APPLICATION OF DATA LINKAGE METHODS AND PROCEDURES AT THE NATIONAL CANCER REGISTRY OF UKRAINE

Anton Ryzhov, Yevgeniy Gorokh

National Cancer Registry of Ukraine

Ukrainian National Cancer Institute

NATIONAL™ CANCER INSTITUTE
FOUNDED IN 1920

ENCR Scientific Meeting and General Assembly

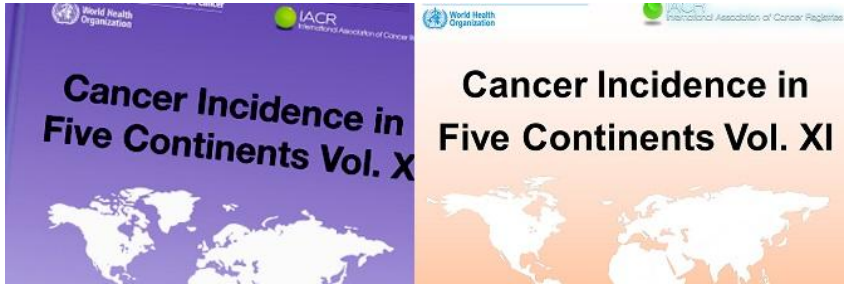
26-28 September 2018 • Copenhagen • Denmark

Ukraine

- Area: 603,000 sq. km
- Population (2017): ~44 mln.
  - median age: 37.4 (m), 43.7 (f)
  - life exp.: 67.4(m); 77.1(f)
- 27 administrative units (380,000 – 4,400,000 inhabitants)
- GDP per capita (2017 est.): 8,700$

State Cancer Registration System of Ukraine & NCRU

- Introduced in 1953
  - Passive case finding
  - Active mortality and follow up
- National Cancer Registry of Ukraine (NCRU)
  - Consists of **27** regional PBCRs
  - Officially established in 1996
  - National coverage since 2002
  - **26** hospital-based cancer registries since 2015



ENCR EUROPEAN NETWORK OF CANCER REGISTRIES MEMBER REGISTRY

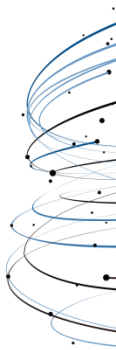
State Cancer Registration System of Ukraine & NCRU

- The state cancer registration system of Ukraine is working over 65 years based on the principle of compulsory reporting of the pre-determined medical paper forms according to the patient's place of residence
- Wherever a cancer patient was diagnosed or treated, the information on the cancer case is transmitted to the patient's "oblast oncological center" with PBCR
  - "Oblast" means large province where patient permanently lives, further subdivided into "rayons" (i.e. districts)

NATIONAL™ CANCER INSTITUTE
FOUNDED IN 1920

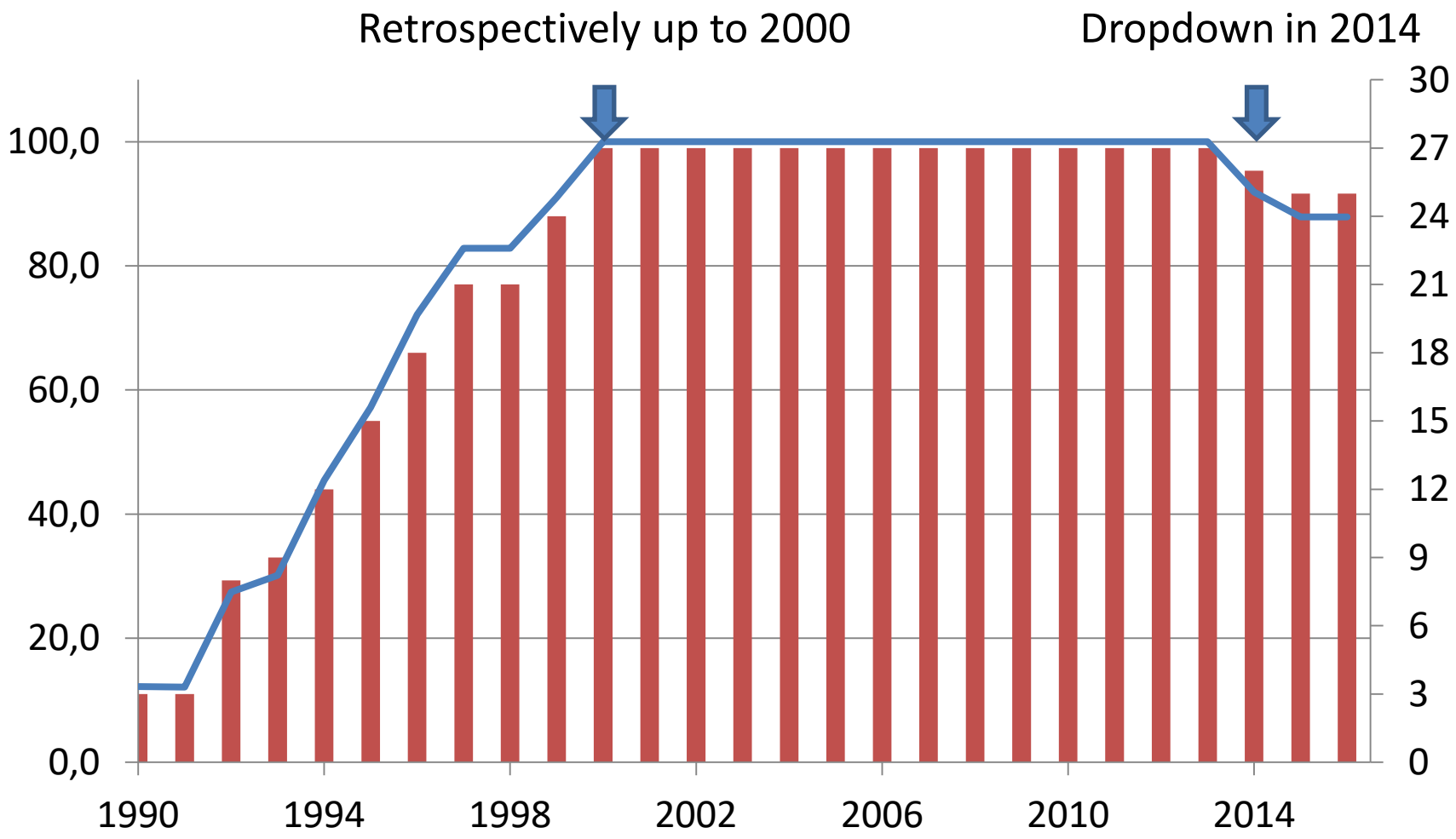
State Cancer Registration System of Ukraine & NCRU

- NCRU compiles all regional databases annually to produce nation-wide statistics
  - 4.5 million records (2018)
- Change in population coverage (2014) due to AR Crimea occupation and war conflict in the East of Ukraine, involving
  - ~15-20% of population of Ukraine (2014 est.)
  - ~13.3% of area of Ukraine
  - More than 1mln internal migrants
- Health care reform towards E-Health

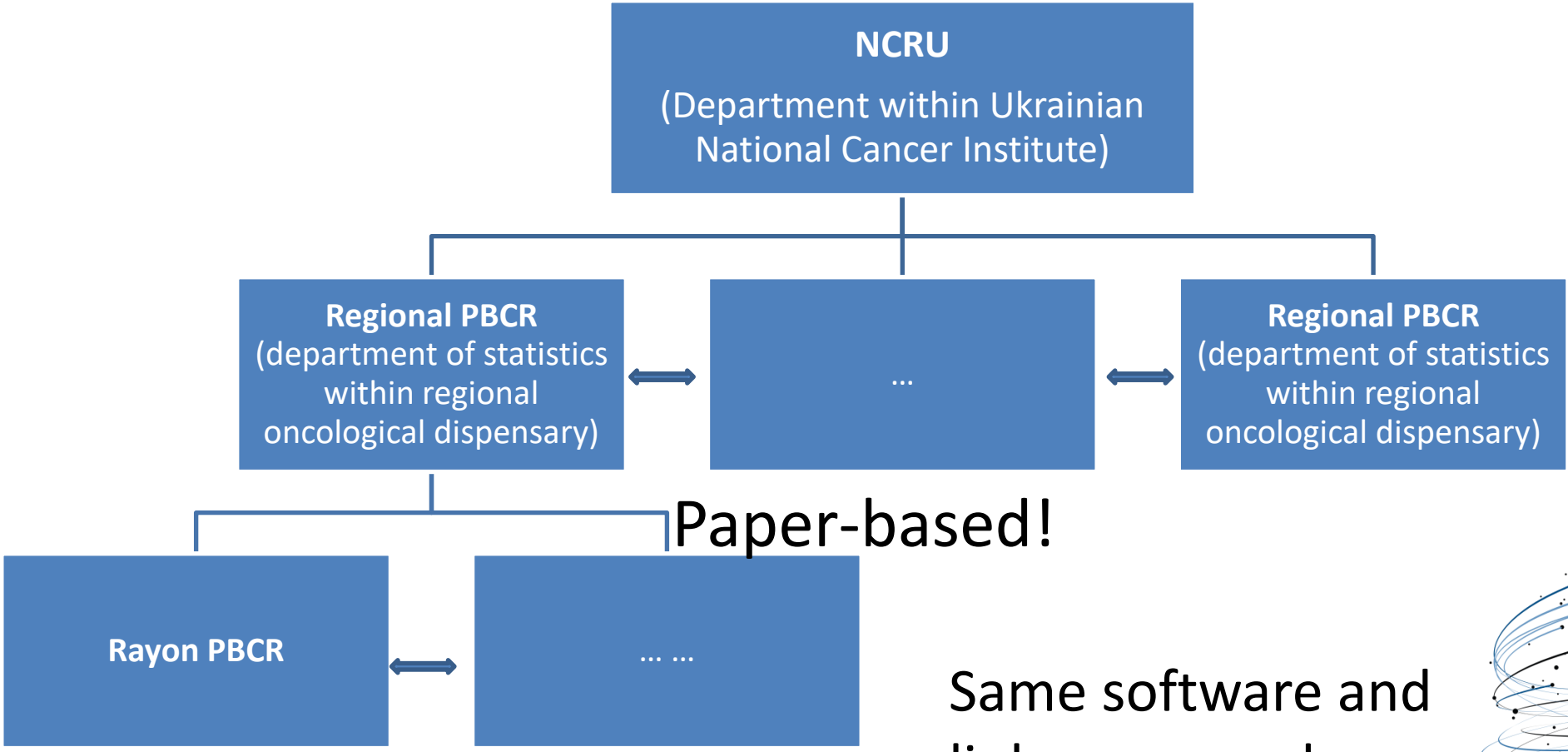


NATIONAL™ CANCER INSTITUTE
FOUNDED IN 1920

NCRU population coverage (incidence)



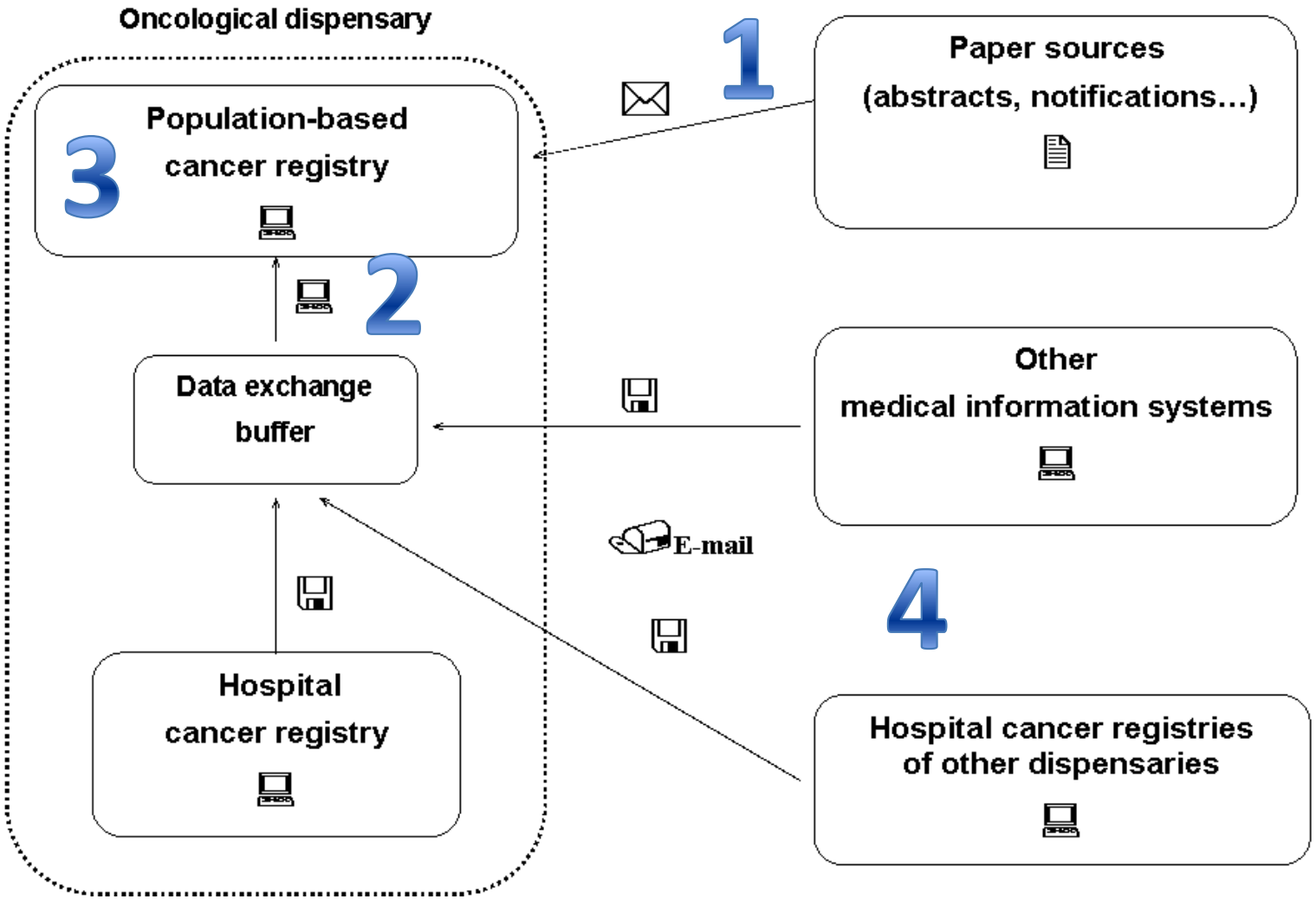
Structure of the NCRU



Same software and linkage procedures

NATIONAL™ CANCER INSTITUTE
FOUNDED IN 1920

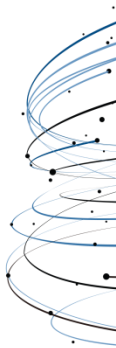
NCRU: data flows and linkage points



NATIONAL™ CANCER INSTITUTE
FOUNDED IN 1920

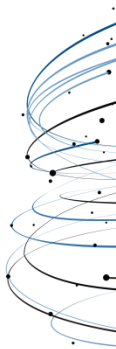
Linkage points: data input and update (1)

- Data items: Last/Middle/First name, Birth year
- New case with single tumor?
- Multiple primary?
- Prevalence case changing residence within a PBCR's area?



Linkage points: HBCR and PBCR (2)

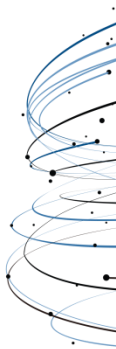
- HBCR data is more reliable and complete
- From 28% to 74% of new cases are diagnosed in the oncological institutions where HBCR is operating
  - Automated transfer of information
  - Reduce of costs for running PBCR
  - Improvement of quality of the PBCR's data
- From 60% to 90% of patients receive in- or out-patient treatment in the nearest oncological institutions
  - Information on treatment received
  - Vital status and last date of contact update



NATIONAL™ CANCER INSTITUTE
FOUNDED IN 1920

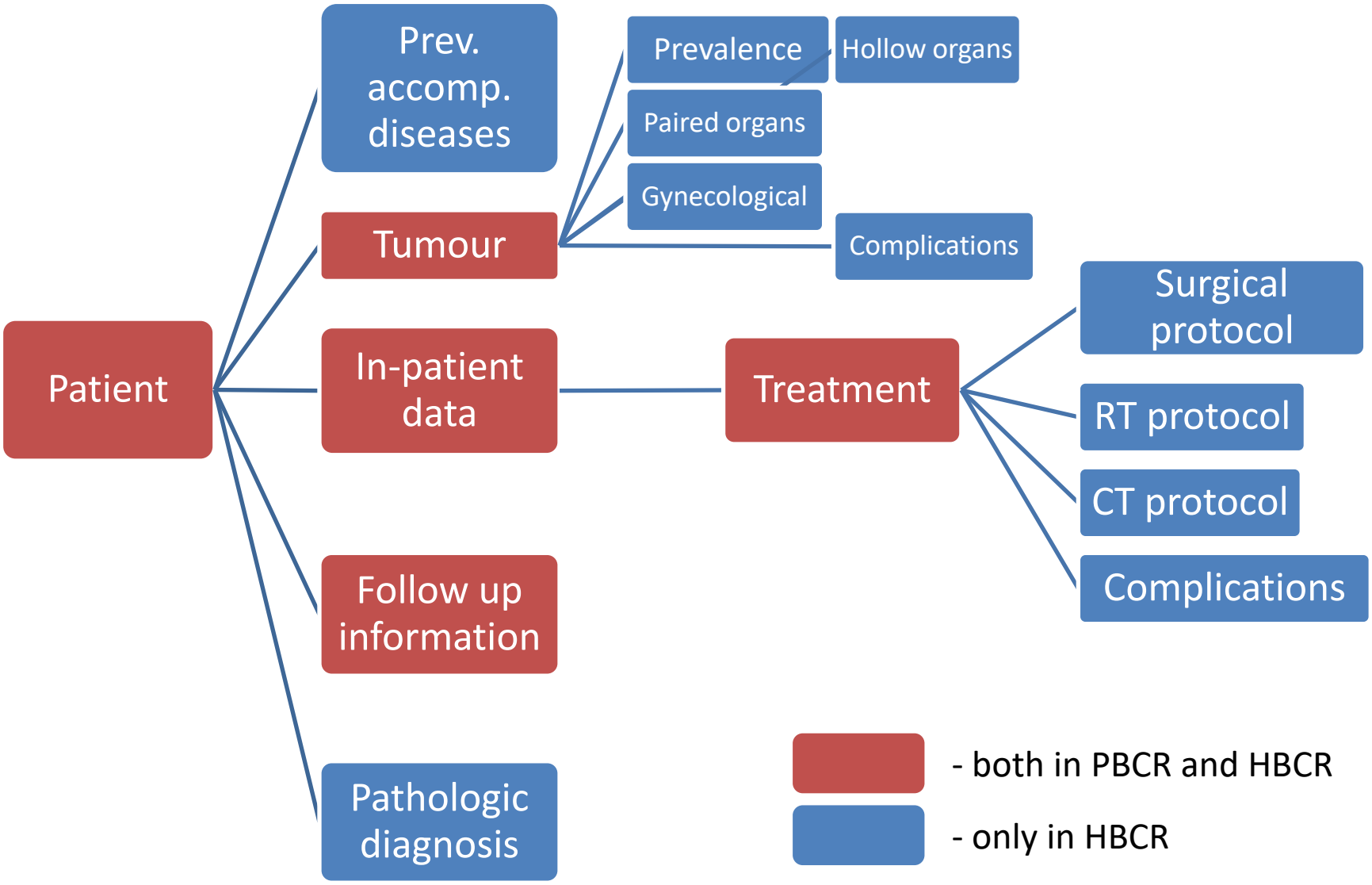
Linkage points: HBCR and PBCR (2)

- Data items to match:
  - all non-empty variables in linking record(s)
- Multiple primaries?
- Residents / migrants?
- Same software and technology
  - Data may be obtained from any institution with HBCR software installed
  - E-medical records and notifications exchange



NATIONAL™ CANCER INSTITUTE
FOUNDED IN 1920

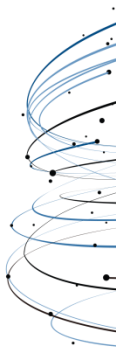
Linkage points: HBCR and PBCR (2)



NATIONAL™ CANCER INSTITUTE
FOUNDED IN 1920

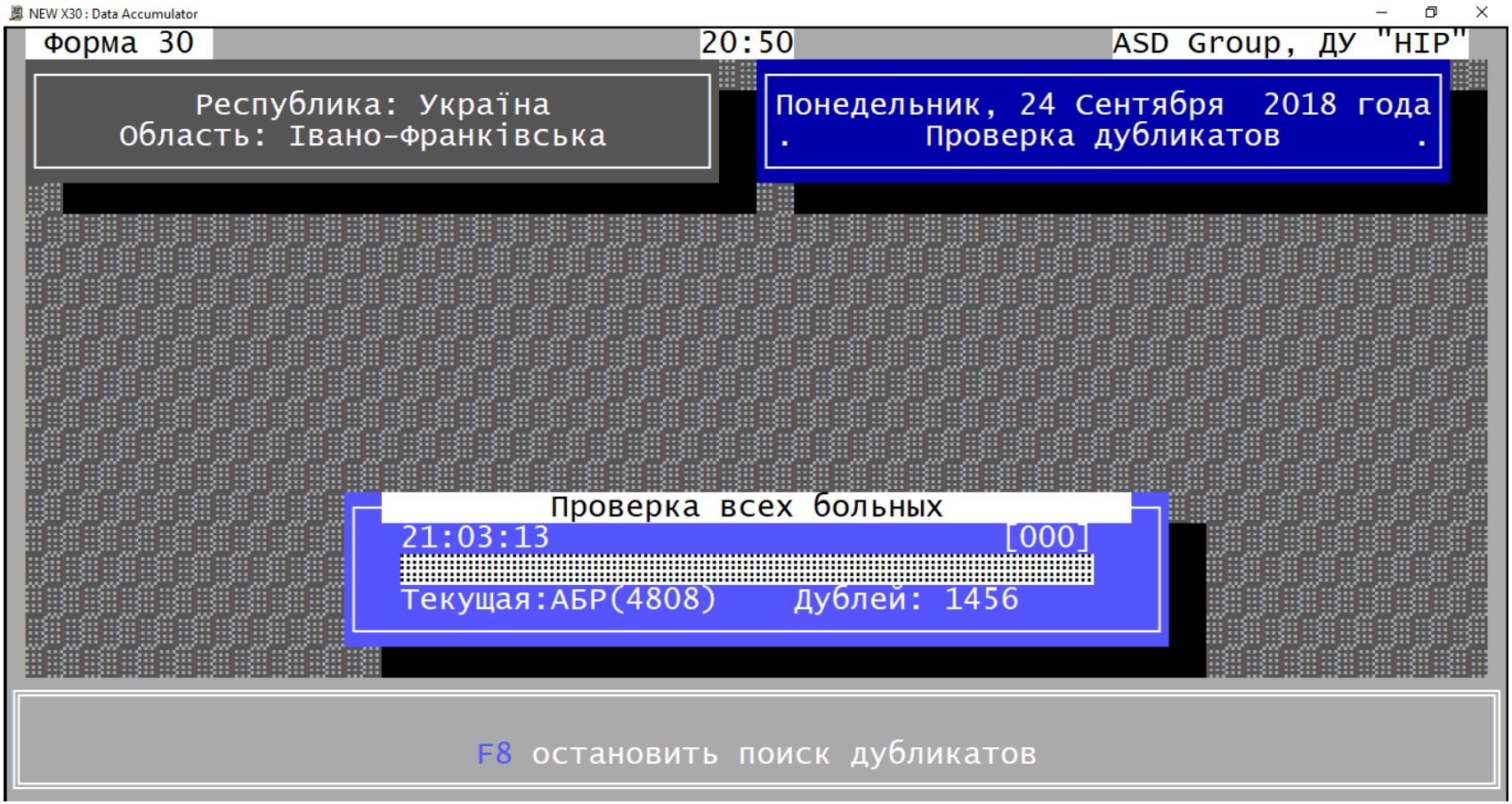
Linkage points: Search for duplicates (3)

- Different blocks to compare:
  - Parts of First/Middle/Last Name
  - Birth year
- List of “false positives”
- Automated IARC/IACR multiple primaries check



NATIONAL™ CANCER INSTITUTE
FOUNDED IN 1920

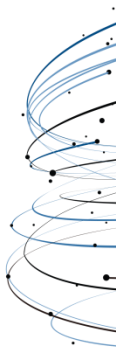
Linkage points: Search for duplicates (3)



NATIONAL™ CANCER INSTITUTE
FOUNDED IN 1920

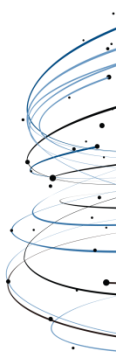
Linkage points: External cohorts (4)

- Custom blocks to compare:
  - Parts of the First/Middle/Last Name
  - Birth year
  - Region codes
  - Address
  - ICD-10 code
  - Etc.



Linkage points: External cohorts (4)

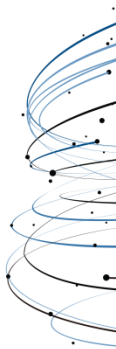
- Synching names and formats of data items in data sets;
- Cleaning and standardization of First/Middle/Last names:
  - upper-case and Ukrainian character conversion;
  - replacing common misprints, Ukrainian and local spellings, synonymic names using NCRU checklists;
  - replacing typical Russian and Ukrainian surname endings and letter combinations using special checklist



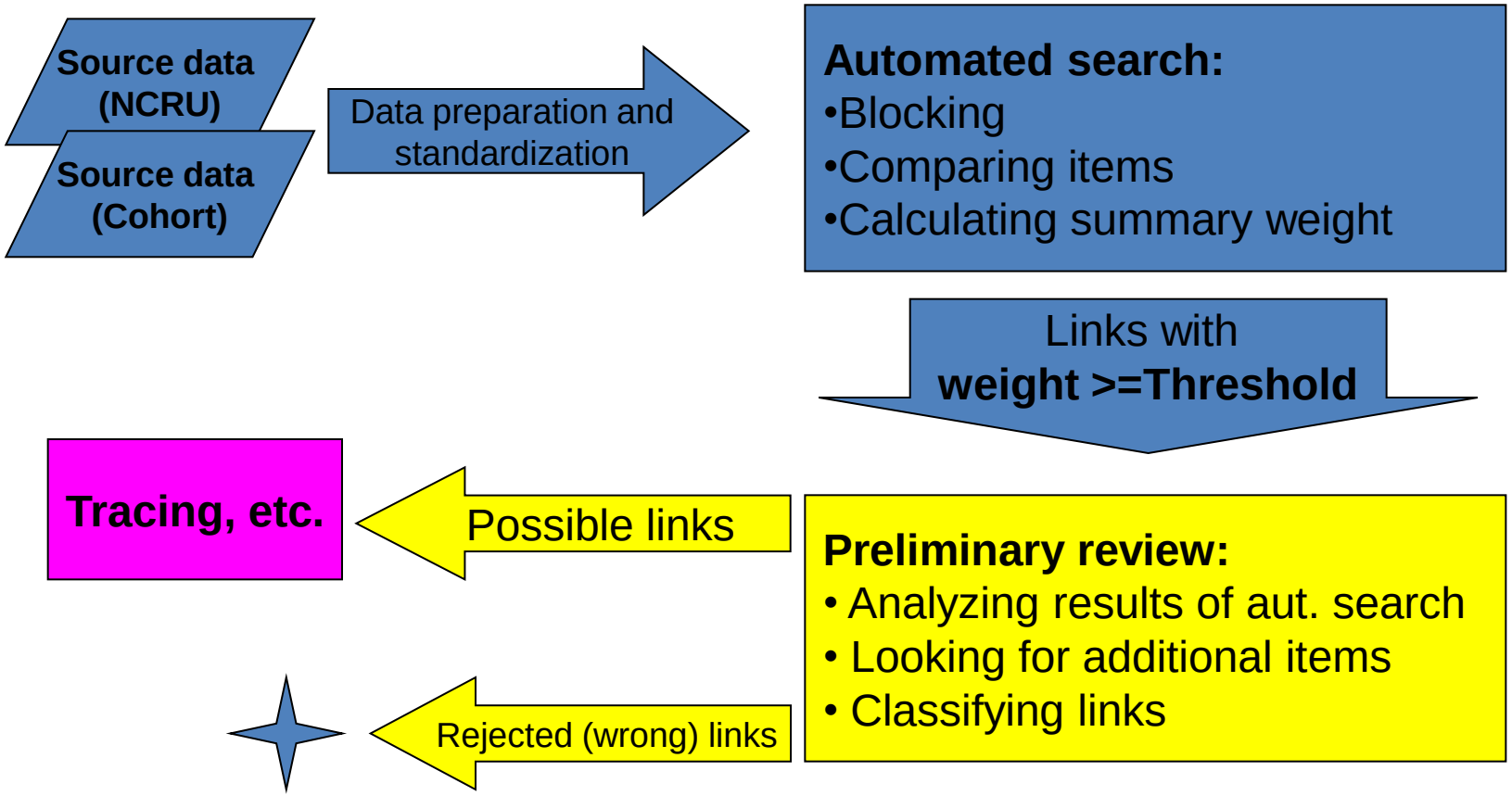
NATIONAL™ CANCER INSTITUTE
FOUNDED IN 1920

Linkage points: External cohorts (4)

- standardization of the ADDRESS strings;
  - partial Ukrainian-Russian translation;
  - parsing and removing non-significant lexemes (e.g. "Str.", "street", etc.)
- synchronizing of numeric regional codes used at the NCRU and the Cohort database;
- extracting the Year of Birth from the Date of Birth
- Etc.



Linkage points: External cohorts (4)

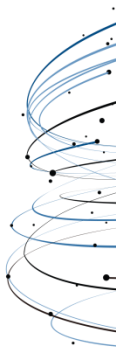


NATIONAL™ CANCER INSTITUTE
FOUNDED IN 1920

Linkage points: External cohorts (4)

	FIELD1	FIELD2	LFUNCTION	LWEIGHT	LDECREASEW
•	SURNAME	SURNAME	LINK_NAME	100	100
•	NAME	NAME	LINK_NAME	100	100
•	PATRONYM	PATRONYM	LINK_NAME	100	100
•	BIRTHDATE	BIRTHDATE	LINK_DATE	100	60
•	ADRES	ADRES	LINK_ADDR	100	0
•	MESTO	MESTO	LINK_MESTO	50	50
•	SURNAME	SURNAME	LINK_LEV	100	0
•	CHERNOB	CHERNOB	LINK_ALL	100	0
•	ADRES	F3_ADRES	LINK_ADDR	100	0

- ✓ Each function returns value between 0 and 1, as a measure of similarity
- ✓ For a positive value, LWEIGHT used as a multiplier, and a result of comparison is added to the summary weight.
- ✓ For a zero value (negative result of comparison), LDECREASEW subtracted from the summary weight.



Linkage points: External cohorts (4)

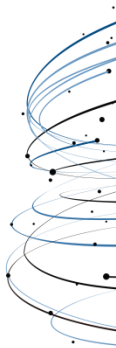
NCRU data used	TIME	All links (found after autom. Search)	Rejected after review	"Probable links"	"Possible links"
2000-2004, 5143 records	July, 2005	961	769	96	96
2005, 1099 records	May, 2006	73	44	21	8

Study: Chernobyl Liquidators (clean-up workers) Cohort And the National Cancer Registry of Ukraine (2006-2007), Leukaemia in male birth cohort

NATIONAL™ CANCER INSTITUTE
FOUNDED IN 1920

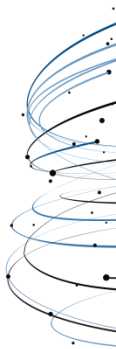
Linkage points: External cohorts (4)

- Ukrainian-American thyroid cohort
- Clinical survival studies: patients' follow up
- Regional death registries
- Future plans: linkage with E-health subregistries



Results

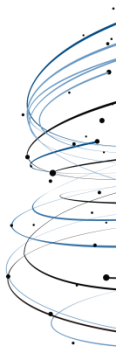
- Same technology may be applied to regular NCRU's database at the national level
  - Extract and convert into required format
  - External software
- The blocking packages and set of data items and linking functions to be used in linkage procedures were determined to produce a single-value weight
- Patient's personal information (first, last name, sex, birth date, place of residence) as well as tumor description (date of diagnosis) were compared



NATIONAL™ CANCER INSTITUTE
FOUNDED IN 1920

Results

- Applying data linkage procedures at the national level the percentage of potential duplicate records in the whole dataset was less than 2.5%, out of which
  - 11.1% of records had follow-up status “alive” in both matching records
  - 2.8% had status “dead” in one of the matching records



NATIONAL™
CANCER INSTITUTE
FOUNDED IN 1920

Thank you!