

Study URI

Kerstin Forsberg and Daniel Goude, Science and Enabling Units IT, Gothenburg

PhUSE Connect Europe, 2018, Emerging Trends and Technologies



Agenda

1

Why do we need a Study URI?

2

What is a Study URI? (clue I = Identifier)

3

What a Study URI is good for!

4

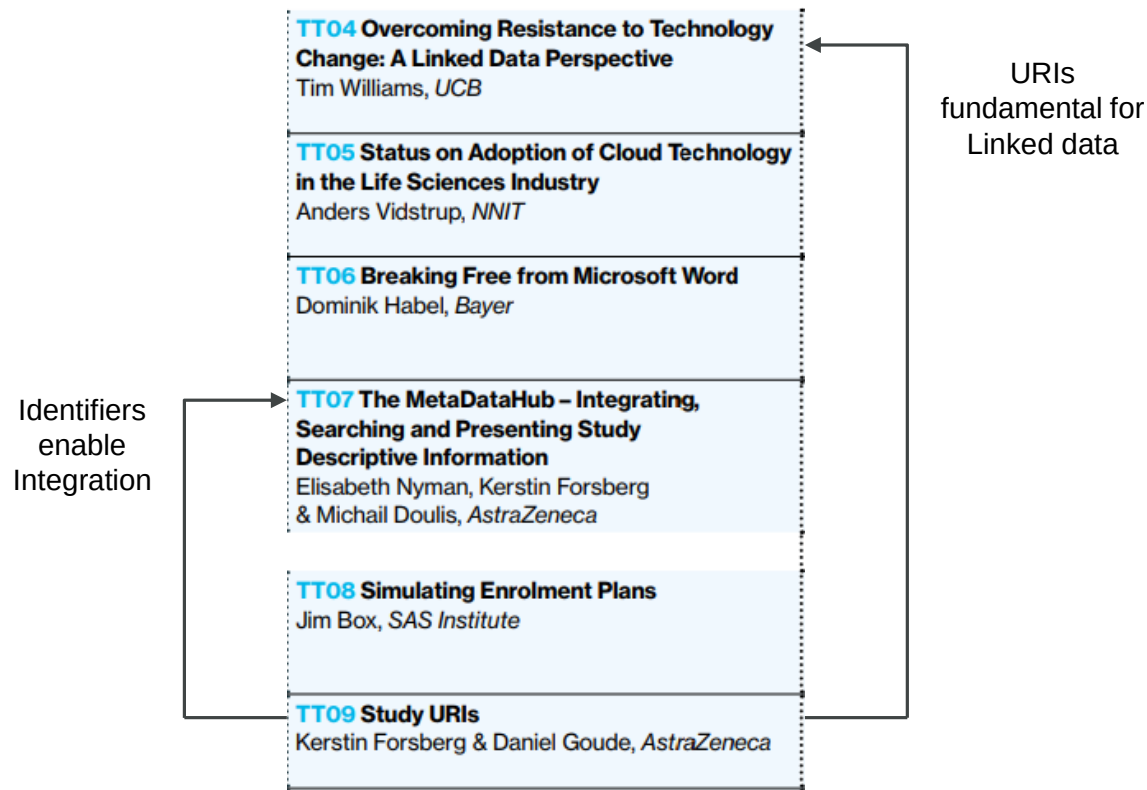
Challenges when assigning URIs for 20.000+ studies

5

Ideas for next steps / Fundamental enablers



Our presentation in relation to previous in this stream



Lifecycle of medicines

We are one of only a handful of companies to span the **entire life-cycle** of a medicine from **research and development** to **manufacturing and supply**, and the **global commercialisation** of primary care and speciality care medicines.



Lifecycle of medicines



Changing pharma landscape

- **Reuse and transparency**
- **Collaboration projects**
- **Mergers and Acquisitions**
- **Divestments**



Clinical studies are key entities/resources with long lives



Broad scope, including “all” studies (20.000+)



Clinical studies have long life cycles



Rosuvastatin (CRESTOR®) was developed by Shionogi under the code name **S-4522** at the end of the 1980s.

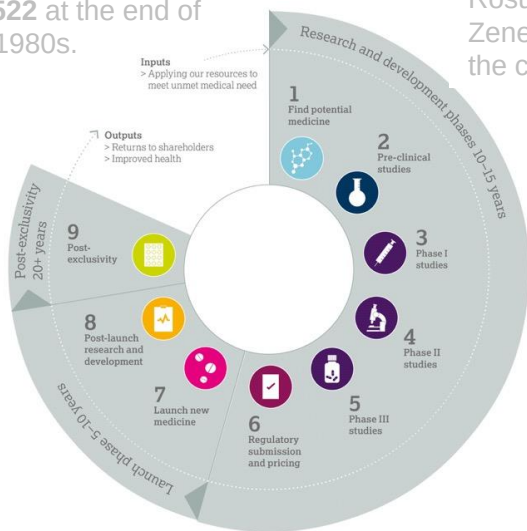
Zeneca was formed in June 1993 from Imperial Chemical Industries (ICI).

Rosuvastatin acquired by Zeneca in June 1998 under the code name of **ZD4522**.

In 1999 British Zeneca Swedish Astra AB merged to form AstraZeneca.

The code name for Rosuvastatin in AstraZeneca became **AZD4522**.

The **AURORA** study (phase III), together with **CORONA** and **JUPITER**, were outcome trials conducted between 2003 and 2008 as part of the GALAXY program



In 2015 **AURORA** was one of many studies of interest in a collaboration project with the Montreal Heart Initiative (MHI)



Clinical studies have many synonyms



Rosuvastatin (CRESTOR®) was developed by Shionogi under the code name **S-4522** at the end of the 1980s.

Zeneca was formed in June 1993 from Imperial Chemical Industries (ICI).

Rosuvastatin acquired by Zeneca in June 1998 under the code name of **ZD4522**.

In 1999 British Zeneca Swedish Astra AB merged to form AstraZeneca.

ZENECA (ICI)
4522IL/0096

The code name for Rosuvastatin in AstraZeneca became **AZD4522**.

AstraZeneca
D3562C00096

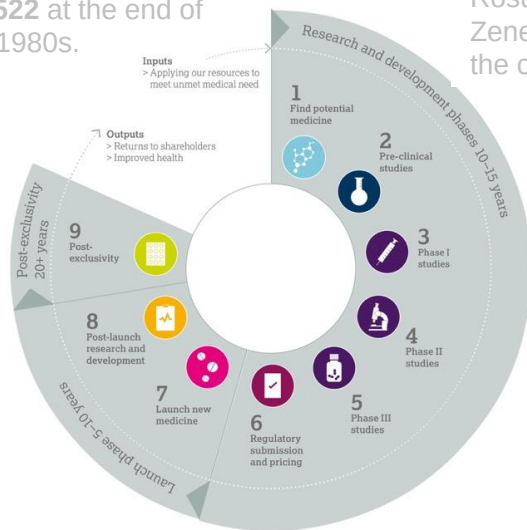
The **AURORA** study (phase III), together with **CORONA** and **JUPITER**, were outcome trials conducted between 2003 and 2008 as part of the GALAXY program

EudraCT
2004-001741-15

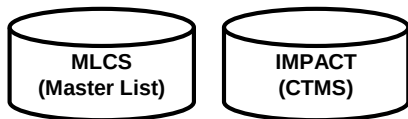
“study 96”

In 2015 AURORA was one of the many studies of interest in a collaboration project with the Montreal Heart Initiative (MHI)

ClinicalTrials.gov
NCT00240331



Legacy clinical study (meta)data



ZENECA (ICI)
4522IL/0096

AstraZeneca
D3562C00096

AURORA

EudraCT
2004-001741-15

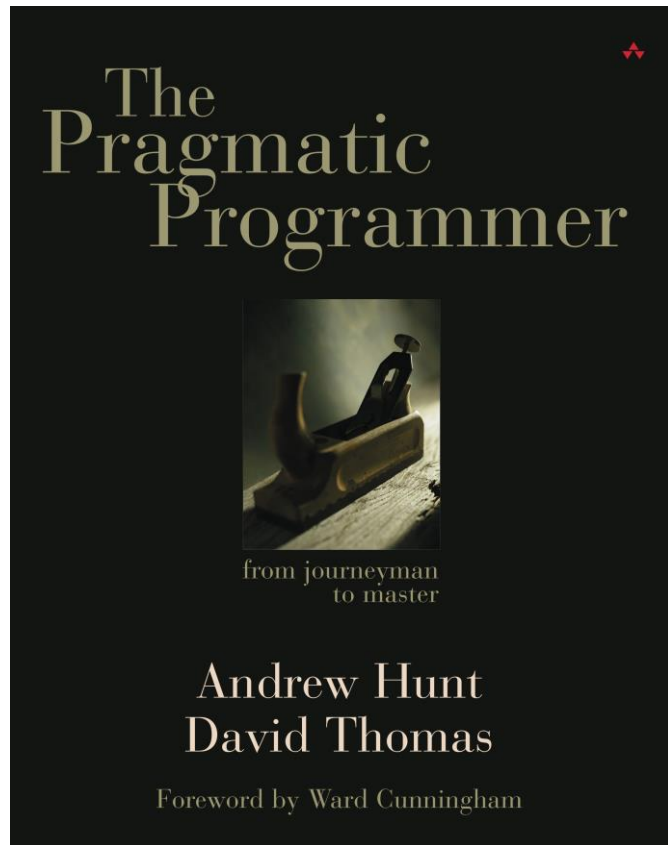
“study 96”

ClinicalTrials.gov
NCT00240331



The Pragmatic Programmer

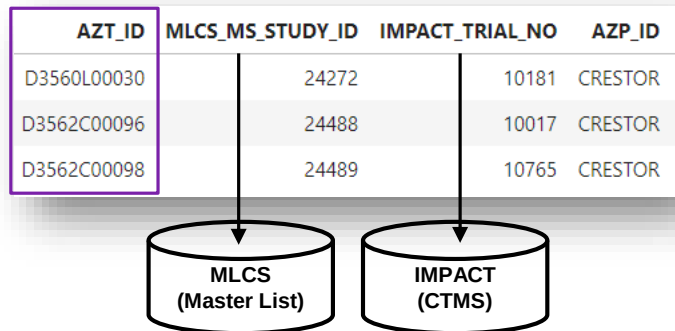
- Care About Your Craft
- Don't assume it -- Prove It
- Don't Live with Broken Windows
- Don't Use Manual Procedures
- Use Tracer Bullets to Find the Target
- Iterate the Schedule With the Code
- Some Things Are Better Done Than Described
- DRY – Don't Repeat Yourself
- Always Use Source Control
- and so on...



Source: <https://pragprog.com/>



Legacy clinical study (meta)data



ZENECA (ICI)
4522IL/0096

AstraZeneca
D3562C00096

AURORA

EudraCT
2004-001741-15

“study 96”

ClinicalTrials.gov
NCT00240331



Data Integration

AZT_ID	MLCS_MS_STUDY_ID	IMPACT_TRIAL_NO	AZP_ID
D3560L00030	24272	10181	CRESTOR
D3562C00096	24488	10017	CRESTOR
D3562C00098	24489	10765	CRESTOR

ZENECA (ICI)
4522IL/0096

Study Codes

Code value	Code type	Source	Source column	Source ID column	Source ID	Code URL
D3562C00096	AZ Trial ID	MLCS	AZT_ID	AZT_ID	D3562C00096	https://data.pharma.abc/clinicaltrial/D3562C00096
4522IL/0096	Other	MLCS	MS_STUDY_NAME	MS_STUDY_ID	24488	
AURORA	Acronym	MLCS	MS_STUDY_ACRONYM	MS_STUDY_ID	24488	
D3562C00096	DCode	MLCS	MS_STUDY_SYNONYMS	MS_STUDY_ID	24488	
4522IL/0096	Other	IMPACT	TRIAL_ALIAS_CODE	TRIAL_NO	10017	
D3562C00096	DCode	IMPACT	SPECIAL_NAME	TRIAL_NO	10017	
2004-001741-15	EudraCT	IMPACT	REGULATORY_REF	TRIAL_NO	10017	https://www.clinicaltrialsregister.eu/ctr-search/search?query=2004-001741-15
NCT00240331	NCT	IMPACT	REGULATORY_REF	TRIAL_NO	10017	https://clinicaltrials.gov/ct2/show/NCT00240331

AstraZeneca
D3562C00096

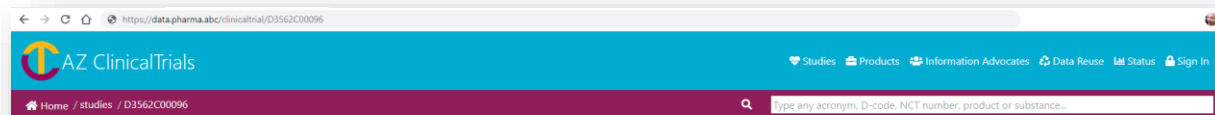
AURORA

EudraCT
2004-001741-15

ClinicalTrials.gov
NCT00240331



AZ ClinicalTrials



D3562C00096 (AURORA)

A study to evaluate the Use of Rosuvastatin in subjects On Regular haemodialysis: an Assessment of survival and cardiovascular events (AURORA). A double blind, randomised, multicentre, phase IIb, parallel-group study to compare the effects of rosuvastatin (10mg oral) with placebo on assessment of survival and cardiovascular events when given to subjects with end-stage renal failure on chronic haemodialysis treatment

The purpose of this study is to se...
haemodialysis. (From Brief Summ...



ZENECA (ICI)
4522IL/0096

Study Codes

Code value	Code type	Source	Source column	Source ID column	Source ID	Code URL
D3562C00096	AZ Trial ID	MLCS-NG	AZT_ID	AZT_ID	D3562C00096	https://data.pharma.abc/clinicaltrial/D3562C00096
4522IL/0096	Other	MLCS	MS_STUDY_NAME	MS_STUDY_ID	24488	
AURORA	Acronym	MLCS	MS_STUDY_ACRONYM	MS_STUDY_ID	24488	
D3562C00096	DCode	MLCS	MS_STUDY_SYNONYMS	MS_STUDY_ID	24488	
4522IL/0096	Other	IMPACT	TRIAL_ALIAS_CODE	TRIAL_NO	10017	
D3562C00096	DCode	IMPACT	SPECIAL_NAME	TRIAL_NO	10017	
2004-001741-15	EudraCT	IMPACT	REGULATORY_REF	TRIAL_NO	10017	https://www.clinicaltrialsregister.eu/ctr-search/search?query=2004-001741-15
NCT00240331	NCT	IMPACT	REGULATORY_REF	TRIAL_NO	10017	https://clinicaltrials.gov/ct2/show/NCT00240331

AstraZeneca
D3562C00096

AURORA

EudraCT
2004-001741-15

ClinicalTrials.gov
NCT00240331



AZ ClinicalTrials Web UI: Study Page

Combined information from different databases

← → ↻ https://data.pharma.abc/clinicaltrial/D3562C00096

AZ ClinicalTrials Studies Products Information Advocates Data Reuse Status Sign In

Home / studies / D3562C00096 🔍 Type any acronym, D-code, NCT number, product or substance...

D3562C00096 (AURORA)

A study to evaluate the Use of Rosuvastatin in subjects On Regular haemodialysis: an Assessment of survival and cardiovascular events (AURORA). A double blind, randomised, multicentre, phase IIb, parallel-group study to compare the effects of rosuvastatin (10mg oral) with placebo on assessment of survival and cardiovascular events when given to subjects with end-stage renal failure on chronic haemodialysis treatment

The purpose of this study is to see if rosuvastatin helps to reduce the number of heart attacks, strokes and cardiovascular deaths in patients undergoing haemodialysis. (From Brief Summary / clinicaltrials.gov)

CRESTOR®

Description

Metrics & Milestones

Reuse

Study Codes

AZ Trial ID: [D3562C00096](#)

Other: [4522IL/0096](#)

Acronym: [AURORA](#)

D-Code: [D3562C00096](#)

EudraCT: [2004-001741-15](#)

NCT: [NCT00240331](#)

[Details](#)

Common Study Data

Clinical plan description: CRESTOR

Outcomes: Renal (ESRD)

LSPC study: 0

Description: AURORA

Study status: Reported

Category: R&D Sponsored

Phase code: IIIB

Phase description: Phase 3

Delivery model: Internal

[Details](#)

Product

AZP ID: [CRESTOR](#)

Associated Product Information:

CRESTOR, Crestor, Zahron, Romazic, Rosucard, Roswera, Suvaradio, OTHER CV/GI, Other Project CV/GI, ROSUVASTATIN

Information advocate:

[So Man@astrazeneca.com](#)

[Details](#)

Indications

Indication group description:

Cardiovascular, Renal and Metabolism(CV)

Indication description:

Hypercholesterolaemia

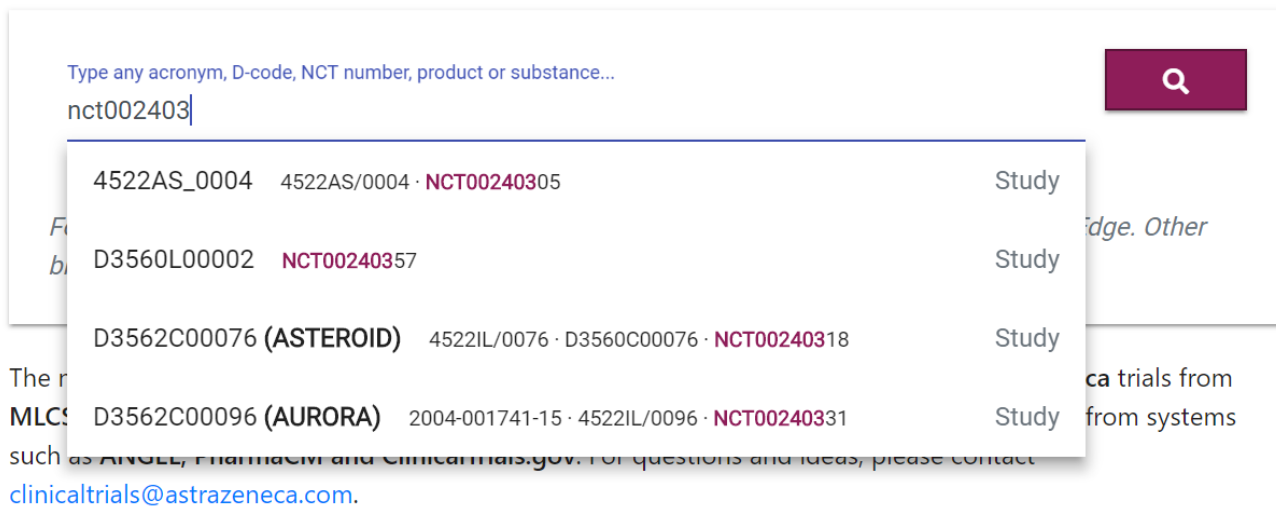
[Details](#)



AZ ClinicalTrials

Search

The most used feature in our web application user interface (UI) is a “Google-look-a-like” search box with autocompletion features.



The screenshot displays the AZ ClinicalTrials search interface. At the top, a search bar contains the text "nct002403" and a magnifying glass icon. Below the search bar, a dropdown menu lists several search results. Each result includes a study identifier, a study name in parentheses, and a study type. The results are as follows:

Study Identifier	Study Name	Study Type
4522AS_0004 · 4522AS/0004 · NCT00240305		Study
D3560L00002 · NCT00240357		Study
D3562C00076 (ASTEROID) · 4522IL/0076 · D3560C00076 · NCT00240318		Study
D3562C00096 (AURORA) · 2004-001741-15 · 4522IL/0096 · NCT00240331		Study

AZ ClinicalTrials Web UI: Study Page

Links to external study pages

AZ ClinicalTrials

Home / studies / D3562C00096

D3562C00096 (AURORA)

A study to evaluate the Use of Rosuvastatin in subjects On Regular haemodialysis: an Assessment of survival and cardiovascular events (AURORA). A double blind, randomised, multicentre, phase IIIB, parallel-group study to compare the effects of rosuvastatin (10mg oral) with placebo on assessment of survival and cardiovascular events when given to subjects with end-stage renal failure on chronic haemodialysis treatment.

The purpose of this study is to see if rosuvastatin helps to reduce the number of heart attacks, strokes and cardiovascular deaths in patients undergoing haemodialysis. (*From Brief Summary / clinicaltrials.gov*)

Study Codes

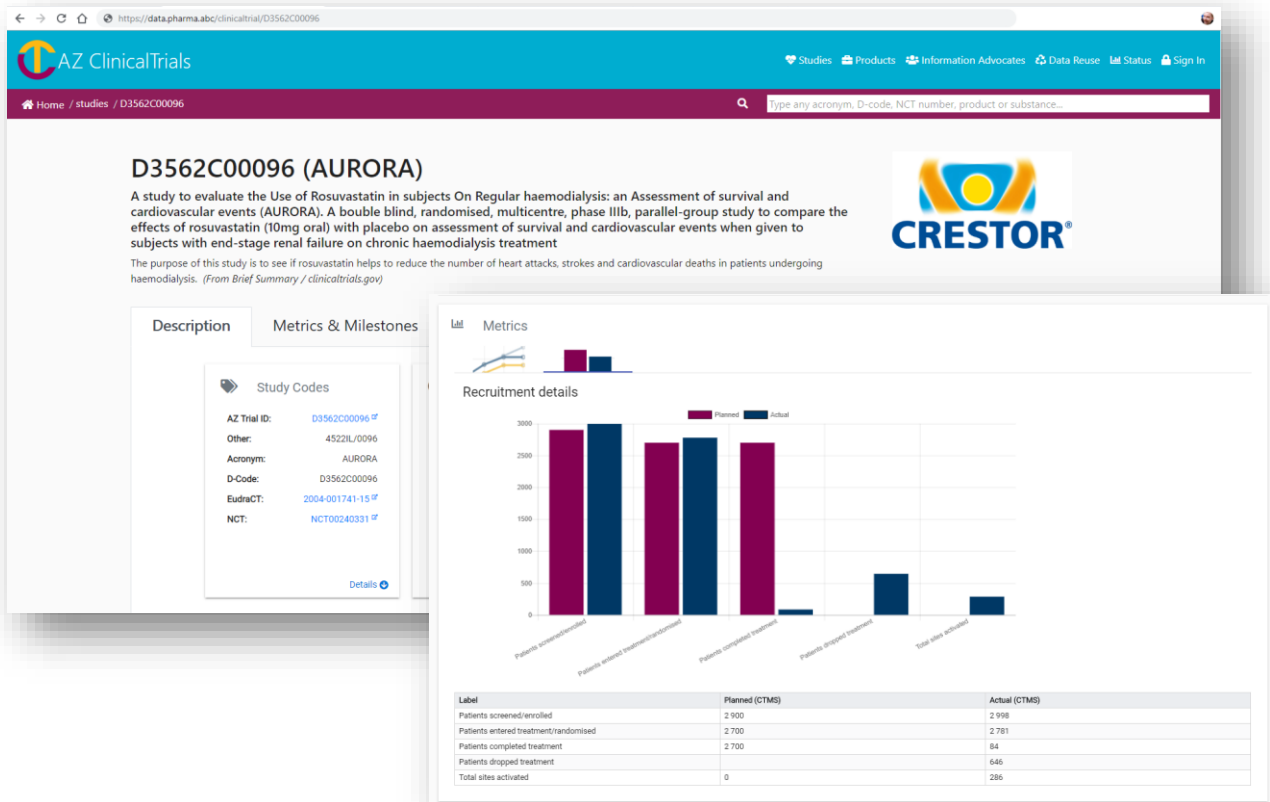
Code value	Code type	Source	Source column	Source ID column	Source ID
D3562C00096	AZ Trial ID	MLCS-NG	AZT_ID	AZT_ID	D3562C00096
4522IL/0096	Other	MLCS	MS_STUDY_NAME	MS_STUDY_ID	24488
AURORA	Acronym	MLCS	MS_STUDY_ACRONYM	MS_STUDY_ID	24488
D3562C00096	DCode	MLCS	MS_STUDY_SYNONYMS	MS_STUDY_ID	24488
4522IL/0096	Other	IMPACT	TRIAL_ALIAS_CODE	TRIAL_NO	10017
D3562C00096	DCode	IMPACT	SPECIAL_NAME	TRIAL_NO	10017
2004-001741-15	EudraCT	IMPACT	REGULATORY_REF	TRIAL_NO	10017
NCT00240331	NCT	IMPACT	REGULATORY_REF	TRIAL_NO	10017

<https://www.clinicaltrialsregister.eu/ctr-search/search?query=2004-001741-15>

<https://clinicaltrials.gov/ct2/show/NCT00240331>

AZ ClinicalTrials Web UI: Study Page

Metrics & Milestones details and graphs



AZ ClinicalTrials Web UI: Study Page

Information to support Reuse



D3562C00096 (AURORA)

A study to evaluate the Use of Rosuvastatin in subjects On Regular haemodialysis: an cardiovascular events (AURORA). A double blind, randomised, multicentre, phase III effects of rosuvastatin (10mg oral) with placebo on assessment of survival and cardiac subjects with end-stage renal failure on chronic haemodialysis treatment

The purpose of this study is to see if rosuvastatin helps to reduce the number of heart attacks, strokes and cardiovascular events in subjects on regular haemodialysis. (From Brief Summary / clinicaltrials.gov)

Description

Metrics & Milestones

Reuse



Study Codes

AZ Trial ID: [D3562C00096](#)
Other: [4522IL/0096](#)
Acronym: [AURORA](#)
D-Code: [D3562C00096](#)
EudraCT: [2004-001741-15](#)
NCT: [NCT00240331](#)

[Details](#)

Common Study Data

Clinical plan description: CRESTOR
Outcomes: Renal (ESRD)
LSPC study: 0
Description: AURORA
Study status: Reported
Category: R&D Sponsored
Phase code: IIB
Phase description: Phase 3
Delivery model: Internal

[Details](#)

Description

Metrics & Milestones

Reuse



Data Reuse Overview

Consent Review (iCART): [Link](#)
Internal Sharings: [2](#)
External Sharings: Not yet available
New Data Reuse Proposal: [Link](#)

[Details](#)

Data & Document Locations

SDS:	86
SDS2:	8
EntimICE:	4
CARTDEV:	43
Clinii:	5
GEL:	9

[Details](#)

Human Biological Samples (ABBA)

Total Subjects:	1 219
Total Samples:	3 188
Source Countries:	20
Expiry date(s):	06-OCT-202306-OCT-2028
Previous sample requests:	3
Genetic samples available:	YES
Range of DNA (µg):	0.831 - 627.1

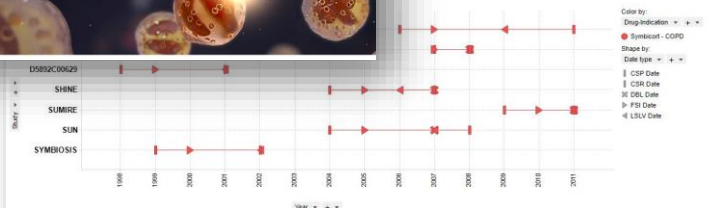
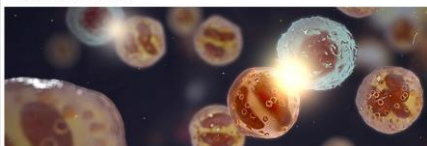
[Details](#)

AZ ClinicalTrials Programmatic access

Other applications and notebooks

The MetaDataHub - Integrating, searching and presenting study descriptive information

Elisabeth Nyman, Kerstin Forsberg & Michail Doulis
PHUSE EU Connect 2018



AZT ID	Alias	Study Title	Primary Objective	Primary Variable	Inclusion Criteria	Exclusion Criteria
D598C00001	Alisa	A 12-Week Double-blind, Double-dummy, Randomized, Parallel-group, Multicenter Efficacy and Safety Study of SYMBICORT (pDEI 2 x 160/4.5 µg bid and 2 x 160/4.5 µg bid) compared to placebo and formoterol TID 2x4.5 µg bid with regard to FEV1 pre-dose and 1-hour post-dose.	To show that SYMBICORT (pDEI 2x160 4.5 µg bid) is effective in patients with COPD, when compared to placebo and formoterol TID 2x4.5 µg bid	FEV1 pre-dose and 1-hour post-dose	- A clinical diagnosis of COPD with symptoms for more than 2 years	- A history of asthma - Patients taking oral steroids
D598C00002	SHINE	A 6-Week Double-blind, Double-dummy, Randomized, Parallel-group, Multicenter Efficacy & Safety Study of SYMBICORT (pDEI 2 x 160/4.5 µg bid & 80/4.5 µg bid) compared to placebo and formoterol TID 2x4.5 µg bid with regard to FEV1 pre-dose and 1-hour post-dose.	To show that SYMBICORT (pDEI 2x160 4.5 µg bid & 80/4.5 µg bid) is effective in patients with COPD, when compared to placebo and formoterol TID 2x4.5 µg bid	FEV1 pre-dose and 1-hour post-dose	- A clinical diagnosis of COPD with symptoms for more than 2 years	- A history of asthma - Patients taking oral steroids
D598C00003	803	A Phase IIIb, 12-Week, Double-blind, Double-dummy, Randomized, Parallel-group, Multicenter Efficacy and Safety Study of SYMBICORT (pDEI 160/4.5 µg bid & 80/4.5 µg bid) compared to placebo and formoterol TID 2x4.5 µg bid with regard to FEV1 pre-dose and 1-hour post-dose.	To compare the efficacy of SYMBICORT (pDEI 160/4.5 µg bid & 80/4.5 µg bid) with regard to FEV1 activation SD with that of formoterol TID	Number of COPD exacerbations FEV1 pre-dose and 1-hour post-dose The randomized	- A current clinical diagnosis of COPD with symptoms for more than 2 years - A history of asthma	- Patients taking oral steroids - A history of asthma

AZT ID	Alias	Study Name	Phase	Length	Unit	CSQ1 Date	CSQ2 Date	CSQ3 Date	P/SI Date	LSLV Date	CSQ link	CSQ link	Enrolled	Randomized
D598C00001	Alisa	D598C00001 HCT602	III	12	month	2004-02	2004-02	2004-01	2007-11	2008-02-08	https://clinicaltrials.gov/study/NCT00109008	https://clinicaltrials.gov/study/NCT00109008	2316	1078
D598C00002	SHINE	D598C00002 HCT602	IIA	6	month	2004-02	2004-02	2007-11	2007-03	2006-12	https://clinicaltrials.gov/study/NCT00109008	https://clinicaltrials.gov/study/NCT00109008	2381	1075
D598C00003	803	D598C00003 Exacast	III	12	month	2004-02	2004-02	2007-11	2007-01	2006-10	https://clinicaltrials.gov/study/NCT00109008	https://clinicaltrials.gov/study/NCT00109008	2381	1278
D598C00004	SHINE	D598C00004 HCT602	IIA	3	month	2004-02	2004-02	2007-11	2007-01	2006-11-03	https://clinicaltrials.gov/study/NCT00109008	https://clinicaltrials.gov/study/NCT00109008	2381	1278

jupyter 2018-04-05 AZCT compared to CT.gov Last Checkpoint: Yesterday at 10:07 AM (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

Not Trusted Python 3

AZCT compared to CT.gov

This is explorative work to compare internal information about studies with external information. This is done using the R&D Data Hub and consist of the following steps using the R&D Data Hub.

- Get Internal information about studies via AZ ClinicalTrials (AZCT)
- Get External information about studies via ClinicalTrials.gov (CT.gov)
- Cross reference the identifiers by look ups
- Merge the two look ups into a Cross reference table
- Combine the Cross reference table with the internal and the External information

Later on this will potentially be incorporated in the R&D Data Hub as one or more standard views (Canonical or ApplicationSpecific)

Source: <https://github.com/azclinet/azclinet/blob/master/notebooks/browse/2018/2018-04-05%20AZCT%20compared%20to%20CT.gov>

```
In [21]: import pandas as pd
from numpy import nan
import requests
pd.set_option('max_colwidth', 800)
from pandas import ExcelWriter
```

Internal information about studies via AZ ClinicalTrials (AZCT)

Information Advocate info

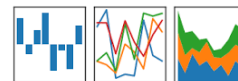
IncI name: internal via PRD look-up, using a AZCT endpoint

```
In [22]: IAs = pd.read_json('https://clinicaltrials-api.astrazeneca.net/v2/nicmg/information_advocates')
IAs = IAs[['IA_ID', 'displayname']]
IAs = IAs.rename(columns={'displayname': 'CSD_NAME'})
```

Basic Study info from AZCT

Combining data from three Canonical Views

- STUDY
- PRODUCT
- STUDY_MILESTONE
- STUDY_TIMELINE



Behind all of this: The Clinical Study URI

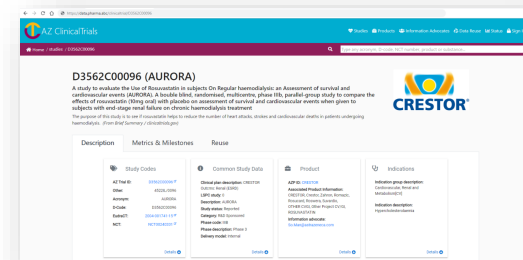
URI = *Uniform Resource Identifier* *)

Generalised example
not the actual AZ one

Example URI: <https://data.pharma.abc/clinicaltrial/D3562C00096>

- **Global namespace:** https://data.pharma.abc
- **Resource type:** clinicaltrial
- **AstraZeneca Trial ID (AZT_ID):** e.g. D3562C00096

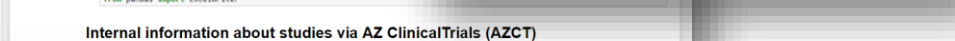
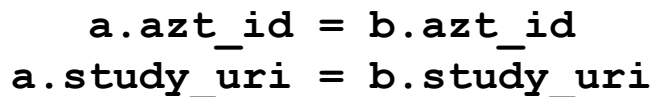
A Study URI resolves
to a Study Page URL



*) **IRI** = *Internationalized Resource Identifier* works using a Unicode typeset, URIs works on ASCII characters



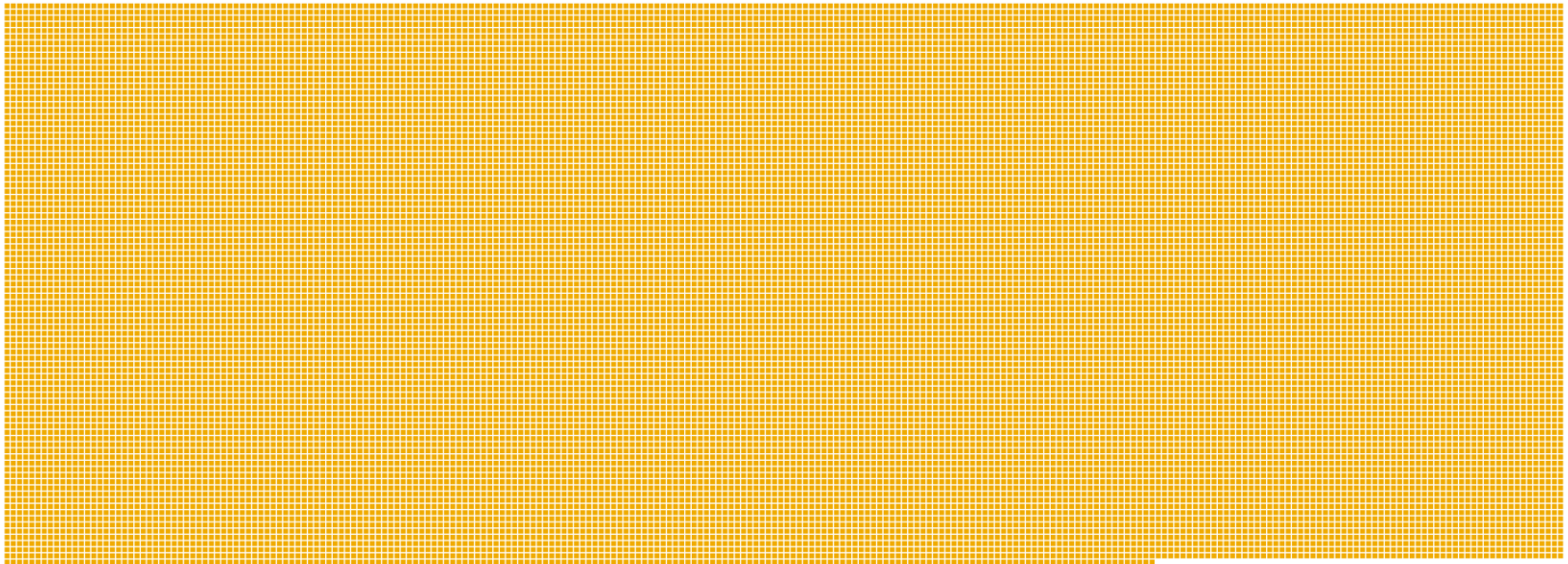
“The marriage point”



DEMO TIME!



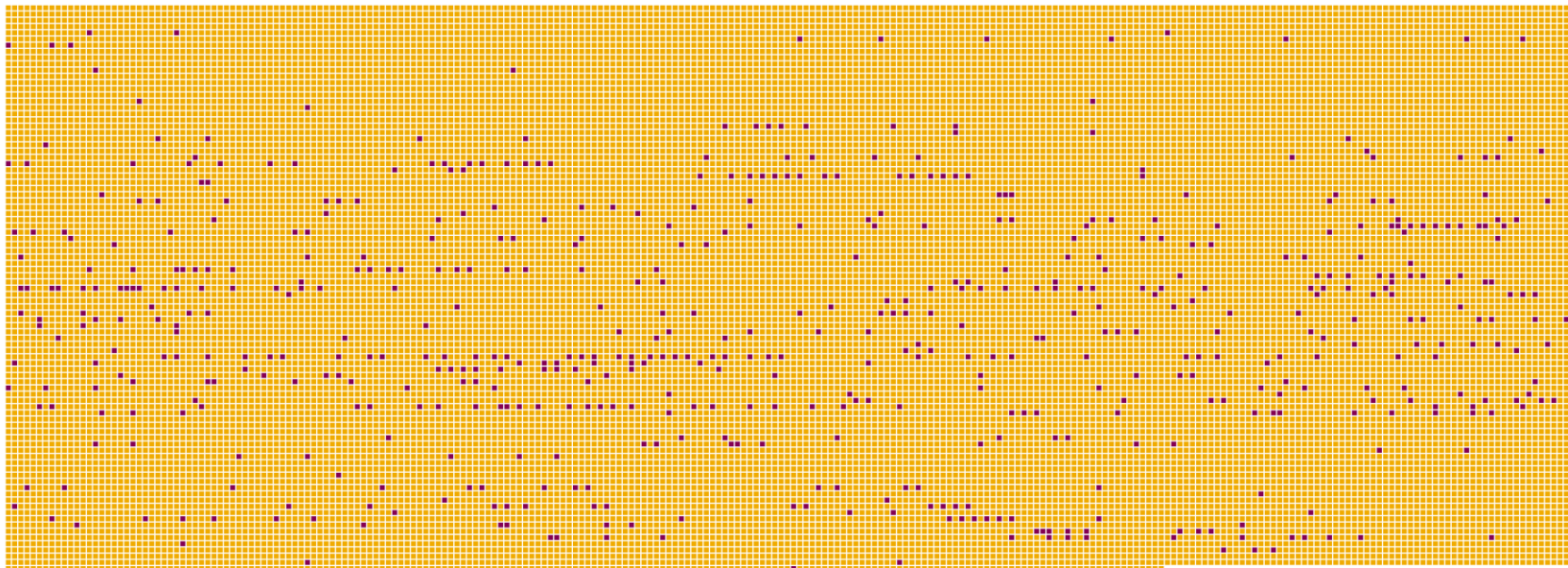
All AZ ClinicalTrials Studies



Each block represents a STUDY entity in AZ ClinicalTrials

< Sort > C- C+

Studies with an Acronym

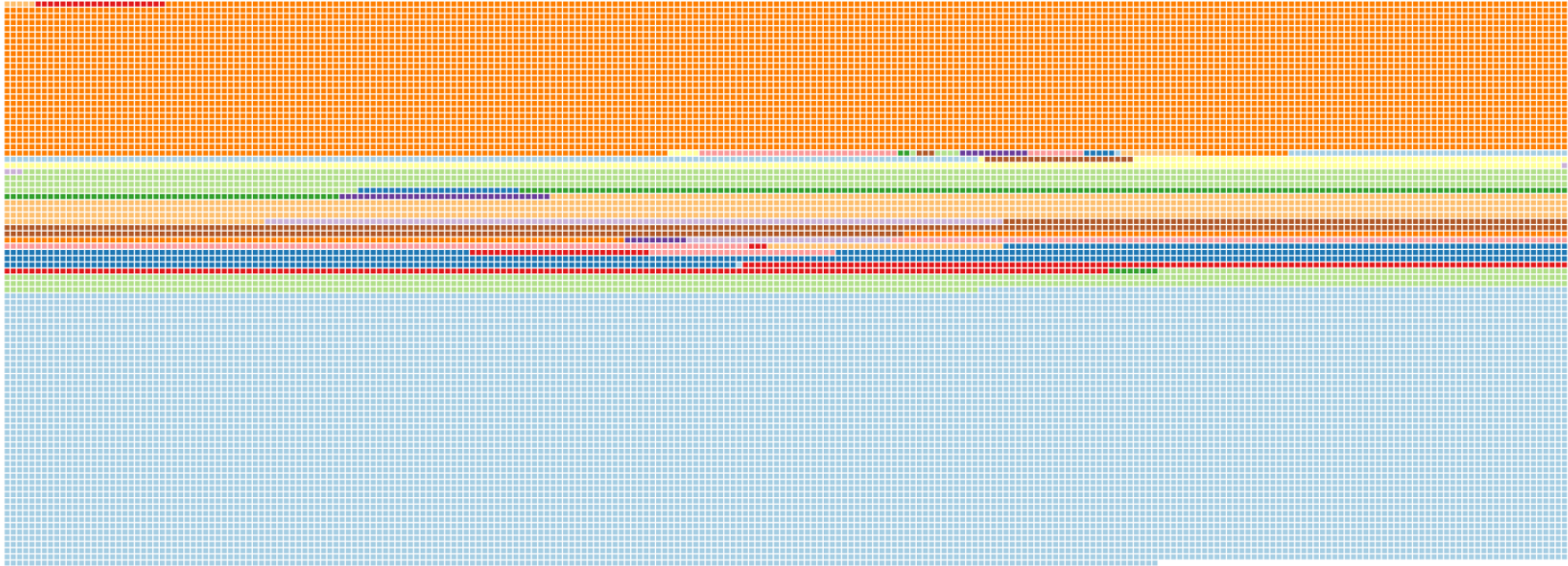


< Sort > C- C+

Studies with an Acronym

ACRONYM: AURORA
AZT_ID: D3562C00096
SYNONYMS: 2004-001741-15, 4522IL/0096,
AURORA, D3562C00096,
NCT00240331
CODE ASTRAZENEGA_D_CODE,
MATCHES: EUDRACT_CODE, NCT_CODE,
ZENEGA_CODE

Studies with any Code Matches



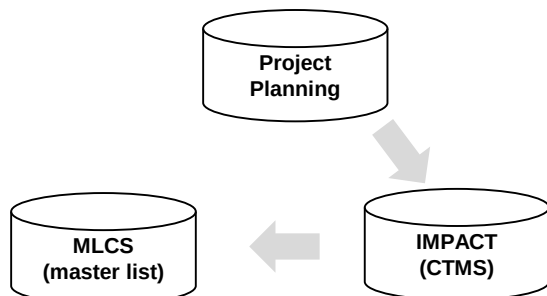
Detailed issues when assigning Study URIs

With a broad scope, "all" studies (20.000+ studies)

- 1. Multiple codes in different database columns**
- 2. Acronyms not consistently recorded**
3. Implicit connection between sponsor and public identifiers
4. Duplicated study records (same study code used)
5. Acquired studies

Detailed issues when assigning Study URIs

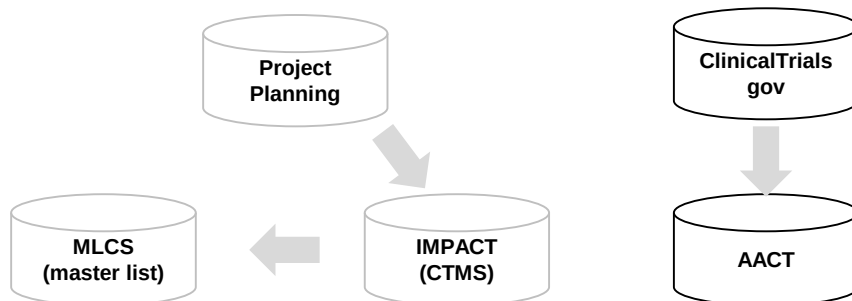
1. Multiple codes in different database columns



MLCS				IMPACT		
		Primary Column	Secondary Column			
MS_STUDY_ACRONYM	MS_STUDY_ID	MS_STUDY_NAME	MS_STUDY_SYNONYMS	TRIAL_NO	Primary Column	Secondary Column
					TRIAL_ALIAS_CODE	SPECIAL_NAME
AURORA	24488	4522IL/0096	D3562C00096	10017	4522IL/0096	D3562C00096
CORONA	24489	D3562C00098	4522IL/0098	10765	D3562C00098	D3562C00098
JUPITER	24272	4522US/0011	D3560L00030	10181	4522US/0011	D3560L00030

Detailed issues when assigning Study URIs

2. Acronyms not consistently recorded



MLCS				IMPACT		
MS_STUDY_ACRONYM	Primary Column		Secondary Column	TRIAL_NO	Secondary Column	
	MS_STUDY_ID	MS_STUDY_NAME	MS_STUDY_SYNONYMS		TRIAL_ALIAS_CODE	SPECIAL_NAME
AURORA	24488	4522IL/0096	D3562C00096	10017	4522IL/0096	D3562C00096
CORONA	24489	D3562C00098	4522IL/0098	10765	D3562C00098	D3562C00098
JUPITER	24272	4522US/0011	D3560L00030	10181	4522US/0011	D3560L00030

NCT_ID	SPONSOR	ACRONYM	CODE_VALUE	CODE_TYPE
NCT00206310	AstraZeneca		4522IL/0098	org_study_id
NCT00206310	AstraZeneca		CORONA	secondary_id
NCT00206310	AstraZeneca		D3562C00098	secondary_id
NCT00239681	AstraZeneca		4522US/0011	secondary_id
NCT00239681	AstraZeneca		D3560L00030	org_study_id
NCT00239681	AstraZeneca		Jupiter	secondary_id
NCT00240331	AstraZeneca		4522IL/0096	org_study_id
NCT00240331	AstraZeneca		D3562C00096	secondary_id

NIH U.S. National Library of Medicine		Find Studies ▾	About Studies ▾
ClinicalTrials.gov			
Home > Search Results > Study Record Detail			
AURORA: Crestor 10mg Versus Placebo in Subjects With End-stage Renal Disease (ESRD)			

Ideas for next steps

Beyond application of specific Study URIs

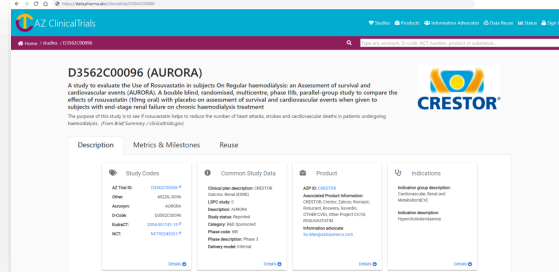
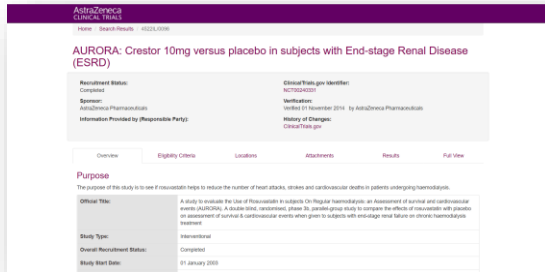
- Enterprise policy to **assign URIs** to more master and reference entities such as Product, Projects and Indications
- Enterprise infrastructure to **resolve URIs**, both internally and externally

Ideas for next steps

Beyond application of specific Study URIs

- Enterprise policy to **assign URIs** to more master and reference entities such as Product, Projects and Indications
- Enterprise infrastructure to **resolve URIs**, both internally and externally

<https://data.astrazeneca.com/clinicaltrial/D3562C00096>



Instead of today's external URL:

<https://astrazenecagrouptrials.pharmacm.com/ST/Submission/View?id=117>



Fundamental Enablers

Proactive project

- Clear Study Data Entity ownership
- Domain knowledge
- Small, agile team – close collaboration with business partners



**Shared, Global and Persistent Identifiers
for Clinical Studies (Study URI)**



4,000 IT experts shaping the future of biotech innovation

- International company with strong belief in diversity
- Global career development and work with inspiring colleagues
- Be part of the Fourth Industrial Revolution, connecting digital technology with science
- Developing life-changing medicines for the benefit of society

If it's time to re-evaluate your aspirations, **search AstraZeneca Careers**, we will be glad to tell you more!

