



SAIL User Forum

Sep 15th 2022



SAIL User Forum - Sep 2022

Agenda

1.¶	Welcome to SAIL user forum¶
1.1¶	Introduction and welcome — thank you for feedback received following last forum. Information on new confluence page.¶
2.¶	Talks¶
2.1¶	Talk 1: NCRN (National Cardiovascular Research Network) and associated CVD research. Julian Halcox, Libby Ellins and Daniel Harris.¶
¶	
2.2¶	Talk 2: Atrial fibrillation and stroke in older adults living in care homes: analysis of prevalence, treatment and outcomes using the SAIL Databank. Stephanie Harrison and Leona Ritchie.¶
¶	
2.3¶	Talk 3: BHF Data Science Centre CVD-COVID-UK: Cardiovascular Reproducible research and projects. Hoda Abbasizanjani and Fatemeh Torabi.¶
¶	
2.4¶	Talk 4: Myocardial Infarction and stroke subsequent to urinary tract infection (MISSOURI). Harry Ahmed and Victoria Best.¶

3.¶	Data acquisition and meta-data team¶
3.1¶	Cynthia McNerney -- Updates on new data sources / planned acquisition to SAIL.¶
4.¶	Analytical team¶
4.1¶	Hannah Davies -- Updates on SAIL analytical team services and information. ¶
4.2¶	Ian Farr -- SAIL project audits¶
4.3¶	Sarah Rees -- New WSDS (single view) going live in SAIL¶
5.¶	Technical services¶
5.1¶	Richard Hier -- Updates on SAIL technical services and developments.¶
6.¶	Marketing¶
6.1¶	Chris Roberts -- Marketing — Updates on what's been in the news.¶
7.¶	Time for questions and discussion from the SAIL community¶
8.¶	Feedback and close¶
8.1¶	Contact us if you are interested in presenting at the next/future user forum.¶

Confluence page for all SAIL User Forum details: <https://docs.hiru.swan.ac.uk/display/SUF>

SAIL User Forum - Sep 2022

1. Welcome to SAIL user forum



Some general house keeping points:



SAIL User forum's will not be recorded.



All slides will be shared after the forum via confluence.



Q&A is welcome via the chat or at designated points over video/audio.



Please mute microphones when not asking questions or presenting.

SAIL User Forum - Sep 2022

1. Welcome to SAIL user forum

What to expect from the “new” SAIL user forums moving forward...



Forum frequency & timings – bi-monthly and on maintenance windows.



Forum chairs – Ashley Akbari & Chris Orton.



Forum structure & content – standing agenda, input from community.



Forum contact – questions/suggestions sailbusinessadmin@chi.swan.ac.uk

SAIL User Forum - Sep 2022

2. Talks

Cardiovascular research

Four talks (*15mins each*):

Talk 1: NCRN (National Cardiovascular Research Network) and associated CVD research.

Julian Halcox, Libby Ellins and Daniel Harris.

Talk 2: Atrial fibrillation and stroke in older adults living in care homes: analysis of prevalence, treatment and outcomes using the SAIL Databank.

Stephanie Harrison and Leona Ritchie.

Talk 3: BHF Data Science Centre CVD-COVID-UK: Cardiovascular Reproducible research and projects.

Hoda Abbasizanjani and Fatemeh Torabi.

Talk 4: Myocardial Infarction and stroke subsequent to urinary tract infection (MISSOURI).

Harry Ahmed and Victoria Best.



SAIL User Forum - Sep 2022

3. Governance and Legal team



Cynthia Mcnerney (*Governance & Legal team*)

- Updates on new data sources / planned acquisition to SAIL.

SAIL User Forum - Sep 2022

4. Analytical team



Hannah Davies, Ian Farr and Sarah Rees (*Analytical team*)

- Updates on SAIL analytical team services and information.
- SAIL project audits.
- New WDSD (single view) going live in SAIL.

SAIL Project Audits

Ian Farr - SAIL Analytical Services

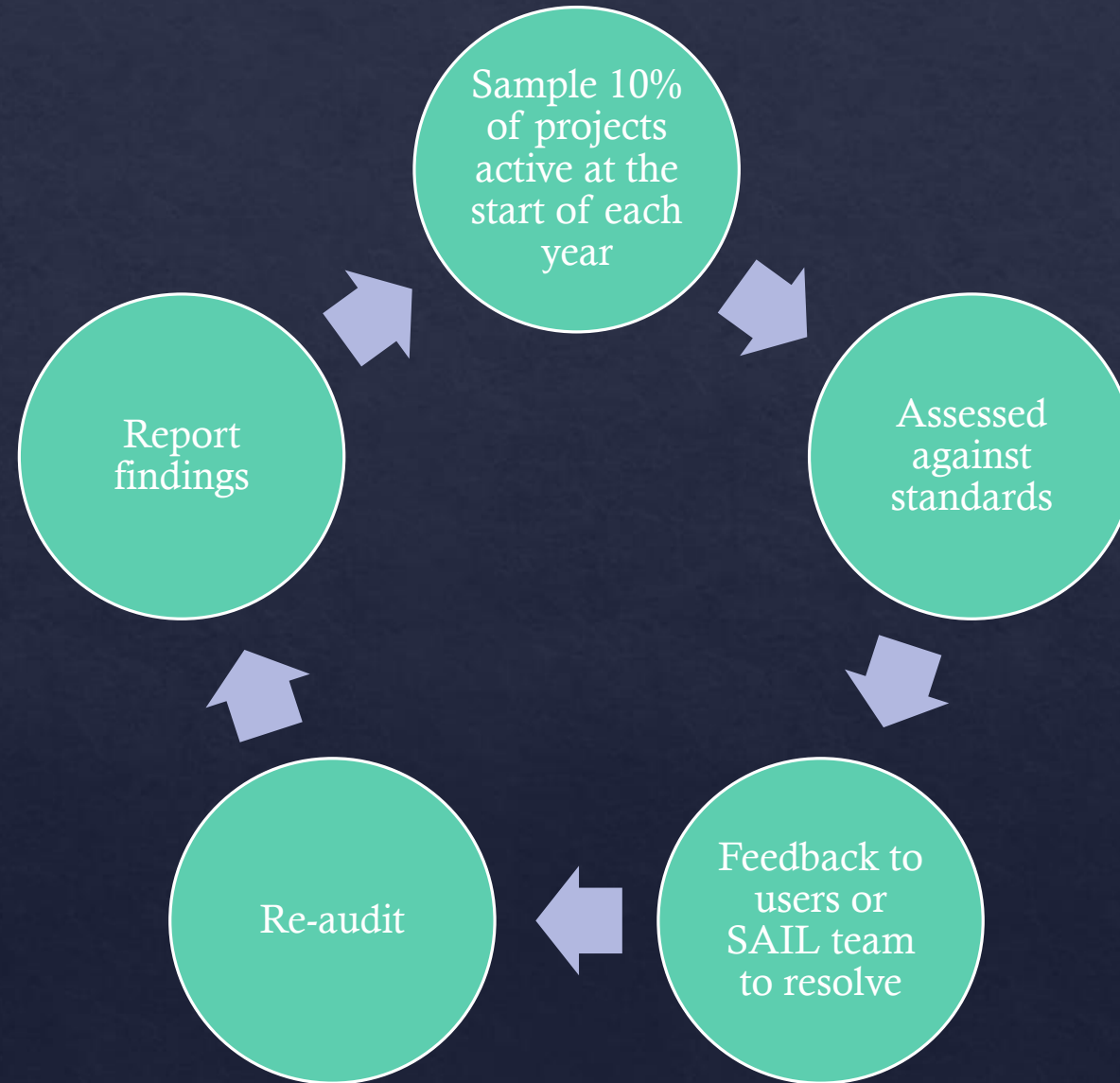
Contents

- ◇ **Background** – what is the need for the audits
- ◇ **Process** – how the audits are conducted
- ◇ **Standards** – what things are assessed
- ◇ **Communication** – How researchers are informed
- ◇ **Support** – where to go to get more information

Background

- ◆ **Not doing it for fun!**
- ◆ **Not doing it to upset people or projects!**
- ◆ Requirement of SAIL Databank information governance accreditations
- ◆ Ensure research is conducted ethically, safely, legally and in compliance with the protocol, conditions of Information Governance Review Panel (IGRP) approval and SAIL policies and procedures
- ◆ Raise awareness of requirements and promote researchers' accountability
- ◆ This means that we can continue to offer over 70 datasets for research purposes

Process



Standards

- ◆ Relate to SAIL policies around:
 - ◆ Data access & permissions
 - ◆ Governance & documentation
 - ◆ Data provisioning
 - ◆ Files requested out
 - ◆ File storage
 - ◆ Publication
- ◆ These relate to researchers
and SAIL Databank

Have only users from organizations mentioned been given access?
If ethical approval is needed, have we obtained evidence of that?
If the project requires permission for use of datasets, do we have evidence that this has been obtained?
Permissions equal or within IGRP dates
Is the end date equal to the end date on IGRP (or latest extension/amendment)?
Have users been granted access to project views only?
Do the project views match what is described in the IGRP form?
If the project is bringing in new data, does the data match what was requested on the IGRP form?
Does the data have any items that are too granular or possibly identifiable?
Are the approved file out requests all within the scope of the analysis plan?
Ensure that there is no reason for concern that the file out process may have been bypassed.
(Project active for a long time, working a lot on the data, but not requesting results out.
Publication that does not seem supported by requests).
Are there any specific restrictions listed here in what will be done/published, and if so, was this followed?
Search code files for inappropriate access (direct access to SAIL datasets, other projects, etc.).
Randomly inspect some files from the project folder/subversion, is code and data are within scope of IGRP?
Search for publications (if project is mature enough).
Does it look like users are not working in their project area? (Not very many files in S drive, SVN, Git)?

Communication

- ◇ Usually just need clarification or awareness.
- ◇ We noticed some common issues due to lack of awareness of our requirements and best practice. So we have updated our essential resources
 - ◇ Do's and Don'ts Youtube- <https://youtu.be/ROGanZXbU8c>



Essential Resources

Guide for new SAIL users and Getting Started.

Efficient SQL for well-behaved, effective queries.

SAIL Output User Guide for guidance on acceptable SAIL outputs. [Make a output request here.](#)

Concept Library for shared definitions to use in your research.

Support

- ◇ [Helpdesk - https://help.saildatabank.com/](https://help.saildatabank.com/)
- ◇ [How to access SAIL services - https://docs.hiru.swan.ac.uk/display/SATP/How+to+access+SAIL+services](https://docs.hiru.swan.ac.uk/display/SATP/How+to+access+SAIL+services)
- ◇ [SAIL website - https://saildatabank.com/contact/](https://saildatabank.com/contact/)
- ◇ i.w.farr@Swansea.ac.uk

Thank you for your attention 😊

Welsh Demographic Service Dataset (WDSD) – single table



Key points

- DHCW introduced new, improved WDS – three-table version no longer maintained
- No longer need to join tables – all-in-one
- Person ID no longer required
- New cleaning rules – no overlaps between GP registrations (SAIL tables already do this)
- Change of some field names and some fields discontinued
- ‘Active To’ date for open registrations null rather than 9999-01-01
- Scripts based on cleaned tables – field names and structures are the same
- Scripts based on uncleaned tables (AR_PERS, AR_PERS_GP, AR_PERS_ADD) will need some amending

Old WDS	
COLNAME	TYPENAME
ALF_E	BIGINT
WOB	DATE
DOD	DATE
GNDR_CD	CHARACTER
RALF_E	BIGINT
LSOA2001_CD	VARCHAR
LSOA2011_CD	VARCHAR
PRAC_CD_E	BIGINT
FROM_DT	DATE
TO_DT	DATE
AVAIL_FROM_DT	DATE
UPRN_QAS_MATCH_CD	VARCHAR
ROW_STS	CHARACTER
RALF_STS_CD	CHARACTER

New WDS	
COLNAME	TYPENAME
ALF_E	BIGINT
WOB	DATE
DEATH_DT	DATE
GNDR_CD	CHARACTER
RALF_E	BIGINT
LSOA2001_CD	VARCHAR
LSOA2011_CD	VARCHAR
PRAC_CD_E	BIGINT
ACTIVEFROM	TIMESTAMP
ACTIVETO	TIMESTAMP
AVAIL_FROM_DT	DATE

Note the change in keywords:

DOD -> DEATH_DT

FROM_DT -> ACTIVEFROM

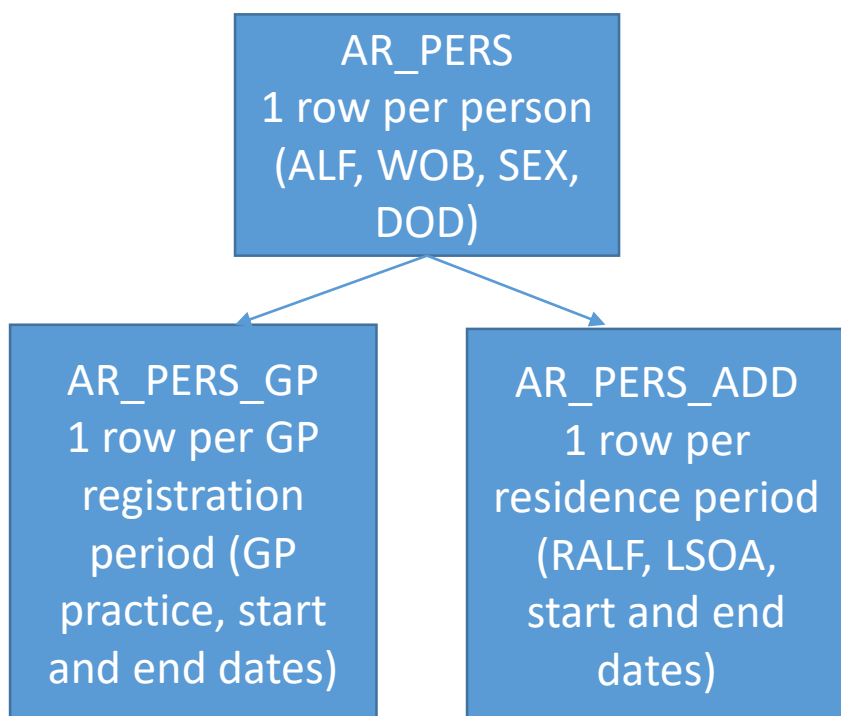
TO_DT -> ACTIVETO

◀ Note changes in data type:

FROM_DT (DATE) -> ACTIVEFROM (TIMESTAMP)

TO_DT (DATE) -> ACTIVETO (TIMESTAMP)

Compare old and new table structures



ALF_E	WOB	DEATH_DT	GNDR_CD	LSOA	PRAC_CD_E	ACTIVE_FROM	ACTIVE_TO
1111111111	01/01/2000	[null]	1	W1	333	01/01/2000	31/12/2002
1111111111	01/01/2000	[null]	1	W2	333	01/01/2003	31/12/2003
1111111111	01/01/2000	[null]	1	W2	444	01/01/2004	31/12/2004
1111111111	01/01/2000	[null]	1	W2	555	01/01/2005	31/12/2005
2222222222	02/02/2002	02/02/2022	1	[null]	666	02/02/2002	31/12/2002
2222222222	02/02/2002	02/02/2022	1	W3	666	01/01/2003	31/12/2003
2222222222	02/02/2002	02/02/2022	1	W4	666	01/01/2004	02/02/2022
3333333333	03/03/2003	[null]	2	W5	777	03/03/2003	[null]
4444444444	04/04/2004	[null]	1	W6	888	04/04/2004	[null]
5555555555	05/05/2005	[null]	2	W7	999	05/05/2005	[null]

(not real data, obv!)



Provisioning

- Sept end onwards
 - Old version phased out – time TBC
 - Scheduled refreshes will be new version
 - Request via amendment process
 - Projects will continue to receive:
 - Uncleaned WDS table (new single-table version)
- Cleaned tables produced by SAIL:
- Cleaned geography-based tables (RALF/LSOA/Wales – depending on IGRP)
 - Cleaned GP coverage tables (GP data flag, GP practice)
 - New row-per-person table to replace AR_PERS

Metadata and data quality checks

- Metadata for new single view table, and cleaned tables <https://www.healthdatagateway.org/>
- DQ checks are being run on new version in parallel with three-table version:
 - Check new version
 - Compare with old version

PER_RESIDENCE_GPREG		✕
New single-view WDS table from Digital Health and Care Wales (DHCW), which supersedes the three-view WDS tables.		
Variables		Expand all
WSPERSONID_PE	Person identifier [encrypted] within the GP practice. The single view WDS person ID system...	▼
ALF_E	Encrypted Anonymised Linking Field (ALF). This field is derived from by double encryption ...	▼
WOB	The date of the Monday that occurs prior, or on, the actual date of birth. Access to this ...	▼
DEATH_DT	The date [if known] on which a person died or is officially deemed to have died.	▼
GNDR_CD	This is the sex (gender) of person, employee or patient. Can potentially assist in correct...	▼
RALF_PE	This is the encrypted residence identifier [Residence Anonymised Linking Field], derived f...	▼
LSOA2001_CD	The ***2001*** code for the Local Super Output Area [LSOA] of residence; this is a code cr...	▼
LSOA2011_CD	The ***2011*** code [different from 2001] for the Local Super Output Area [LSOA] of reside...	▼
PRAC_CD_PE	This is the encrypted code of the patient's registered General Practitioner (GP) Practice...	▼
ACTIVEFROM	Start of period at residence and GP practice.	▼
ACTIVETO	End of period of residence and GP practice. 9999-01-01 or NULL if current or end date not ...	▼
AVAIL_FROM_DT	Date when the data made available i.e. date of loading.	▼

SAIL User Forum - Sep 2022

5. Technical services



Richard Hier (*Technical services team*)

- Updates on SAIL technical services and developments.

SAIL User Forum - Sep 2022

6. SAIL Marketing & Comms



Chris Roberts (*Marketing & comms team*)

- Marketing – Updates on what's been in the news.

SAIL User Forum - Sep 2022

7. Time for questions and discussion from the SAIL community



Anyone at the meeting today – any questions?



SAIL User Forum - Sep 2022

8. Feedback



We will request feedback after each forum, we are keen to hear how we can improve the forums moving forward and tailor them to be more impactful & useful to the SAIL community.

Please contact the SAIL email with any feedback or requests for changes please.

SAIL User Forum - Sep 2022

8. Feedback



Are you interested in presenting at a future SAIL user forum?

Do you have any requests for areas of research / developments you would like to hear about at the next forum or in the future?

Please contact SAIL by **Oct 16th**

sailbusinessadmin@chi.swan.ac.uk

SAIL User Forum - Sep 2022

8. Feedback

Confluence page for all SAIL User Forum details: <https://docs.hiru.swan.ac.uk/display/SUF>

SAIL User Forums for 2022 will be on *(other SAIL/SeRP maintenance window dates on the website):*

Thursdays (10:00 - 12:00)

- 17th November 2022

All the dates are available on confluence, and downloadable calendar.



Any further questions...

Thank you for coming



SAIL User Forum - Sep 2022

Questions from the forum

- Cynthia McNerney update: Updates on new data sources / planned acquisition to SAIL.
- Updates to data sources have continued over the past months, and new acquisitions are ongoing.
- WCSU (Welsh Cancer) data has now been updated and is up to date until the end of 2019.
- SBSW, SBTW and SCSW (Bowel, Cervical and breast screening) data have been updated in SAIL.
- CAFE and CAFW (CAFCASS England and Cymru) data have been updated in SAIL.
- SAIL has recently acquired FMAN (family man) data from the ministry of justice. (We would be able to enable linkage between CAFCASS England data and family man data for the first time).
- SAIL also have other acquisitions happening with the Data First, agreement for Prison and Probation data has been signed off, just waiting for update on the Meta data and to agree a date of transfer. So Prison and Probation data is coming soon. Working on Crowned and Magistrate court data, hopeful that all these data would be by the autumn which is good news.
- Lots of work going on with Office for National Statistics (ONS). ONS at the moment is putting an over arching agreement to make available data into SAIL that are in Secure Research Service (SRS). The priority at the moment is putting in an agreement to get the COVID Infection survey data, population survey among others. Also working with ADR Wales (Faye Gracie) leading discussion with the census team at ONS about the acquisition of census 2021 data.

Anybody who wants to know about data and when new datasets are coming in, information about this goes out in the SAIL news letter, updated at this forum bi-monthly, and over the SAIL website in real-time.

SAIL User Forum - Sep 2022

Questions from the forum

- Richard Hier update: Technical team.
- There have been some questions recently around some software changes that we made for the template for the desktop. This is a process that SAIL / SERP are going through, and as there are more users the system needs to have a better communication, so in the next few weeks we are going to start communicating when we are going to be making changes to the desktops and what those changes are.
- We would be doing this in advance and then we can come back to this group if we have to make any decisions around any versions of application and maybe run a poll for people to give a response on.
- A list of upcoming, planned and implemented changes will be found on confluence.

SAIL User Forum - Sep 2022

Questions from the forum

- General Q&A.
- Q. Is there a good way of communicating with other SAIL users and who is using software – such as R/Python users?
- A. (Multiple people) The wiki inside of SAIL contains details, and there is also communication channels outside of SAIL such as on Slack, Teams etc. which may be useful for people to join. If there are technical issues or questions these should be directed to the helpdesk.
- Q. Our project has been waiting for enhanced computing power for a while now (and have the funds for this). Please could you advise what steps need to be taken for us to get increased power (to stop SAIL crashing when we run queries in larger datasets)? Either step's we need to take or SAIL needs to take.
- A. (Richard Hier) We have had some challenges with people being on leave over the summer. But the team is back now and we would have some positive updates to give at the next SAIL User forum. For further support please log a help desk queries and they would get back immediately on any specific queries.

SAIL User Forum - Sep 2022

Questions from the forum

- General Q&A.
- Q. Using Gitlab for project work, but when a project ends the access is lost. What is the purpose of Gitlab if you shut it down when the project ends? Thought it was a continuous repository.
- A. (Richard Hier) GitLab group is associated with the SAIL project and the users/permissions/access with that SAIL project, so that is why it is linked to the project duration.
- Q. There are various GitLab repositories that have R packages and other information, what are these and are they maintained/updated? And how do we know what is available and when they are updated/released?
- A. (Hannah Davies) The R Packages on the R'drive and in GitLab are the ones that we design to help researchers, we are currently designing some more so we can probably look and just double check to see if they are still available and valid, I believe they are which is why they are still there, but we would put it in as a task to check. I believe that we have a new package for new users, an introductory pack that will include elements of the R'Packages. For anything that is new, it will either come to this forum, or we will send out communication on it.

SAIL User Forum - Sep 2022

Questions from the forum

- General Q&A.
- Q. On the topic of package management, I usually install my own packages and keep an eye on the version being used, then activate the environment and run the analysis, sometimes it does not allow access to the environments. Wanted to know if there has been a change in how that package management (PIP) is been operated in the background.
- A. (Alex Lee) The issue with PIP is to do with HEP proxy variable, Please put a ticket in and instructions can be given on how to bypass it (you need to manually reset).
- Q. It will be useful for the GitLab in SAIL to have a space that people can share between users/projects?
- A. (Hannah Davies) There will be developments coming soon around this. We are looking into safe and secure methods for sharing between all users in SAIL (as a replacement to the 0000 S drive folder), along with options to share between SAIL project groups, more info coming soon but will likely be by request via helpdesk to set up and initiate permission. Hopefully that should be coming in the next few weeks. It will be like a subproject, like a project space within Git where you who has requested it become the owner and basically responsible for what is then contained within it, and you can decide who you want to share it with. (They would get read access).
- Q. Once I have requested the repository to be share with any user, would I still be able to push changes to that repository?
- A. (Hannah Davies) Yes, you would be the rewriter of that, still need to test out the process internally before we roll it out.

SAIL User Forum - Sep 2022

Questions from the forum

- General Q&A.
- Q. Could SAIL offer inductions for new starters?
- A. (Hannah Davies) We have a new user guide in Progress – will announce this at a future SAIL user forum.
- Further update from Richard Hier on behalf of the technical team
- We are in the process of backing up data, this is a request to all SAIL users if they could review and delete any tables that are no longer required as part of their project W schemas (or S drive folders), to reduce the back up.

Funded by



Ymchwil Iechyd
a Gofal Cymru
Health and Care
Research Wales



Certified by



IS632731

Supported by



In partnership with



Accredited by



www.SAILDatabank.com

Powered by

