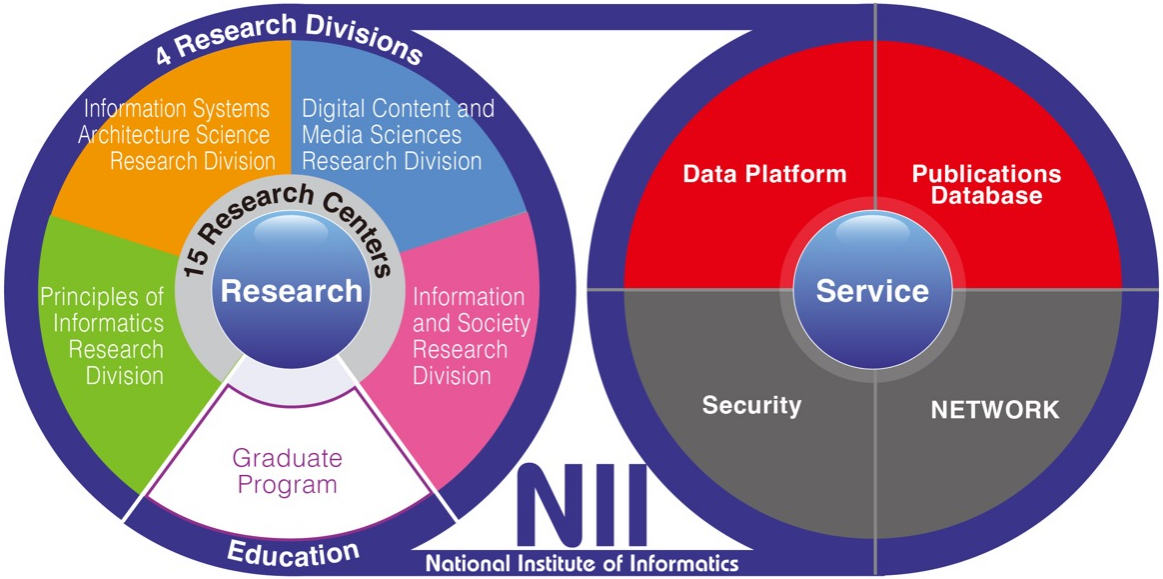
Cloud Service Adoption Support in Academic Communities

Kento Aida

Center for Cloud Research and Development
National Institute of Informatics

National Institute of Informatics

- National Institute of Informatics (NII) seeks to **advance integrated research and development activities in information-related fields**, including networking, software, and content.
- As an inter-university research institute, NII promotes the creation of **a state-of-the-art academic-information infrastructure** (the Cyber Science Infrastructure, or CSI) that is essential to research and education within the broader academic community



Supporting Academic Research Infrastructure

- NII builds and operate the **Science Information NETwork (SINET),** high speed (400Gbps), high reliability and multifunctionality network.
- Leveraging the SINET, NII provides an **authentication federation platform**, **cloud adoption and utilization support**, and **academic content platforms** as well as develops **NII Research Data Cloud** to promote open science. Through those services, NII is working to maintain and provide the **Scientific Research Digital Platform**.
- Furthermore, NII **Security Operation Collaboration Services** contribute to building the framework enabling national universities and other academic institutions to respond quickly to cyber security

https://www.nii.ac.jp/en/service/

Center for Cloud Research and Development

Cloud

NII provides universities and research institutions with information and consulting services to support the adoption of clouds, as well as tools to support the use of clouds for education and research, and the development of wide-area data collection and analysis programs.

The center aims at establishment of the national advanced R&E infrastructure taking advantage of clouds. We research and develop advanced cloud infrastructure technologies such as inter-cloud, which utilizes multiple cloud platforms in a federated manner.

- Research and development of infrastructure technologies such as cloud architecture and middleware.
- Research and development of advanced IoT (Internet of Things) system infrastructure technologies using the Mobile SINET and clouds.
- Research and development of cloud and IoT application software in cooperation with researchers in universities.
- **Support for cloud adoption and use in academic organizations**.

https://www.nii.ac.jp/en/research/centers/ccrd/

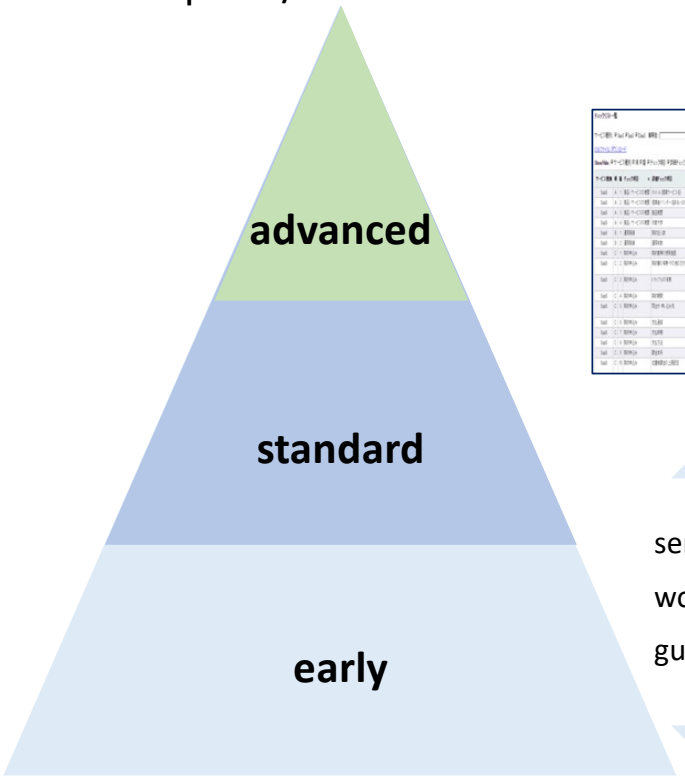
GakuNin Cloud

https://cloud.gakunin.jp/

GakuNin Cloud

Support universities to adopt/use cloud services

support universities in early/standard/advanced phases of cloud adoption/use

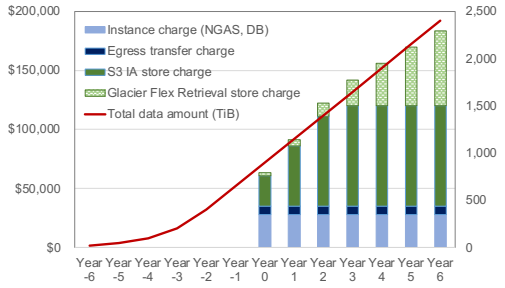


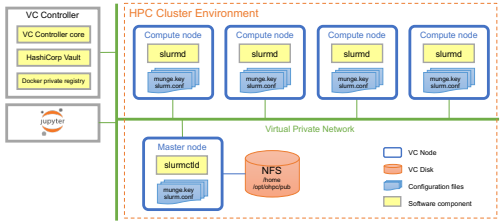
cloud service checklist

seminar/ workshop/ guideline

ondemand configuration/ gateway/ maturity survey

PoC





GakuNin Cloud Use Cases

https://cloud.gakunin.jp/

GakuNin Cloud

Support universities to adopt/use cloud services

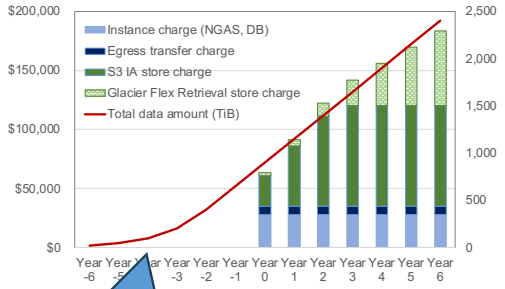
support univers d phases of cloud adoption/use

Cloud Service Checklist

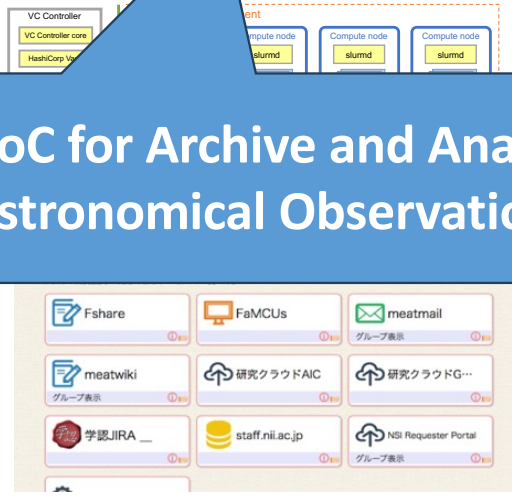
cloud service checklist

ondemand configuration/ gateway/ maturity su

PoC



PoC for Archive and Analysis of Astronomical Observation Data



advanced

standard

early

seminar/ workshop/ guideline

GakuNin Cloud

use case 1

Cloud Service Checklist

Gakunin Cloud Adoption Support Service

NII collects, disseminates, and shares standard processes and information required when universities and institutes adopt and use cloud services.

- plan cloud adoption
- develop specification and run procurement

- review checklist responses
- conduct individual consultation, etc.

- provide checklist responses
- propose products and services

- **Access to checklist responses**
- **Request for individual consultation**
- **Others (Sharing information through workshops/seminars**

cloud service checklist

- **Provide checklist responses to universities and research institutes**
- **Understand the needs of universities and research institutes**
- **Others (sharing information, attend workshops, etc.)**

Cloud Service Checklist

- The checklist comprehensively includes checkpoints, which should be considered at the time of cloud adoption of universities and research institutes.
- NII collects responses from cloud service providers and share the responses among universities/research institutes.

	Checkpoints	# of detailed checkpoints	Contents
A	Overview	4	Service name, service overview, etc.
B	Historical records	2	Number of customers, service launch date, etc.
C	Contracts	8	Payment method, license system, etc.
D	Authorization/authentication	3	Shibboleth and Gakunin support, multi-factor authentication, etc.
E	Availability/reliability	4	Service availability, planned outage, etc.
F	Customer support	5	Help desk, support turnaround time, etc.
G	Network and communication	9	SINET connection, network encryption, etc.
H	Management functions	12	Dashboards, statistics, etc.
I	Software environment	4	Supported software, software stack, proven applications, etc.
J	Scalability	5	Resource limit, limit of number of instances, etc.
K	Data center	7	Physical security, data center location, etc.
L	Security	11	Security policy, incident response, etc.
M	Data management	9	Data redundancy, log management, etc.
N	Backup	6	Backup service, restore etc.
O	Trustworthiness of provider	6	Subcontracting to third parties, personal information protection, etc.
P	Terms and conditions	6	Responsibility sharing, liability for damages, etc.
Q	Data administration	3	Rights of data usage, data sanitization, etc.
R	Data migration	4	Support for data migration, compatibility of images, etc.
S	Certification	4	Business continuity, security, etc.

Examples of Checklist Items

Reliability

SLA	✓ Is service level agreement published? • the published value
Data durability	✓ Is data durability is published? • the published value
Scheduled maintenance	✓ Is scheduled maintenance planned? • the frequency and the average down time
Notification of maintenance	✓ Is the procedure for notifying maintenance information published?

Data and log management

Data redundancy	✓ Are data redundantly stored? • methods for data redundancy
Access control (file)	✓ Can the user configure access control for files? • methods of access control
Encryption	✓ Are stored data encrypted? • methods for encryption
Log	✓ Can the user access log files (e.g. a system log, a security log, an access log) ? • a list of log files

Examples of Checklist Items (cont'd)

Data center

Data center location	✓ Is a country that hosts user data published? • the name of country
Data center location	✓ Can the user designate a country or a region of datacenter to use?

Terms and conditions

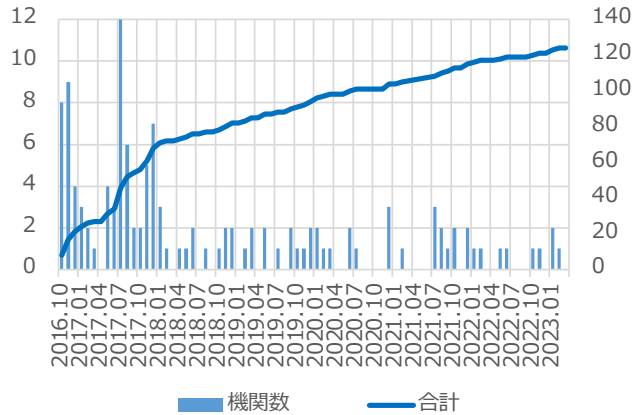
Governing law	✓ Is the contract governed by and interpreted in accordance with the laws of Japan? ✓ the name of country (if it is governed by a foreign law)
Jurisdiction	✓ Is a court with jurisdiction determined? • the name of the court

Misc.

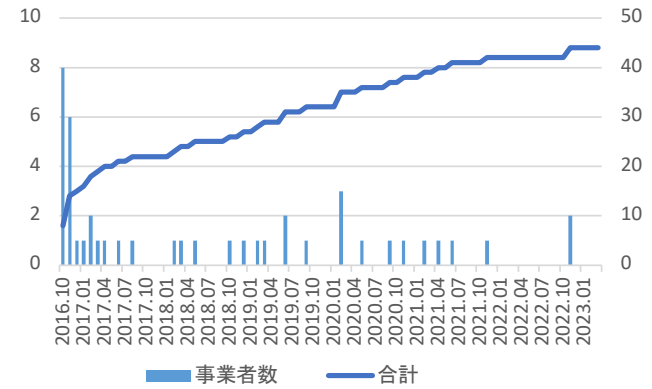
SINET connection	✓ Are DCs directly connected to SINET?
SAML	✓ Authentication via SAML? ✓ Gakunin ready?
Interoperability	✓ Does the service provide APIs that have interoperability to other cloud services?

Statistic

university members



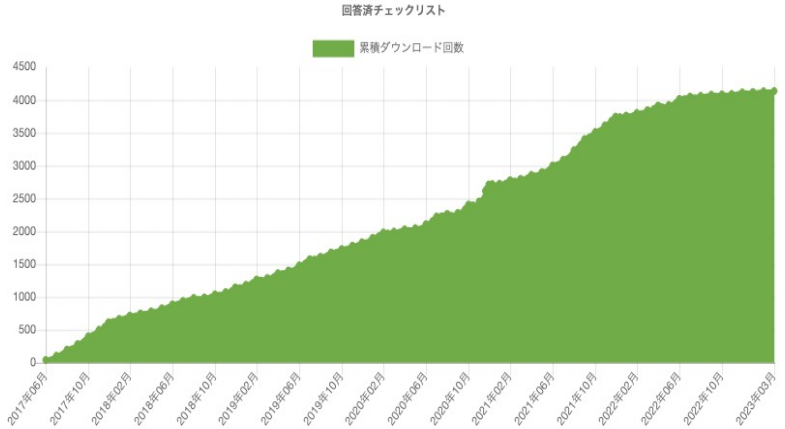
cloud provider members



service categories in checklist

Ver.	#providers	#services		
		SaaS	IaaS	others(Paas..)
5.1	24	13	18	2
4.1	5	8	1	0
3.0	2	1	2	0
sum	31	22	21	2

checklist responses downloads



Checklist for Online Meeting Services

- check items selected from the original checklist
- additional items related to functionalities of online meeting services

Category	id	Detailed checkpoints	Response type
Network and communication	G10	**End-to-end encryption (E2EE)**	Yes / No (with detailed description)
Software integration	OA1	**LTI support**	Yes / No
	OA2	**Integration with LMS (including proven cases)**	Yes / No (with detailed description)
	OA3	Integration of individual licenses	Yes / No
Online meeting functionality	OB1	Maximum number of concurrent participants	description
	OB2	Network bandwidth requirement	description
	OB3	Video recording	Yes / No (with detailed description)
	OB4	**Attendance registration**	Yes / No (with detailed description)
	OB5	**Attendance restriction**	Yes / No (with detailed description)

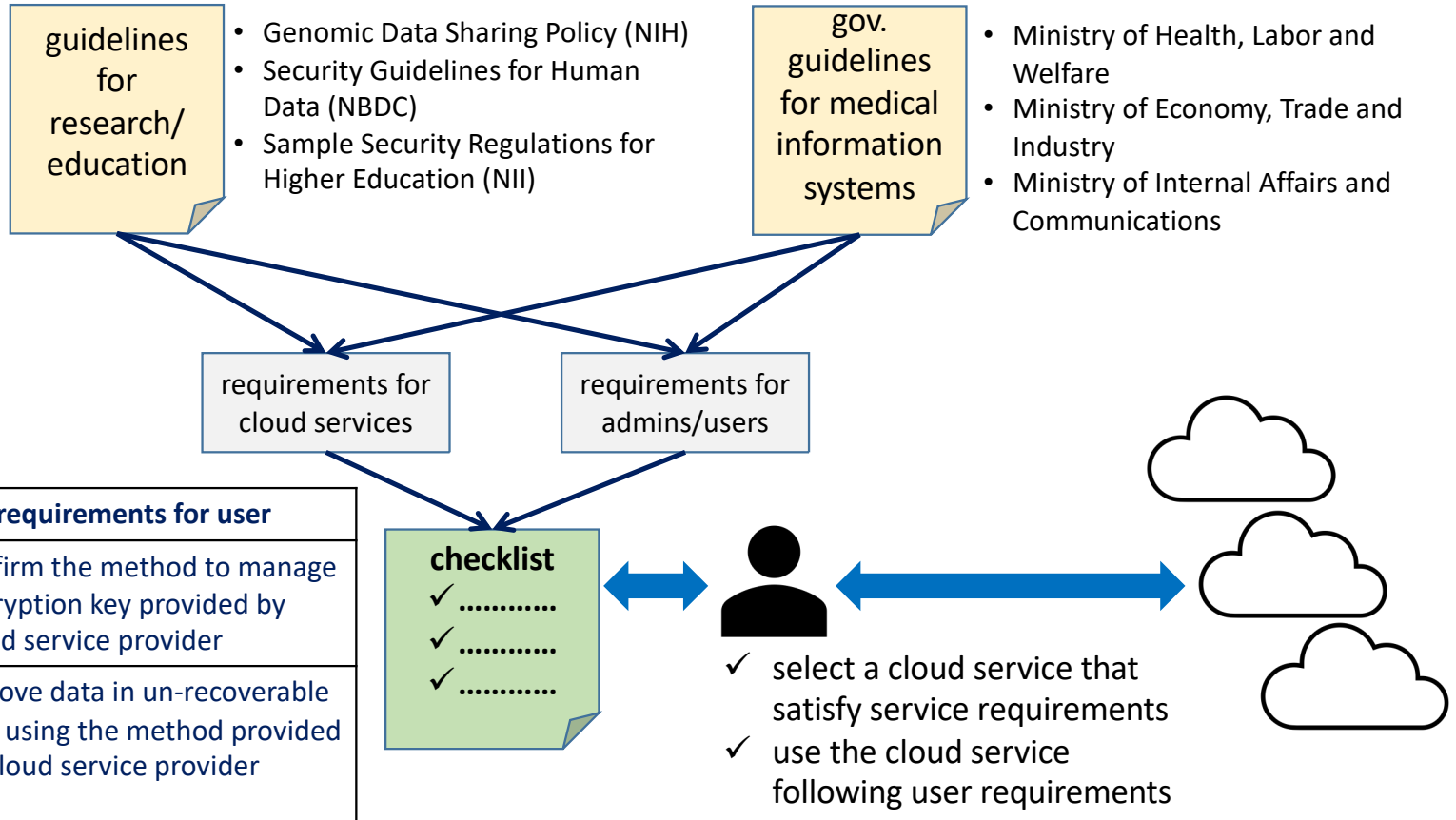
Responded services

- Amazon Chime (published in July 2020)
- Cisco Webex Meetings (published in July 2020)
- Google Meet (published in July 2020)
- Microsoft Teams (published in July)
- NTT SMART Communication & Collaboration Cloud (published in May 2020)
- Zoom (published in August 2020)

Checklist for Genome Research

new checklist for selecting/using suitable cloud services satisfying (data management) requirements

assuming store/analyze human genome data on cloud

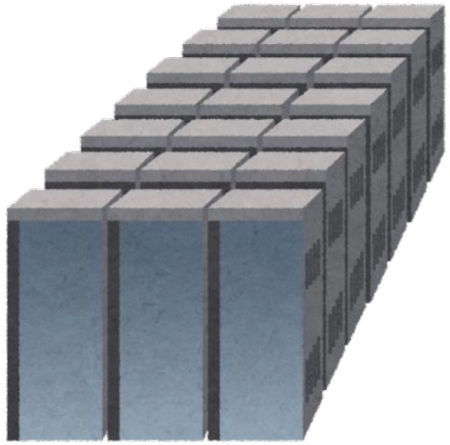


requirements for cloud service providers	requirements for user
✓ show users the method to manage encryption key	✓ confirm the method to manage encryption key provided by cloud service provider
✓ show users the method to remove data in un-recoverable way ✓ guarantee the removed data is un-recoverable (Issuing a certificate of data removal is preferred.)	✓ remove data in un-recoverable way using the method provided by cloud service provider

Japan Agency for Medical Research and Development (AMED)
Grant Number JP21km0405501

Checklist for HPC

- Check items focusing on HPC resources are added to the original checklist.
  - software license
  - **snapshots for recovering from system failure**
  - **benchmark results**
  - logs for research purpose
  - **container support**
  - **remote job submission**
  - privilege to install software
  - resource spec. (computing node, queue)
  - network (connection to the internet, VPN)



collaboration with RIKEN R-CCS

Spreadsheet to Make Spec. Document (work in progress)

- The spreadsheet helps administrators to make a specification document for procurement.
  - Texts for the specification document are generated from inputs (e.g., yes/no and additional comments) to the cloud service checklist.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	チェック項目	項番		詳細チェック項目	記入要領	回答方法	SaaS	IaaS	IDaaS	仕様書における規定	仕様書の文言	付帯条件候補	具体的な付帯条件	付帯条件付き仕様書文言	備考・コメント（なぜこう判断したか等の備忘）
17	サポート関連	F	1	サポート窓口	サポートについて、記述回答欄に以下を記入してください。サポートプラン（有償・無償など）毎に異なる場合はそれぞれについて記入してください。 ・窓口（例：メール、電話、チャット、など） ・受付時間帯（例：平日 9:00-17:00、24時間365日、など） ・回答時間（例：無償の標準プランの場合は1営業日以内、有償の〇〇プランの場合は4時間以内、など） ・対応言語（例：日本語のみ、日本語と英語、など）	記述	○	○	○	必要(付帯条件あり)	サポートについて以下の条件で行われること。			サポートについて以下の条件で行われること。電話または電子メール、あるいはそれらに準ずる電子的方法によること。24時間365日問合せが発行可能であること。24時間365日問合せが発行可能であること。発行された問合せは、平日の営業時間（例として10:00～17:30等）に対応されること。日本語で対応すること。	
18												窓口	電話または電子メール、あるいはそれらに準ずる電子的方法によること。24時間365日問合せが発行可能であること。		
19												受付時間帯	24時間365日問合せが発行可能であること。		
20												回答時間	発行された問合せは、平日の営業時間（例として10:00～17:30等）に対応されること。		
21												対応言語	日本語で対応すること。		
22												その他			
23		F	2	重要情報の通知	サービス停止、障害、保守実施、非互換を伴う仕様変更などの通知手順が定められているか「Yes/No」欄を選択してください。 「Yes」の場合、その方法（ウェブページに掲載（可能ならばURLを記入）、電子メール、契約時に書面で交付など）を記述回答欄に記入してください。	Yes / No（記述あり）	○	○	○	必要	サービス停止、障害、保守実施、非互換を伴う仕様変更などの通知手順が定められていること。	条件の詳細		サービス停止、障害、保守実施、非互換を伴う仕様変更などの通知手順が定められていること。	
24		F	3	導入時の教育プログラム	サービス導入時に大学・研究機関が教育プログラムを受けることができるか「Yes/No」欄を選択してください。 「Yes」の場合、有償・無償で内容が異なる場合はそれぞれについて記述回答欄に記入してくださ	Yes / No（記述あり）	○	○	○	必要	サービス導入時に本学が教育プログラムを受けることができること。	プログラムの内容・受講人員数など		サービス導入時に本学が教育プログラムを受けることができること。	

checklist items

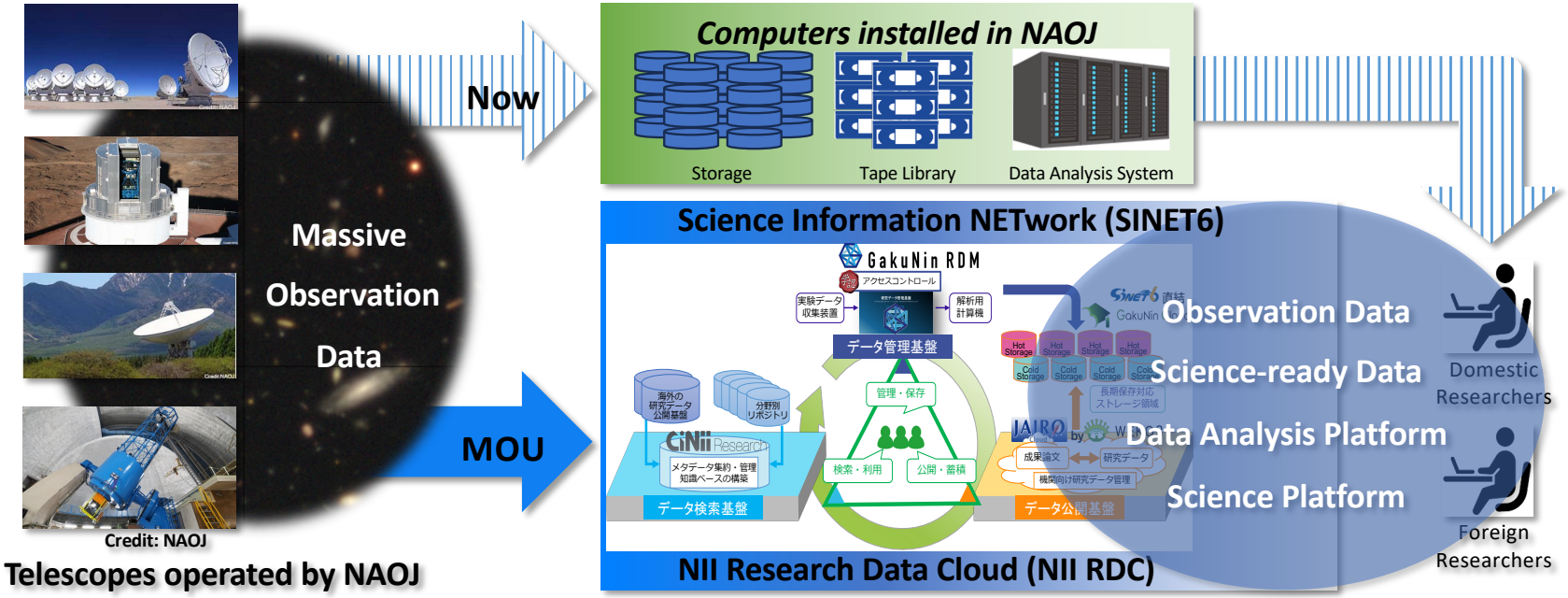
GakuNin Cloud

use case 2

PoC for Archive and Analysis of Astronomical Observation Data

Collaboration with NAOJ regarding Astronomical Observation Data on Cloud

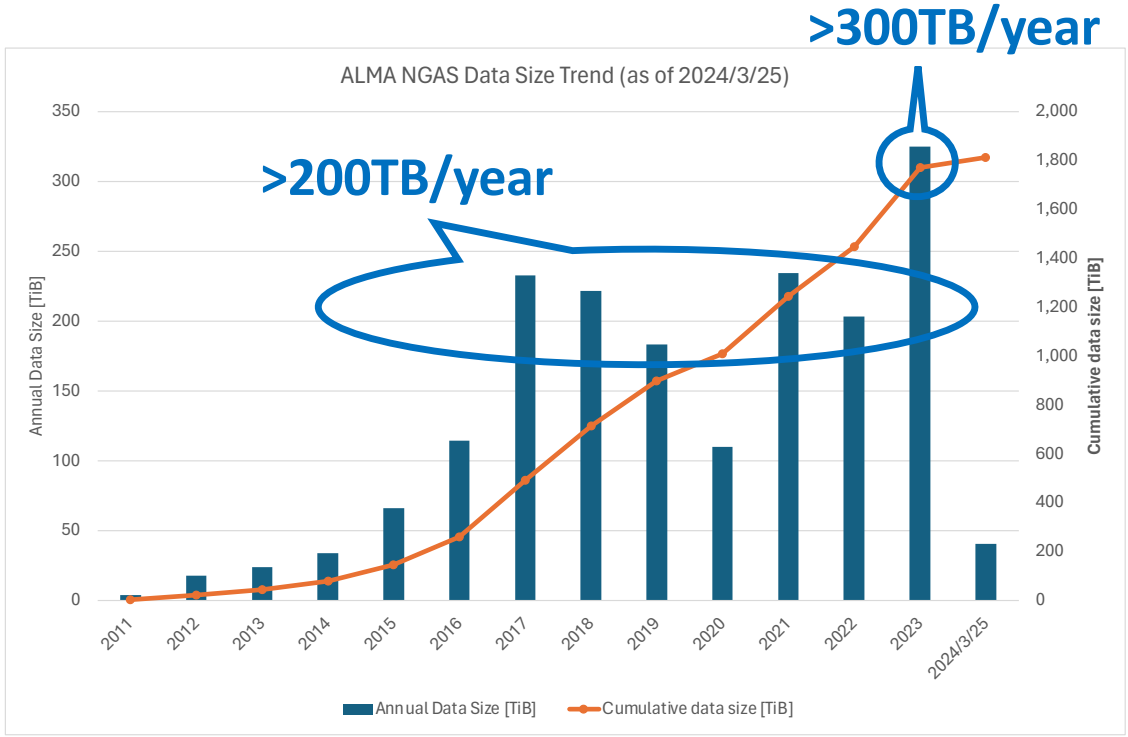
- Astronomy Data Center (National Astronomical Observatory of Japan : NAOJ) and Center for Cloud Research and Development (National Institute of Informatics : NII) concluded an engagement in long-term archiving and publishing of growing astronomical observation data, as well as data analysis… (October 18, 2023)



George Kosugi@NAOJ

ALMA

- Atacama Large Millimeter/submillimeter Array (ALMA)
  - astronomical interferometer installed in Atacama of Chile
  - 66 antennas (12m x 54, 7m x 12)



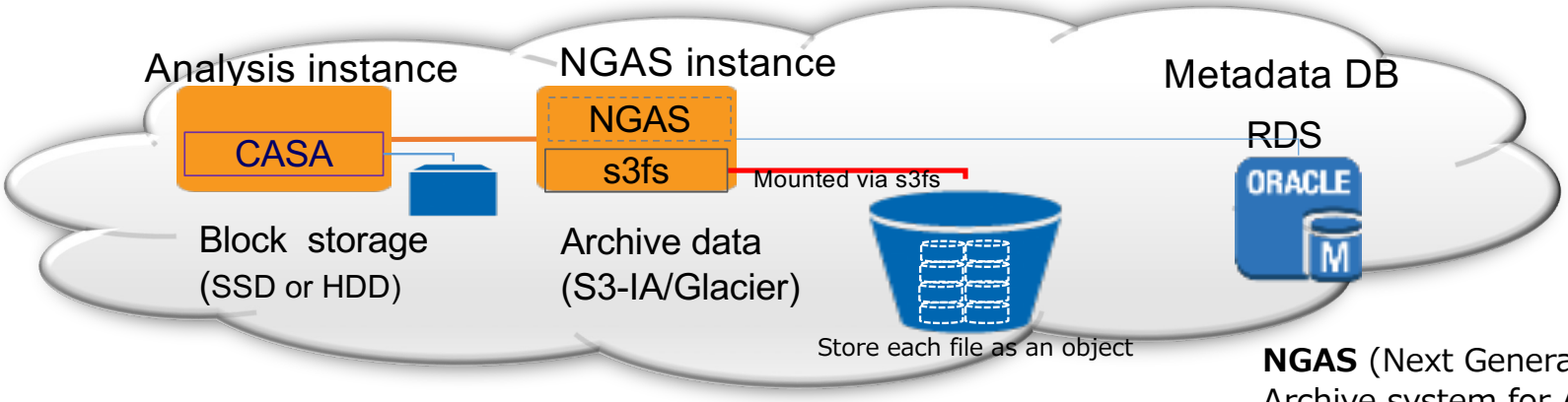
Issues

- ✓ storage system for long-term data archive
- ✓ computing system for data analysis

How can we utilize cloud service to solve the issues?

Archive and Analysis for Astronomical Data on Clouds

PoC for designing platforms for ALMA telescope data archive and analysis using **public clouds** (NAOJ and NII)



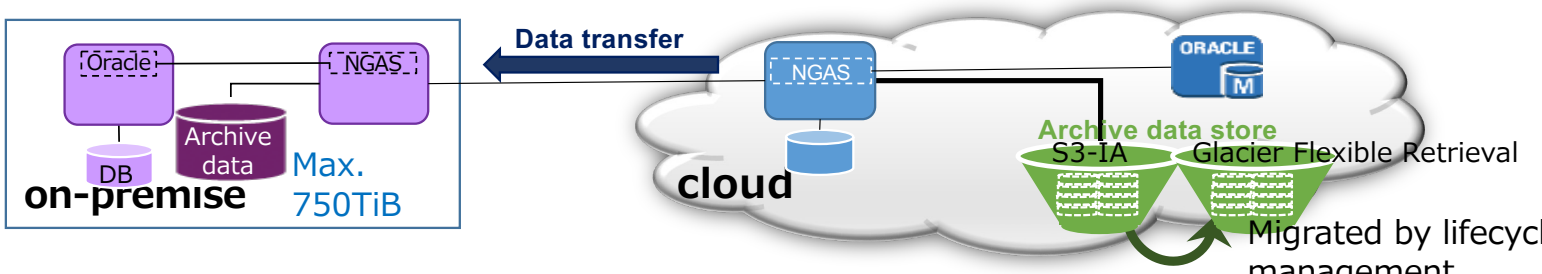
NGAS (Next Generation Archive System): Archive system for ALMA telescope data

Acquire practical information for:

- Making decisions about storing and analyzing research data in public cloud services
- Designing an overall hybrid cloud architecture

Cost Estimation of Hybrid Cloud for ALMA Archive System

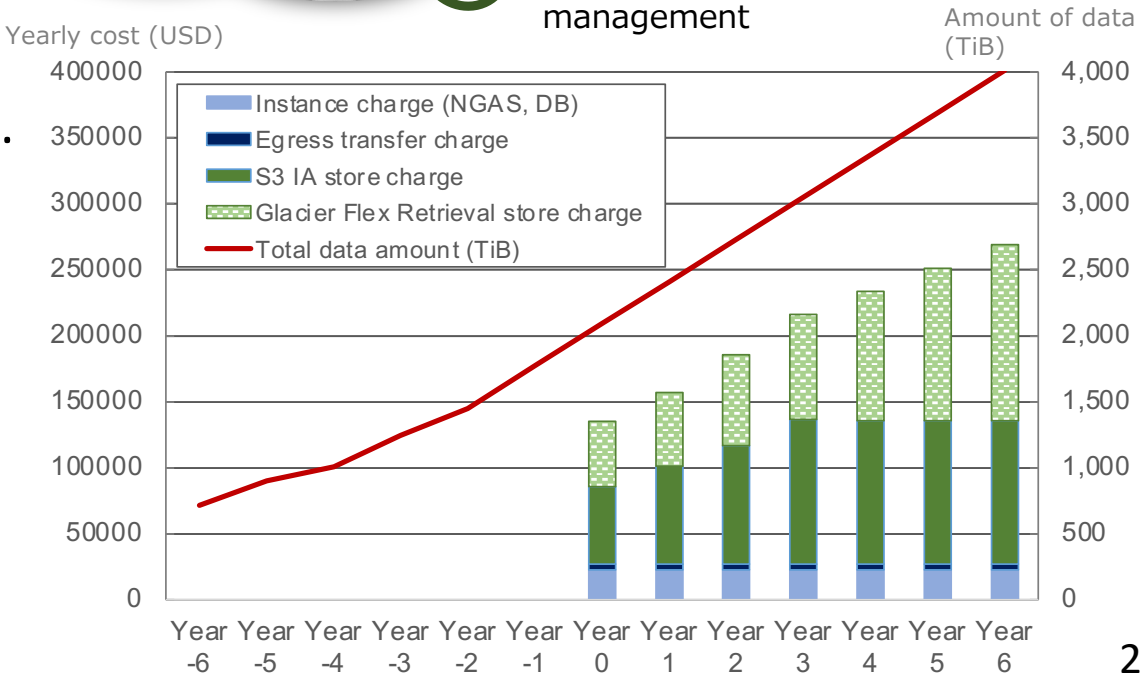
Tiered storage including Glacier, S3 IA, and on-premise storage



Tiered storage architecture restrains the extreme increase of the cloud storage charge.

[Assumptions derived from actual data usage statistics]

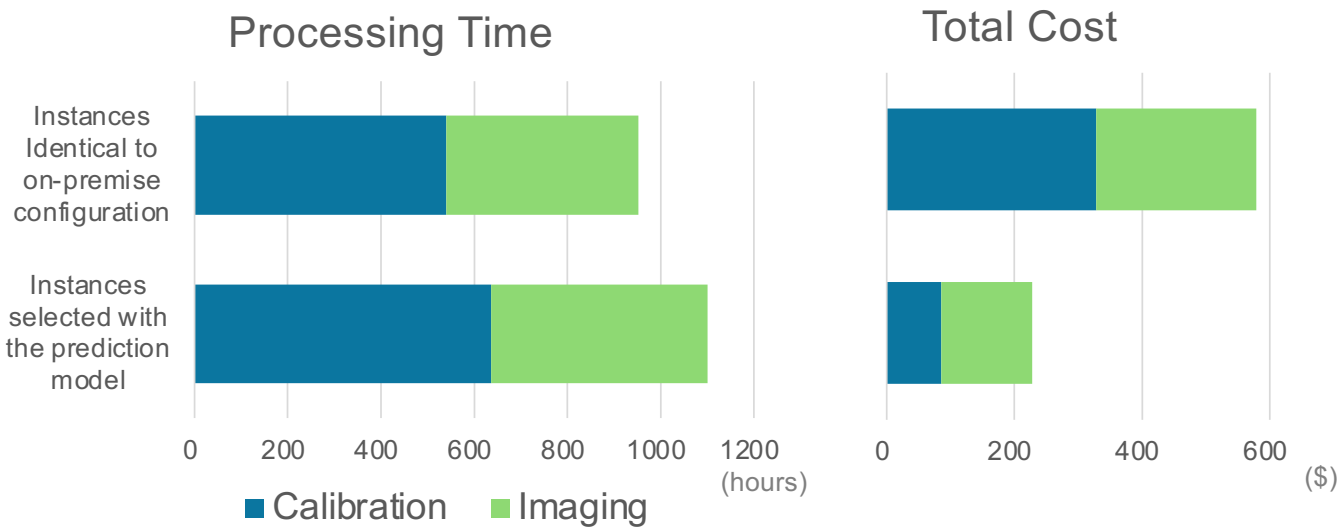
- Amount of on-premise storage: 960 TB.
- Amount of archive data:
  2,092TiB in year 0; increases 320 TB/year.
- Datasets are stored
  in on-premise storage for 3 years,
  in S3 IA for 2 years,
  and then migrated to Glacier Flexible Retrieval
- Amount of downloads is 550 TB/year;
  12% from S3 IA
  6% from Glacier Flex Retrieval



Computing Resource Optimization for Data Analysis

Estimate optimal resources (instance types) through correlation analysis between observation parameters and required resources (CPU/memory)

- The result of the analyses of 372 observation data-sets (7m ALMA antenna) on the instances chosen with our methodology showed that the processing time increased by 15%. However, the total cost decreased by 60% compared to the instances identical to on-premise servers (a simplistic “lift and shift” case).

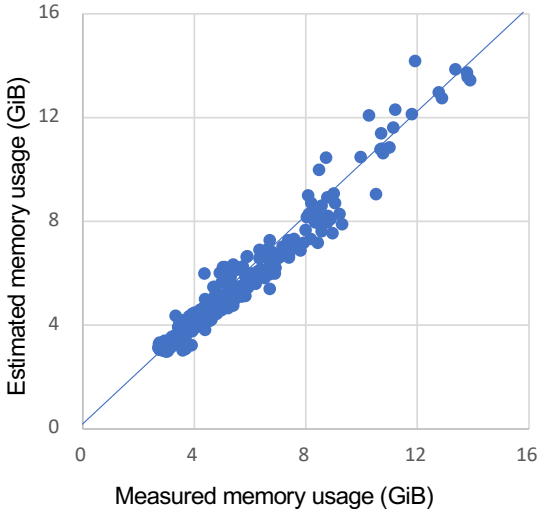


- w/ resource optimization (choose instance type with minimum cpu/memory required by applications)
  - for calibration
    - c6i.large (1core/2vcpus/4GB) $0.107/h or
    - m6i.large (1core/2vcpus/8GB) $0.124/h or
    - r6i.large (1core/2vcpsu/16GB) $0.152/h
  - for imaging
    - c6i.xlarge (2cores/4vcpus/8GB) $0.214/h or
    - c6i.2xlarge (4cores/8vcpus/16GB) $0.428/h
- w/ resource optimization (choose instance type identical to the on-premise configuration)
  - r6i.2xlarge (4cores/8vcpus/64GB) $0.608/h

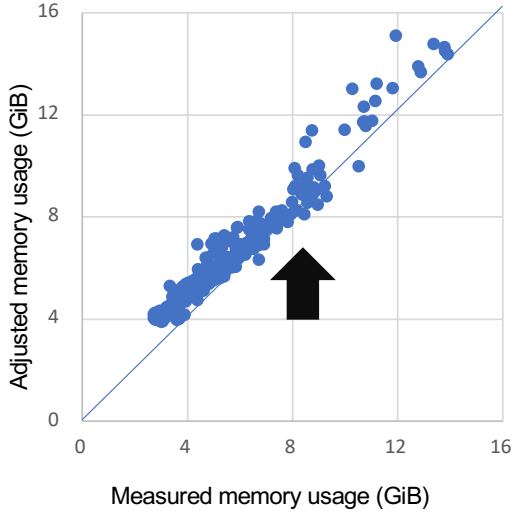
Comparison of total processing time and total cost between analyses with instances identical to on-premise configuration and analyses with estimated instances

Selection of Optimal Server Instances for Analyses

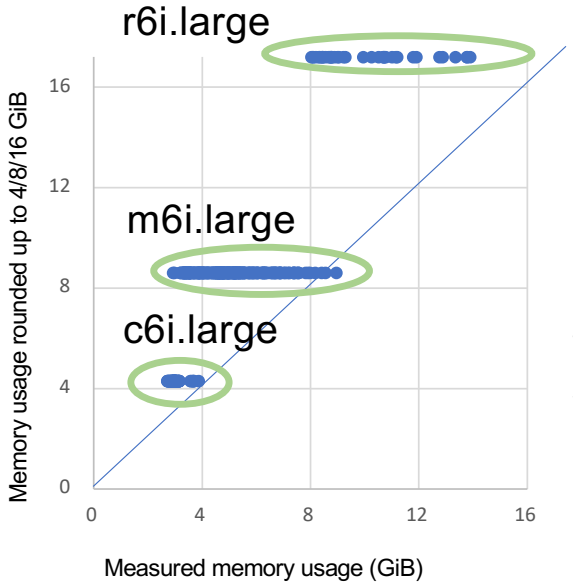
- For 7m antenna observations*, we developed a methodology to choose the cloud server instances based on the resource usage estimation (number of cores and memory capacity) according to the observation parameters.
- Example: Selection of instances considering memory usage in the calibration stages



Comparison between measured memory usage and memory usage estimated with the model



Comparison between measured memory usage and memory usage adjusted to compensate for the underestimated cases



Comparison between measured memory usage and memory usage rounded up to the cloud instance specification values (4/8/16GiB)

*ALMA telescope consists of 7m and 12m antennas. As a first step, we tried to establish a model for the observations with 7m ones because the analysis execution times are relatively short, performance is determined mainly by CPU and memory, and the effect of disk I/O is small.

International Collaboration

GÉANT NII Collaboration

Objective

- to continue to exchange information and know-how between GÉANT and NII/SINET about cloud delivery and adoption processes, building further on the activities of previous years.

Goal

- to improve the cloud activities of both organisations aimed at building advanced and highly efficient IT infrastructures in both academic communities and making cloud services safe and easy to use for the Research and Education communities from both organisations.

GÉANT NII Collaboration Work Activities

Continue to **exchange information and explore topics of cloud adoption processes and use cases**

- The topics include sharing information on vendor engagements, cloud adoption, cloud management platforms and tools, technical integrations (with cloud providers and institutions), supporting data science (or big data) applications, national and community-based clouds, real-time communication solutions, and online services for Research & Education.

To **work together to make the collaboration more visible externally**. The activities to include:

- plan shared presentations at GÉANT Cloud Forums
- Show collaboration at external cloud meetings
- plan joint efforts on the global scale.

Parties: NII/SINET staff (Kento Aida) and GÉANT (leads Maria Ristkok, David Heyns)

- Online meeting once a quarter : November, August, April in 2024
- topics from NII : cloud checklist update, academic cloud workshops …
- topics from Geant: OCRE update, eduMeet, cloud workshops, …

Conclusions

- NII promotes the creation of a state-of-the-art academic-information infrastructure that is essential to research and education within the broader academic community.

- GakuNin Cloud service provides universities and research institutions with information and consulting services to support the adoption of clouds.

- GÉANT and NII collaborate to exchange information and know-how about cloud delivery and adoption processes.

latest R&D projects https://ccrd.nii.ac.jp/sc24/